

Université de Montréal

**Analyse épigénétique intégrative pour identifier de  
nouveaux biomarqueurs dans la leucémie myéloïde aiguë  
causée par des translocations chromosomiques  
de type *KMT2A***

Par Thomas Milan

Programme de Biologie Moléculaire, option Générale  
Faculté de médecine

Thèse présentée à la Faculté de Médecine  
en vue de l'obtention du grade de Philosophiae Doctor  
en Biologie Moléculaire  
option Générale

Juin, 2021

© Thomas Milan, 2021

Université de Montréal  
Faculté de Médecine

Cette thèse intitulée :

**Analyse épigénétique intégrative pour identifier de nouveaux biomarqueurs dans la  
leucémie myéloïde aiguë causée par des translocations chromosomiques  
de type *KMT2A***

Présentée par :  
Thomas Milan

A été évaluée par un jury composé des personnes suivantes :

Julie Lessard  
Président-rapporteur

Brian Wilhelm  
Directeur de recherche

Benoit Laurent  
Membre du jury

François Mercier  
Examineur externe

David Hipfner  
Représentant de la doyenne

## Résumé

La leucémie est une forme de cancer qui affecte les cellules du système hématopoïétique. Selon la lignée cellulaire affectée et la vitesse de développement du cancer, la leucémie peut être myéloïde ou lymphoïde, aiguë ou chronique, respectivement. Chez les enfants, elles sont souvent caractérisées par la présence de translocations chromosomiques, impliquant notamment le gène *KMT2A*. L'impact biologique de ces fusions de gènes, connues pour être des perturbateurs épigénétiques, est encore mal compris.

Afin d'étudier spécifiquement les conséquences de la présence de fusion impliquant le gène *KMT2A*, un modèle leucémique humain chez la souris a été mis en place. Le modèle utilisé consiste à induire de manière rétrovirale l'expression d'une fusion oncogénique dans des cellules souches hématopoïétiques et progénitrices d'un unique donneur sain. Ces cellules sont ensuite injectées dans des souris immunodéficientes pour produire une leucémie aiguë myéloïde ou lymphoïde après quelques semaines. L'utilisation de ce modèle leucémique vise à définir les gènes qui sont régulés de manière épigénétique et essentiels dans le processus de leucémogénèse médié par une translocation chromosomique faisant intervenir le gène *KMT2A*.

La première partie des travaux cartographie les changements génétiques et épigénétiques à chacun des stades de la leucémogénèse causée par la fusion *KMT2A-MLL3*. Nous avons cartographié les changements épigénétiques tels que la méthylation de l'ADN (Methyl-seq), les modifications des histones (ChIP-seq) et l'accessibilité de la chromatine (ATAC-seq), puis les avons corrélés avec les niveaux d'expression des gènes (RNA-seq). Nous avons observé que les leucémies myéloïdes aiguës présentent un phénotype global d'hypométhylation tandis que les changements d'expression après l'addition de la fusion ont mis en évidence l'inactivation de gènes associés aux cellules souches et des altérations dans d'autres gènes impliqués dans la leucémogénèse tels que *S100A8/9*. Nos données d'ATAC-seq ont montré qu'il y avait relativement peu de changements spécifiques à la leucémie myéloïde aiguë et que la grande majorité correspondait à des régions de chromatine ouvertes et à des régions contenant des motifs pour des facteurs de transcription précédemment observés dans d'autres types de cellules sanguines. L'analyse des marques d'histones associées à des promoteurs actifs suggère également un potentiel rôle du récepteur CCR1 et de son ligand spécifique CCL23. Finalement,

nos résultats suggèrent que la transformation leucémique par la fusion *KMT2A-MLLT3* implique des modifications épigénétiques minimales qui requièrent également la coopération des réseaux transcriptionnels utilisés dans les cellules sanguines normales.

La deuxième partie de cette thèse s'intéresse à la fusion de gènes *KMT2A-MLLT4*, une translocation chromosomique peu étudiée mais pour laquelle le pronostic vital des patients est connu pour être défavorable et pire que celui des patients porteurs de la fusion *KMT2A-MLLT3*. L'extension de notre modèle à la fusion *KMT2A-MLLT4* nous permet d'appliquer les mêmes approches que précédemment et de détailler les différences génétiques et épigénétiques entre ces deux fusions, jusqu'à maintenant jamais caractérisées. Nous avons pu observer une baisse globale d'expression dans un groupe de gènes intervenant dans les processus ribosomiaux et traductionnels. Par ailleurs, *PROM1 (CD133)* fait office de potentiel candidat biomarqueur permettant la distinction entre ces deux translocations chromosomiques tandis que le gène *LPL* pourrait jouer un rôle dans la leucémogénèse médiée par la fusion de gènes *KMT2A-MLLT4*.

En conclusion, l'étude des mécanismes à chacun des stades du développement leucémique nous a fourni une meilleure compréhension des changements épigénétiques intervenant dans le processus de leucémogénèse causé par des réarrangements de type *KMT2A*. Une meilleure caractérisation de la pathophysiologie de la leucémie pourrait permettre d'explorer des avenues thérapeutiques plus ciblées.

## Mots-clés

Leucémie, épigénétique, *KMT2A-MLLT3*, *KMT2A-MLLT4*, méthylation de l'ADN, marques d'histones, chromatine, facteurs de transcription, RNA-seq, ChIP-seq, Methyl-seq, ATAC-seq

## Abstract

Leukemia is a form of cancer that affects blood cells. Depending on the affected cell lineage and the rate at which the cancer grows, leukemia can be myeloid or lymphoid, or acute or chronic, respectively. In children, they are often characterized by the presence of chromosomal translocations, in particular involving the *KMT2A* gene. The biological impact of these gene fusions, known to be epigenetic disruptors, is still poorly understood.

To study the consequences of the presence of gene fusions involving *KMT2A*, we have developed a human leukemia model. The model consists of transducing hematopoietic stem and progenitor cells (CD34<sup>+</sup>) from a single healthy donor with a retrovirus bearing an oncogenic fusion. These cells are injected into immunodeficient mice to produce acute myeloid or lymphoid leukemia after a few weeks. By using this model, we aim to define genes that are epigenetically regulated and essential in the process of leukemogenesis mediated by *KMT2A* gene fusions.

The first part of this thesis characterized the genetic and epigenetic changes at each step of leukemogenesis caused by *KMT2A-MLLT3* gene fusion. We investigated epigenetic changes such as DNA methylation (Methyl-seq), histone marks (ChIP-seq), and chromatin accessibility (ATAC-seq) and correlated these with expression changes (RNA-seq). We observed that acute myeloid leukemias exhibit a profound hypomethylation phenotype while expression changes after addition of the fusion highlighted the loss of stem cell associated genes and alterations in other genes implicated in leukemogenesis such as *SI00A8/9* in the early stages of leukemic transformation. Our ATAC-seq data showed that there were relatively few changes specific to acute myeloid leukemia and that the vast majority corresponded to open chromatin regions and clusters of transcription factors previously seen in other types of blood cells. Examination of ChIP-seq data for active histone marks revealed that leukemia specific expression of the chemokine *CCL23* can enable autocrine signalling through its cognate receptor, *CCR1*. Our results suggest that *KMT2A-MLLT3* induces minimal changes in the epigenome while co-opting the normal transcriptional machinery to drive leukemogenesis.

The second part of this thesis focuses on *KMT2A-MLLT4* gene fusion, another chromosomal translocation for which the vital prognosis of patients is known to be worse than that of patients

carrying the *KMT2A-MLLT3* fusion. The extension of our model to the *KMT2A-MLLT4* fusion allows us to apply the same approaches and to characterize the genetic and epigenetic differences between these two different leukemias. We were able to observe a dramatic decrease in the expression level of a group of genes involved in ribosomal and translational processes. Furthermore, *PROM1* (*CD133*) acts as a potential biomarker candidate which might be used to make the distinction between these two leukemias. *LPL* gene might play a role in leukemogenesis mediated by *KMT2A-MLLT4* gene fusion.

In conclusion, studying the mechanisms at each stage of leukemic development has provided us with a better understanding of the epigenetic changes involved in the process of leukemogenesis mediated by *KMT2A* rearrangements. A better characterization of the pathophysiology of leukemia could make it possible to eventually develop more targeted therapeutic treatments.

## **Keywords**

Leukemia, epigenetics, *KMT2A-MLLT3*, *KMT2A-MLLT4*, DNA methylation, histone marks, chromatin, transcription factors, RNA-seq, ChIP-seq, Methyl-seq, ATAC-seq

# Table des matières

Résumé.....	iii
Abstract .....	v
Table des matières.....	vii
Liste des tableaux .....	xii
Liste des figures.....	xiii
Liste des abréviations.....	xv
Remerciements .....	xxiv
CHAPITRE 1 .....	1
1 Introduction.....	2
1.1 ORGANISATION GÉNÉRALE DES CHAPITRES .....	2
1.2 LE SYSTÈME SANGUIN.....	4
1.2.1 <i>L'hématopoïèse : un processus assurant l'homéostasie sanguine</i> .....	4
1.2.2 <i>À la base de la hiérarchie : les cellules souches hématopoïétiques</i> .....	5
1.2.3 <i>Un processus finement régulé</i> .....	8
1.2.3.1 Cytokines.....	8
1.2.3.2 Facteurs de transcription.....	9
1.3 ÉPIGÉNÉTIQUE.....	10
1.3.1 <i>Méthylation de l'ADN</i> .....	10
1.3.2 <i>Modifications des histones</i> .....	11
1.4 LA LEUCÉMIE MYÉLOÏDE AIGUË.....	13
1.4.1 <i>Définition</i> .....	13
1.4.2 <i>Manifestations cliniques et méthodes de diagnostic</i> .....	13
1.4.3 <i>Classification de la leucémie myéloïde aiguë</i> .....	14
1.4.4 <i>Facteurs de pronostic des leucémies myéloïdes aiguës</i> .....	16
1.4.5 <i>Le paysage génétique des leucémies myéloïdes aiguës</i> .....	17
1.5 LES LMA CAUSÉES PAR UNE FUSION DE GÈNES DE TYPE <i>KMT2A</i> .....	20
1.5.1 <i>Le gène KMT2A sauvage</i> .....	20

1.5.1.1	Historique .....	20
1.5.1.2	Structure de la famille MLL/KMT2.....	21
1.5.1.3	Fonction de la protéine KMT2A dans l'hématopoïèse.....	22
1.5.1.4	Structure détaillée de la protéine KMT2A sauvage .....	23
1.5.1.4.1	Partie N-terminale de la protéine KMT2A.....	23
1.5.1.4.2	Partie C-terminale de la protéine KMT2A.....	24
1.5.1.4.3	Clivage de la protéine KMT2A.....	25
1.5.2	<i>Les translocations chromosomiques faisant intervenir le gène KMT2A</i> .....	28
1.5.2.1	Origine moléculaire de la translocation.....	28
1.5.2.2	Caractéristique des partenaires de fusion .....	29
1.5.2.2.1	Classification des partenaires de fusion.....	29
1.5.2.2.2	Une fréquence variable en clinique .....	30
1.5.2.2.3	Partenaires de fusion et pronostics cliniques.....	30
1.5.2.2.4	Un programme d'expression génique en commun .....	32
1.5.2.2.5	Une mécanique qui varie selon les protéines partenaires.....	34
1.5.2.2.5.1	Le cas de la fusion <i>KMT2A-MLLT3</i> .....	34
1.5.2.2.5.2	Le cas de la fusion <i>KMT2A-MLLT4</i> .....	36
1.6	HYPOTHÈSE ET OBJECTIFS DE RECHERCHE .....	37
1.6.1	<i>Problématique</i> .....	37
1.6.2	<i>Hypothèse générale et objectifs</i> .....	38
1.7	RÉFÉRENCES.....	39
<b>CHAPITRE 2 .....</b>		<b>51</b>
<b>2 Étude des changements épigénétiques dans le modèle de leucémies humaines de type</b>		
<b><i>KMT2A-MLLT3</i> .....</b>		<b>52</b>
2.1	PRÉFACE .....	52
2.2	CONTRIBUTIONS .....	52
2.3	RÉFÉRENCE DE LA PUBLICATION.....	53
2.4	ARTICLE.....	54
2.4.1	<i>Abstract</i> .....	55
2.4.2	<i>Introduction</i> .....	56
2.4.3	<i>Results</i> .....	59
2.4.3.1	Transcriptional consequences of KM3 expression.....	59



2.4.3.2	Sequential changes in DNA methylation are not globally correlated with expression changes.	65
2.4.3.3	Epigenetic analysis reveals potential CCR1-CCL23 autocrine signaling in KM3-acute myeloid leukemia.....	74
2.4.3.4	KM3-mediated epigenetic changes are minimal relative to normal blood cells. ....	88
2.4.4	<i>Discussion</i> .....	92
2.4.5	<i>Methods</i> .....	96
2.4.5.1	Patients' samples and model leukemia generation.....	96
2.4.5.2	Human CD34 <sup>+</sup> hematopoietic stem and progenitor cell isolation and cell culture ..	97
2.4.5.3	Chromatin immunoprecipitation, RNA-sequencing and ATAC-sequencing .....	98
2.4.5.4	DNA methylation capture sequencing .....	99
2.4.5.5	Bioinformatics analysis of ChIP-sequencing and ATAC-sequencing .....	100
2.4.5.6	Mapping methyl-sequencing reads and methylation analysis.....	101
2.4.5.7	Data-sharing statement .....	101
2.4.6	<i>Acknowledgments</i> .....	102
2.4.7	<i>Author contributions</i> .....	102
2.4.8	<i>References</i> .....	103
<b>CHAPITRE 3</b>	.....	<b>110</b>
<b>3</b>	<b>Caractérisation des différences génétiques et épigénétiques entre les modèles de leucémies humaines de type <i>KMT2A-MLLT3</i> et <i>KMT2A-MLLT4</i> .....</b>	<b>111</b>
3.1	INTRODUCTION .....	111
3.2	OBJECTIFS DU PROJET .....	115
3.3	MATÉRIELS ET MÉTHODES.....	117
3.3.1	<i>Échantillons de patients leucémiques</i> .....	117
3.3.2	<i>Génération des modèles de leucémies <i>KMT2A-MLLT4</i></i> .....	117
3.3.3	<i>RNA-seq, ChIP-seq et ATAC-seq – Protocoles et analyses bioinformatiques</i> .....	117
3.4	RÉSULTATS .....	119
3.4.1	<i>Les signatures transcriptomiques du modèle <i>KM4</i> sont similaires à celles retrouvées chez les patients leucémiques porteurs de la fusion <i>KM4</i></i> .....	119
3.4.1.1	Les données du modèle proches des observations cliniques .....	119
3.4.1.2	Un modèle utile pour la production de réplicats biologiques .....	122

3.4.2	<i>La comparaison des profils d'expression génique entre les deux modèles leucémiques révèle une baisse d'expression de gènes intervenant dans les mécanismes de traduction.....</i>	124
3.4.2.1	Des signatures distinctes.....	124
3.4.2.2	Une baisse d'expression de certains gènes codant pour des facteurs de transcription pourrait être à l'origine de ces signatures.....	128
3.4.3	<i>L'intégration des données transcriptomiques et épigénétiques permet de révéler la présence de potentiels biomarqueurs spécifiques au modèle KM4.....</i>	132
3.4.3.1	<i>PROM1</i> , potentiel biomarqueur de la LMA médiée par la fusion <i>KM4</i> .....	132
3.4.3.2	<i>LPL</i> est un exemple de gènes régulés de manière épigénétique .....	134
3.5	DISCUSSION.....	138
3.6	RÉFÉRENCES.....	142
<b>CHAPITRE 4 .....</b>		<b>147</b>
<b>4</b>	<b>Leucémie pédiatrique : vers des modèles plus précis .....</b>	<b>148</b>
4.1	PRÉFACE .....	148
4.2	CONTRIBUTIONS .....	148
4.3	RÉFÉRENCE DE LA PUBLICATION.....	148
4.4	ARTICLE.....	149
4.4.1	<i>Abstract</i> .....	150
4.4.2	<i>Introduction</i> .....	151
4.4.3	<i>Mining patient cohorts to understand the disease</i> .....	154
4.4.4	<i>Using cell lines as model systems for pediatric leukemias</i> .....	157
4.4.5	<i>Murine models of pediatric AML</i> .....	161
4.4.6	<i>Human models of pediatric leukemia</i> .....	167
4.4.7	<i>Conclusions</i> .....	173
4.4.8	<i>Acknowledgements</i> .....	174
4.4.9	<i>References</i> .....	175
<b>CHAPITRE 5 .....</b>		<b>191</b>
<b>5</b>	<b>Discussion, conclusion et perspectives .....</b>	<b>192</b>
5.1	UN MODÈLE AVEC DES AVANTAGES ET DES INCONVÉNIENTS.....	192
5.1.1	<i>Contrôler le site d'intégration de la fusion oncogénique</i> .....	192

5.1.2	<i>Se rapprocher encore davantage de la réalité clinique</i> .....	193
5.2	UNE MÉCANIQUE BIOLOGIQUE QUI SOULÈVE ENCORE DES QUESTIONS.....	194
5.3	RECHERCHER DES BIOMARQUEURS POUR UNE MÉDECINE PLUS PERSONNALISÉE .....	196
5.4	CONCLUSION GÉNÉRALE .....	198
5.5	RÉFÉRENCES.....	199
<b>ANNEXE I</b> .....		<b>i</b>
<b>AI</b>	<b>Analyse transcriptomique et génomique du cancer : outils bioinformatiques et enjeux actuels</b> .....	<b>ii</b>
AI.1	PRÉFACE .....	II
AI.2	CONTRIBUTIONS .....	II
AI.3	RÉFÉRENCE DE LA PUBLICATION.....	II
AI.4	ARTICLE.....	III
AI.4.1	<i>Abstract</i> .....	<i>iv</i>
AI.4.2	<i>Introduction</i> .....	<i>v</i>
AI.4.3	<i>RNA-seq in Cancer Genomics</i> .....	<i>vii</i>
AI.4.4	<i>Currently Available Bioinformatic Platforms for RNA-seq</i> .....	<i>ix</i>
AI.4.5	<i>Challenges / Current Gaps</i> .....	<i>xx</i>
AI.4.5.1	<i>Data Visualization / Exploration</i> .....	<i>xx</i>
AI.4.5.2	<i>Data Sharing and Standards</i> .....	<i>xxi</i>
AI.4.5.3	<i>Privacy Issues and Access to Patient Clinical Data</i> .....	<i>xxii</i>
AI.4.5.4	<i>Potential Solutions</i> .....	<i>xxii</i>
AI.4.6	<i>Conclusion</i> .....	<i>xxv</i>
AI.4.7	<i>Additional information</i> .....	<i>xxv</i>
AI.4.8	<i>References</i> .....	<i>xxvi</i>
<b>ANNEXE II</b> .....		<b>xxxiv</b>
<b>AII</b>	<b>Curriculum vitae</b> .....	<b>xxxv</b>
AII.1	PUBLICATIONS .....	XXXV
AII.2	PARTICIPATIONS À DES CONGRÈS SCIENTIFIQUES.....	XXXV
AII.2.1	<i>Présentations orales</i> .....	<i>xxxv</i>
AII.2.2	<i>Présentations par affiche</i> .....	<i>xxxvi</i>

## Liste des tableaux

Tableau 1.1 – Classification FAB des leucémies myéloïdes aiguës. ....	14
Tableau 1.2 – Classification de l’OMS des leucémies myéloïdes aiguës. ....	15
Tableau 1.3 – Classification European LeukemiaNet (ENL) des leucémies myéloïdes aiguës. ....	16
Tableau 1.4 – Résumé de la nomenclature des membres de la famille de protéines KMT2/MLL.....	22
Tableau 1.5 – Classification des différents partenaires de fusion retrouvés dans les leucémies médiées par le gène <i>KMT2A</i> . ....	29
Table 2.1 – Gene expression changes (FPKM) in chromatin modifying enzymes. ....	64
Table 2.2 – Genes with correlated gene expression and active histone modifications. ....	87
Table 2.3 – Summary of pediatric patients in this study. ....	96
Table 2.4 – List of primers used for shRNA-mediated knockdown of <i>ADCY9</i> expression. .....	97
Table 2.5 – List of primers used to prepare assay for transposase-accessible chromatin (ATAC)-sequencing libraries.....	99
Table 4.1 – Mouse models of leukemia. ....	164
Table 4.2 – Human models of leukemia.....	166
Table 4.3 – Screens performed using mouse leukemia models. ....	171
Table 4.4 – Screens performed using human leukemia models.....	172
Table AI.1 – Standalone analysis tools for RNA-seq data. ....	xi
Table AI.2 – Tool suites for RNA-seq analysis.....	xv
Table AI.3 – Data repositories with integrated analysis tools. ....	xix

## Liste des figures

Figure 1.1 – Processus de l’hématopoïèse.....	7
Figure 1.2 – Résumé des huit catégories de gènes fréquemment mutés dans la leucémie myéloïde aiguë.....	19
Figure 1.3 – Schéma récapitulatif de la structure de la protéine KMT2A.....	25
Figure 1.4 – Schéma résumant la formation du complexe multiprotéique de KMT2A et la reconnaissance des cibles de KMT2A.....	27
Figure 1.5 – Schéma simplifié de la structure d’une protéine chimérique issue d’une translocation chromosomique de type <i>KMT2A</i> .....	28
Figure 1.6 – Classification des patients basée sur la classe d’âge et le type de leucémies.	31
Figure 1.7 – Programme d’expression génique des cellules souches hématopoïétiques (HSC) dans les cellules hématopoïétiques non-leucémiques (considérées comme normales) et les cellules leucémiques porteuses d’une translocation de type <i>KMT2A</i> .....	32
Figure 1.8 – Mécanismes moléculaires menant à l’activité oncogénique des protéines de fusion KMT2A-MLLT3.....	35
Figure 1.9 – Mécanismes moléculaires menant à l’activité oncogénique des protéines de fusion KMT2A-MLLT4.....	36
Figure 1.10 – Représentation du modèle leucémique humain de type <i>KMT2A-MLLT3</i> ...37	
Figure 2.1 – Gene expression changes in the <i>KMT2A-MLLT3</i> leukemia model.....	61
Figure 2.2 – Gene set enrichment analysis of <i>KMT2A-MLLT3</i> expressing CD34+ cells...62	
Figure 2.3 – Expression of various transcription factor families in model leukemia stages.....	63
Figure 2.4 – Intra- and inter-individual reproducibility of methyl-sequencing replicates.....	66
Figure 2.5 – Descriptive statistics of methylation data.....	67
Figure 2.6 – DNA methylation changes in model leukemias.....	70
Figure 2.7 – Distribution in functional genomic elements of DMCs associated to the development of KM3 translocated model and patient AMLs.....	71
Figure 2.8 – Distribution in and around CpG islands of DMCs associated to the development of KM3 translocated model and patient AMLs.....	72

Figure 2.9 – DNA methylation in the transcription factor binding sites. ....	73
Figure 2.10 – Epigenetic analysis changes in <i>KMT2A-MLLT3</i> acute myeloid leukemia...76	76
Figure 2.11 – Genome-wide distribution of histone modifications in model leukemias. ..78	78
Figure 2.12 – Expression of the <i>CCL23</i> gene in leukemic and normal blood cells. ....79	79
Figure 2.13 – Overlapping peaks in ATAC-sequencing data from normal and leukemic cells.....89	89
Figure 2.14 – Predicted transcription factor binding sites in ATAC-sequencing peaks...90	90
Figure 2.15 – Single-cell expression of different genes in normal blood cells. ....94	94
Figure 3.1 – Structure de la protéine KMT2A-MLLT4. ....112	112
Figure 3.2 – Mécanismes moléculaires menant à l’activité oncogénique des protéines de fusion KMT2A-MLLT4. ....114	114
Figure 3.3 – Représentation du modèle leucémique humain de type <i>KMT2A-MLLT4</i> . ..115	115
Figure 3.4 – Classification hiérarchique d’échantillons de modèles LMA et de patients LMA basée sur leur signature transcriptomique respective. ....121	121
Figure 3.5 – Nuage de points ( <i>scatterplot</i> ) des valeurs d’expression génique entre deux réplicats biologiques du modèle <i>KM4</i> (en haut) et deux échantillons pédiatriques leucémiques <i>KM4</i> (en bas).....123	123
Figure 3.6 – Analyse transcriptomique des modèles <i>KM4</i> et <i>KM3</i> et caractérisation des processus biologiques impactés.....127	127
Figure 3.7 – Résumé des motifs de sites de fixation des facteurs de transcription retrouvés au niveau des promoteurs (-300 ou -600 pb en amont du TSS) des gènes plus (A) ou moins (C) exprimés dans le modèle <i>KM4</i> comparativement au modèle <i>KM3</i> . ....131	131
Figure 3.8 – Volcano plot représentant les différences d’expression des gènes entre les deux modèles leucémiques. ....133	133
Figure 3.9 – <i>LPL</i> est un potentiel biomarqueur probablement régulé de manière épigénétique et spécifique de la LMA médiée par la fusion <i>KMT2A-MLLT4</i> . ....137	137
Figure 4.1 – Hierarchical clustering of hematopoietic cell lines. ....159	159
Figure 5.1 – Expression des gènes <i>P2RY1</i> et <i>SHARPI</i> dans la cohorte de patients pédiatriques atteints d’une LMA (données TARGET).....194	194

## Liste des abréviations

5mC	5-méthylCytosine
ABL1	<i>Abelson tyrosine-protein kinase 1</i>
ADCY9	<i>Adenylate Cyclase 9</i>
AEP	<i>AF4/ENL/pTEFb complex</i>
AF10/MLLT10	<i>ALL1 (MLL/KMT2A) fused gene from chromosome 10 protein</i>
AF4/MLLT2	<i>ALL1 (MLL/KMT2A) fused gene from chromosome 4 protein</i>
AF6/MLLT4/AFDN	<i>ALL1 (MLL/KMT2A) fused gene from chromosome 6 protein</i>
AF9/MLLT3	<i>ALL1 (MLL/KMT2A) fused gene from chromosome 9 protein</i>
AGM	Aorte-Gonade-Mésonéphros
AHD	<i>ANC1 Homology Domain</i>
AKT	Protéine kinase B
ALL / LLA	<i>Acute Lymphocytic Leukemia / Leucémie Lymphoïde Aiguë</i>
AMKL	<i>Acute Megakaryoblastic Leukemia</i>
AML / LMA	<i>Acute Myeloid Leukemia / Leucémie Myéloïde Aiguë</i>
AML1-ETO	Également sous le nom <i>RUNX1-RUNX1T1</i>
APC	<i>Antigen Presenting Cell</i>
APOC2	<i>ApolipoProtein CII</i>
ARID3B	<i>AT-Rich Interactive Domain-Containing Protein 3B</i>
ASH2L	<i>ASH2 Like Histone Lysine Methyltransferase Complex Subunit</i>
ASXL1	<i>ASXL Transcriptional Regulator 1</i>
ATAC-seq	<i>Assay for Transposase-Accessible Chromatin using sequencing</i>
B4galt1	<i>Beta-1,4-Galactosyltransferase 1</i>
BCL2	<i>BCL2 Apoptosis Regulator</i>
BCLQ	Banque de Cellules Leucémiques du Québec
BCR	<i>BCR Activator Of RhoGEF And GTPase</i>
BM	<i>Bone Marrow</i>
BMI1	<i>BMI1 Proto-Oncogene, Polycomb Ring Finger</i>
BNIPL1	<i>BCL2 Interacting Protein Like</i>
BRD	Bromodomaine
BSA	<i>Bovine Serum Albumin</i>
BWA	<i>Burrows-Wheeler Aligner</i>
cAMP	<i>Cyclic Adenosine MonoPhosphate</i>
CAR-T	<i>Chimeric Antigen Receptor T</i>
Cas9	<i>CRISPR-associated protein 9</i>
CBFA2T3	<i>CBFA2/RUNX1 Partner Transcriptional Co-Repressor 3</i>

CBFB	<i>Core-Binding Factor Subunit Beta</i>
CBP	<i>CREB-Binding Protein</i>
CCL	<i>C-C Motif Chemokine Ligand</i>
CCLE	<i>Cancer Cell Line Encyclopedia</i>
CCR1	<i>C-C Chemokine Receptor type 1</i>
CD	<i>Cluster of Differentiation</i>
CDK9	<i>Cyclin-Dependent Kinase 9</i>
cDNA	<i>complementary DNA</i>
CEBP- $\alpha$	<i>CCAAT Enhancer Binding Protein Alpha</i>
CEBP- $\beta$	<i>CCAAT Enhancer Binding Protein Beta</i>
CEBP- $\epsilon$	<i>CCAAT Enhancer Binding Protein Epsilon</i>
CGWB	<i>Cancer Genome WorkBench</i>
CHD4	<i>Chromodomain Helicase DNA Binding Protein 4</i>
ChIP-seq	<i>Chromatin ImmunoPrecipitation using sequencing</i>
CLP	<i>Progéniteurs Lymphoïdes Communs</i>
CML	<i>Chronic Myelogenous Leukemia</i>
CMP	<i>Progéniteurs Communs de cellules Myéloïdes</i>
COSMIC	<i>Catalogue of Somatic Mutations in Cancer database</i>
CpGi	<i>CpG Island</i>
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
CRLF2	<i>Cytokine Receptor Like Factor 2</i>
CSF	<i>Colony-Stimulating Factor</i>
CSH / HSC	<i>Cellules Souches Hématopoïétiques / Hematopoietic Stem Cells</i>
CSL	<i>Cellules Souches Leucémiques</i>
c-Src	<i>SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase</i>
CTD	<i>C-Terminal Domain</i>
CyP33	<i>CycloPhiline 33</i>
dbSNP	<i>Database for Single Nucleotide Polymorphisms</i>
DCPS	<i>Decapping Enzyme, Scavenger</i>
DEG	<i>Differentially Expressed Genes</i>
DEK	<i>DEK Proto-Oncogene</i>
DIL	<i>Dilute Domain</i>
DMC	<i>Differentially Methylated Cytosines</i>
DMSO	<i>DiMéthylSulfOxide</i>
DNMT1	<i>ADN méthyltransférase 1</i>
DNMT3A	<i>ADN méthyltransférase 3A</i>
DNMT3B	<i>ADN méthyltransférase 3B</i>
DOT1L	<i>DOT1 Like Histone Lysine Methyltransferase</i>
DOTCOM	<i>DOT1-Containing Complex</i>



DPY30	<i>Dpy-30 Histone Methyltransferase Complex Regulatory Subunit</i>
E(z)	<i>Enhancer of Zeste</i>
E2F	<i>E2F Transcription Factor</i>
ECS(ASB)	<i>Elongin B/C-Cullin-SOCS box protein</i>
EDTA	<i>EthyleneDiamineTetraAcetic Acid</i>
EEN	<i>Extra Eleven-Nineteen Leukemia Fusion Gene Protein</i>
ELL	<i>Elongation Factor For RNA Polymerase II</i>
ELN	<i>European LeukemiaNet</i>
ENCODE	<i>Encyclopedia of DNA Elements</i>
ENL/MLLT1	<i>MLLT1 Super Elongation Complex Subunit</i>
EPH	<i>Erythropoietin-Producing Hepatoma Receptor</i>
EPO	<i>Erythropoïétine</i>
EPOR	<i>Erythropoietin Receptor</i>
eQTL	<i>Expression Quantitative Loci</i>
ES	<i>Embryonic Stem</i>
ETV6	<i>ETS Variant Transcription Factor 6</i>
FAB	<i>French-American-British</i>
FHA	<i>ForkHead Associated domain</i>
FISH	<i>Fluorescence In Situ Hybridization</i>
FLC	<i>Fetal Liver Cell</i>
FLT3	<i>FMS-Like Tyrosine Kinase 3</i>
FOS	<i>Fos Proto-Oncogene, AP-1 Transcription Factor Subunit</i>
FPKM	<i>Fragments Per Kilobase of transcript per Million mapped reads</i>
FTHY	<i>Transplanted Fetal THYmic Tissue</i>
GAS7	<i>Growth Arrest Specific 7</i>
GATA-1	<i>GATA Binding Protein 1</i>
GATA-2	<i>GATA Binding Protein 2</i>
GDC	<i>Genomic Data Commons</i>
GDSC	<i>Genomics of Drug Sensitivity in Cancer</i>
GEO	<i>Gene Expression Omnibus</i>
GFI1B	<i>Growth Factor Independent 1B Transcriptional Repressor</i>
GFP	<i>Green Fluorescent Protein</i>
GLIS2	<i>GLIS Family Zinc Finger 2</i>
GM-CSF	<i>Granulocyte-Macrophage Colony-Stimulating Factor</i>
GMP	<i>Progéniteurs Granulocytes/Macrophages</i>
GO	<i>Gene Ontology</i>
GPCR	<i>G Protein-Coupled Receptor</i>

GSEA	<i>Gene Set Enrichment Analysis</i>
GSK-3 $\alpha$	<i>Glycogen Synthase Kinase 3 alpha</i>
GTEX	<i>Gene-Tissue EXpression</i>
GTP	<i>Guanosine TriPhosphate</i>
H2A	<i>Histone H2A</i>
H2B	<i>Histone H2B</i>
H3	<i>Histone H3</i>
H3K27ac	<i>Histone H3 Lysine K27 avec un groupement acétyle</i>
H3K36me2/3	<i>Histone H3 Lysine K36 avec un groupement di/triméthyle</i>
H3K4me1-3	<i>Histone H3 Lysine K4 avec un groupement mono/di/triméthyle</i>
H3K79me1-3	<i>Histone H3 Lysine K79 avec un groupement mono/di/triméthyle</i>
H3K9ac	<i>Histone H3 Lysine K9 avec un groupement acétyle</i>
H4	<i>Histone H4</i>
H4K16ac	<i>Histone H4 Lysine K16 avec un groupement acétyle</i>
H4R3	<i>Histone H4 Arginine R3</i>
hg19	<i>Human Genome version 19</i>
HMG20B	<i>High Mobility Group 20B</i>
Hox	<i>Homeobox</i>
HSS	<i>Hypersensitive Sites</i>
IC50	<i>Half-maximal Inhibitory Concentration</i>
ICGC	<i>International Cancer Genome Consortium</i>
IDH1/2	<i>Isocitrate Dehydrogenase (NADP(+)) 1/2</i>
IL	<i>Interleukine</i>
IRF8	<i>Interferon Regulatory Factor 8</i>
ITD	<i>Internal Tandem Duplication</i>
Itgb3	<i>Integrin Subunit Beta 3</i>
JAK	<i>Tyrosine-Protein Kinase JAK</i>
Jmjd1c	<i>Jumonji Domain Containing 1C</i>
JUN	<i>Jun Proto-Oncogene, AP-1 Transcription Factor Subunit</i>
KCl	<i>Potassium chloride</i>
KDM	<i>Lysine Demethylase</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
KIT	<i>KIT Proto-Oncogene, Receptor Tyrosine Kinase</i>
KLF1	<i>Kruppel Like Factor 1</i>
KM3	<i>KMT2A-MLLT3</i>
KM4	<i>KMT2A-MLLT4</i>

KMT2	<i>Lysine méthyltransférase 2</i>
KO	<i>Knock-Out</i>
KRAS	<i>KRAS Proto-Oncogene, GTPase</i>
LEDGF	<i>Lens Epithelium-Derived Growth factor</i>
LiCl	<i>Lithium Chloride</i>
LMO-2	<i>LIM Domain Only 2</i>
LMPP	<i>Lymphoid-primed MultiPotent Progenitors</i>
LPL	<i>Lipoprotéine Lipase</i>
MACS2	<i>Model-based Analysis of ChIP-Seq version 2</i>
MAF	<i>MAF BZIP Transcription Factor</i>
MECOM	<i>MDS1 and EVI1 Complex Locus</i>
MEIS1	<i>Meis Homeobox 1</i>
MEN1	<i>Menin 1</i>
MEP	<i>Progéniteurs Mégacaryocytes/Érythroïdes</i>
Methyl-seq	<i>Methylation Sequencing</i>
MKL1	<i>Megakaryoblastic Leukemia 1</i>
MLL/KMT2A/ALL-1	<i>Mixed-Lineage Leukemia</i>
MN1	<i>MN1 Proto-Oncogene, Transcriptional Regulator</i>
MOF	<i>Lysine Acetyltransferase 8</i>
MOZ	<i>Lysine Acetyltransferase 6A</i>
MPAL	<i>Mixed-Phenotype Acute Leukemia</i>
MPL	<i>MPL Proto-Oncogene, Thrombopoietin Receptor</i>
MPP	<i>Progéniteurs Primitifs Multipotents</i>
MYH11	<i>Myosin Heavy Chain 11</i>
MZF1	<i>Myeloid Zinc Finger 1</i>
NaCl	<i>Sodium chloride</i>
NCI	<i>National Cancer Institute</i>
NCOR	<i>Nuclear Receptor Co-Repressor</i>
NF-KB	<i>Nuclear Factor Kappa B</i>
NF-Y	<i>Nuclear Factor Y</i>
NGS	<i>Next-Generation Sequencing</i>
NHEJ	<i>Non-homologous end joining</i>
NHGRI	<i>National Human Genome Research Institute</i>
NK	<i>Natural Killer</i>
NOD-SCID	<i>Nonobese Diabetic/Severe Combined Immunodeficiency</i>
NPM1	<i>Nucleophosmin 1</i>
NRAS	<i>NRAS Proto-Oncogene, GTPase</i>

NUP214	<i>Nucleoporin 214</i>
NUP98	<i>Nucleoporin 98</i>
OCT 4	<i>Octamer-binding transcription factor 4</i>
OMS	<i>Organisation Mondiale de la Santé</i>
P2RY1/2	<i>Purinergic Receptor P2Y 1/2</i>
PAFc	<i>Polymerase-Associated Factor complex</i>
PATZ1	<i>POZ/BTB And AT Hook Containing Zinc Finger 1</i>
PAX7	<i>Paired Box 7</i>
PBMC	<i>Peripheral Blood Mononuclear Cell</i>
PBS	<i>Phosphate Buffered Saline</i>
PBX1	<i>PBX Homeobox 1</i>
PcG	<i>Polycomb Group</i>
PCPG	<i>Pediatric Cancer Genome Project</i>
PCR	<i>Polymerase Chain Reaction</i>
PDX	<i>Patient-Derived Xenograft</i>
PHD	<i>Plant HomeoDomain</i>
Phf6	<i>PHD Finger Protein 6</i>
PI3K	<i>Phosphatidylinositol 3-Kinase</i>
PMID	<i>PubMed unique Identifier</i>
PML	<i>PML Nuclear Body Scaffold</i>
PRC1	<i>Polycomb Repressive Complex 1</i>
PREX1	<i>Phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 1</i>
PROM1	<i>Prominin 1</i>
PRR	<i>Poliovirus Receptor-Related protein</i>
PTD	<i>Partial Tandem Duplication</i>
P-TEFb	<i>Positive Transcription Elongation Factor b</i>
RA	<i>RAS-Association</i>
RAD21	<i>RAD21 Cohesin Complex Component</i>
RARA	<i>Retinoic Acid Receptor Alpha</i>
RbBP5	<i>Retinoblastoma-Binding Protein 5</i>
RBM15	<i>RNA Binding Motif Protein 15</i>
RBPJ	<i>Recombination Signal Binding Protein For Immunoglobulin Kappa J Region</i>
RefSeq	<i>The Reference Sequence database</i>
RET	<i>Ret Proto-Oncogene</i>
RNA / ARN	<i>RiboNucleic Acid / Acide RiboNucléique</i>

RNAi	<i>RNA interference</i>
RNA-seq	<i>RNA sequencing</i>
ROCK1	<i>Rho Associated Coiled-Coil Containing Protein Kinase 1</i>
RPKM	<i>Reads Per Kilobase of transcript per Million mapped reads</i>
RPMI	<i>Roswell Park Memorial Institute medium</i>
RPS13	<i>Ribosomal Protein S13</i>
RREB1	<i>Ras Responsive Element Binding Protein 1</i>
rRNA	<i>Ribosomal RNA</i>
RUNX1	<i>RUNX Family Transcription Factor 1</i>
RUNX1T1	<i>RUNX1 Partner Transcriptional Co-Repressor 1</i>
S100A8/A9	<i>S100 Calcium Binding Protein A8/9</i>
SAM	<i>S-AdénosylMéthionine</i>
SCF	<i>Stem Cell Factor</i>
scRNA-seq	<i>Single-cell RNA sequencing</i>
SEPT	<i>Septin</i>
SET	<i>Su(var), Enhancer of zeste, Trithorax</i>
SETD1A/B	<i>SET Domain Containing 1A/B</i>
SHARP1	<i>Basic Helix-Loop-Helix Family Member E41 (BHLHE41)</i>
shRNA	<i>Short Hairpin RNA</i>
siRNA	<i>Small Interfering RNA</i>
SIRT1	<i>Sirtuin 1</i>
SLC17A9	<i>Solute Carrier Family 17 Member 9</i>
SNX27	<i>Sorting Nexin Family Member 27</i>
SPI1/PU.1	<i>Spi-1 Proto-Oncogene</i>
STAG2	<i>Stromal Antigen 2</i>
STAT	<i>Signal Transducer And Activator Of Transcription</i>
STIL	<i>STIL Centriolar Assembly Protein</i>
STK3	<i>Serine/Threonine Kinase 3</i>
TAD	<i>Trans-Activating Domain</i>
TAL1/SCL	<i>TAL BHLH Transcription Factor 1, Erythroid Differentiation</i>
TALEN	<i>Transcription Activator-Like Effector Nucleases</i>
TARGET	<i>Therapeutically Applicable Research to Generate Effective Treatments</i>
TASP1	<i>Taspase 1</i>
TCGA	<i>The Cancer Genome Atlas</i>
TDZ	<i>Thioridazine</i>
TET2	<i>Tet Methylcytosine Dioxygenase 2</i>
TFBS	<i>Transcription Factor Binding Site</i>

TLR4	<i>Toll Like Receptor 4</i>
TP53 / p53	<i>Tumor Protein 53</i>
TPO	<i>Thrombopoïétine</i>
trxG	<i>Trithorax-group</i>
TSS	<i>Transcription Start Site</i>
UCSC	<i>University of California Santa Cruz</i>
WDHD1	<i>WD Repeat And HMG-Box DNA Binding Protein 1</i>
WDR5	<i>WD Repeat Domain 5</i>
WEE1	<i>WEE1 G2 Checkpoint Kinase</i>
WT1	<i>Wilms Tumor 1</i>
ZEB2	<i>Zinc Finger E-Box Binding Homeobox 2</i>

*À ma grand-mère.*

## Remerciements

Mon aventure au doctorat touche à sa fin et aura été incroyablement riche en rencontres. J'ai eu la chance de découvrir le monde de la recherche académique et de me spécialiser dans le domaine de bio-informatique et du séquençage. En plus d'avoir nourri ma curiosité scientifique, j'ai surtout grandi et mûri. Quitter la France, les ami.e.s, la famille et venir m'installer à Montréal aura été plus qu'un tournant dans ma Vie.

Merci **Brian** pour ta confiance et m'avoir permis de rejoindre ton équipe de recherche. Je me rappellerai toujours de notre première rencontre « en vrai » lors du concours de recrutement étudiant à l'IRIC, une après-midi de juin 2015. Une bière à la main, devant un poster. Merci de m'avoir fait confiance, d'avoir cru en moi, de m'avoir transmis ta passion pour la science et la recherche, de m'avoir laissé tant de liberté pour gérer mes horaires et réaliser mes projets, d'avoir fait preuve de patience et d'un soutien sans faille à chacune des étapes de mon parcours. Merci pour les longues discussions dans ton bureau à parler de science, de génomique mais aussi du quotidien en dehors du labo. Tu auras été un boss incroyablement présent, bienveillant et je me sens privilégié d'avoir pu être ton premier étudiant au doctorat.

Merci évidemment à tous mes collègues de la « Wilhelm Family ». **Karine**, merci de m'avoir encadré au touuuuut début de mon projet. Tu m'auras appris la plupart des techniques expérimentales de mon projet et ta rigueur scientifique aura pavé la voie pour le restant de mon doctorat. Merci pour les larmes de rire, les discussions scientifiques (et moins scientifiques !), ta folie contagieuse, les soupers sushis et les bons moments à tes côtés (« OK, BYE ! »). **Élodie**, merci pour tes conseils d'ancienne étudiante au doctorat (you know what it is !), ta rigueur, ton temps passé à relire ma thèse et mes papiers ou à m'aider dans mes manip' de dernière minute, pour les bons moments passés au labo, que ce soit sous la hotte ou pendant nos pauses déjeuners, le concert inoubliable de Muse au Centre Bell (!!!) et surtout ton soutien constant pour m'aiguiller dans le parcours du doctorat rempli d'embûches. **Shahem**, mon voisin de bureau, merci pour ta bonne humeur, nos fous rires de bon matin, nos discussions à refaire le monde et à parler de tout et de rien, travailler notre *français*, et surtout... pour l'accumulation des *post-it* au-dessus de nos bureaux sans réelle signification mais toujours avec plein d'autodérision ! Je me souviendrai toujours de notre congrès à Ottawa : trois journées scientifiques à écouter des



conférences, à (beaucoup trop) manger à chaque pause et à prendre l'apéro dans la piscine de l'hôtel. Quel souvenir... avec évidemment **Safia**, ma voisine bio-informaticienne : merci pour les longues discussions dans le bureau à parler de la vie d'expat' ou de programmation (un peu !) et pour tes bons mots qui me permettaient de relativiser lors de mes périodes de doute. **Valérie**, ma nouvelle voisine de labo lorsque la pandémie est arrivée, merci d'avoir été une oreille attentive durant la deuxième partie de mon doctorat. J'ai adoré pouvoir échanger sur nos vies respectives tous les matins, te partager mes petites victoires mais aussi mes doutes. Merci pour tes conseils toujours affûtés, ton temps passé à relire minutieusement ma thèse et ta bonne humeur. Enfin, merci à **Chithra, Vikie, Blandine, Karla, Benjamin, Sarah, Chloé, Jessica, Hera, Magalie** et tous les membres du labo qui auront croisé mon chemin. Ce fut un réel plaisir de travailler à vos côtés ! Merci aux membres du laboratoire de Vincent, mes voisins de labo et de bureau pour nos échanges et surtout pour les nombreux gâteaux que nous avons dévorés pour célébrer n'importe quelle occasion. Un merci tout particulier à **Karine, Damien et Laia**. Merci aussi à **Jennifer** et **Raphaëlle** d'avoir pris soin de mes centaines de *runs* de séquençage !

L'institut est particulièrement chanceux d'avoir une équipe de feu qui prend soin de sa communauté étudiante, du début jusqu'à la fin du parcours : celle des **affaires académiques**. Votre présence est essentielle ! Les nombreux cafés pris ensemble me manqueront ! Quel plaisir de passer vous voir pour jaser et célébrer avec vous les petites victoires ! **Julie**, merci d'avoir été là quand ça allait moins bien. Tu as joué un rôle important pour me permettre de garder le cap. Merci pour tes conseils, ton oreille attentive, ta bonne humeur et tes nombreuses lettres pour appuyer mes candidatures à des bourses d'implication sociale. **Évelyne**, une rencontre qui aura marqué mon parcours et mon quotidien à l'IRIC. Merci pour les BBQ (parfois enflammés !), ta joie de vivre communicative, l'organisation des événements étudiants jusqu'à pas d'heure, nos fous rires et surtout nos nuits dans des bars peu fréquentables ! On remet ça quand tu veux ! **Pascale**, merci pour les nombreux cafés, ton calme olympien face aux obstacles administratifs (comme l'immigration par exemple !!!) et tes bons mots. T'entendre dire « bonne journéeéééééééééé ! » me manquera ! Un merci particulier à **Valérie** pour ta présence au fil de mon parcours que ce soit à l'IRIC, pendant des événements de recrutement, des partys de Noël (...) ou lors de séances de magasinage dans les rues de Verdun. Un vrai plaisir de te croiser dans les couloirs du 3<sup>ème</sup> ou du sous-sol ! **Patrick**, un *shout-out* particulier pour toi que j'avais rencontré en

octobre 2014 à Paris lors d'un forum international et sans qui je ne serai pas là où je suis aujourd'hui. Merci d'avoir bouleversé ma vie, d'avoir cru en mon dossier de l'époque et de m'avoir permis de rejoindre l'institut.

Un immense merci à mes collègues de **l'association étudiante de l'IRIC** au sein de laquelle je me suis impliqué et épanoui pendant 3 années consécutives dont 2 à la présidence. Merci de m'avoir fait confiance dans nos projets les plus fous ! Je garde que de bons souvenirs de tous les événements que nous avons organisés ensemble. Des projets qui me tenaient à cœur, des projets nouveaux pour tenter de faire bouger les lignes. Merci pour les fous rires, les longues soirées à discuter des projets, les meetings d'asso, les 5 à 7, les partys d'halloween et de Noël, les BBQ, le temps passé dans la salle des archives à ranger les bières et j'en passe ! Quel plaisir d'avoir bossé avec vous autres ! Merci à mes ami.e.s et collègues de l'IRIC pour les nombreuses bières partagées lors des 5 à 7 du vendredi, les bons moments et le soutien moral. Un merci tout particulier à **Nandita** (thanks for our weekly *tasty tongues meeting*, our dances and being my friend after all this time !), **Mélanie** (que de bons souvenirs pendant les événements de recrutement... on retourne à la Rockette quand tu veux !!), **Audrey** (la seule et unique Brigitte !!!), **Virginie & Jean-David** (merci pour nos longues discussions sur la santé mentale et la situation de la communauté étudiante et pour votre confiance dans l'organisation de la semaine de sensibilisation à la santé mentale !), **Rébecca** (la *dramaqueen* !), **Charles** (le *streamer* né !), **Lynda** (la gardienne du fort ; merci pour les petits moments à jaser à la réception, les soupers et les larmes de rire !), **Lucas** (la diva !), **Caroline** (#24h!) mais aussi **Jack, Swati, Marjorie, Jordan, Ema, Justine** et toutes les personnes que j'ai rencontrées pendant mon parcours.

Une mention spéciale à mon amie, collègue, passionnée, *partner-in-crime* **Myreille** qui me suit dans mes projets les plus dingues depuis le début. Merci d'embarquer à chaque fois. *Dans Les Coulisses de la recherche* restera un des plus beaux souvenirs de mon parcours ! C'était si chouette de l'organiser, de l'imaginer, de le construire et de l'améliorer d'année en année. Quelle aventure de fou : le stress de l'organisation, le contact avec le grand public, côtoyer Yanick Villedieu, faire rayonner la recherche, passer dans les médias, aller jusqu'en finale du concours Forces Avenir... On en ressort grandi avec une passion dévorante pour la communication scientifique qui nous anime et nous pousse à aller plus loin. Merci d'être qui tu

es, de rire à mes blagues et de ne jamais juger mes rires en « ultrason » ! J'ai hâte de voir la suite de l'aventure entrepreneuriale dans laquelle on s'embarque ! #DLC

Pendant mon doctorat, j'ai eu la chance de côtoyer l'incroyable équipe des communications de l'IRIC. **Catherine, Cristina, Noémie, Léa, Sandy**, merci de m'avoir pris sous votre aile et de m'avoir fait confiance ! Quel plaisir de bosser avec vous ! J'ai énormément appris à vos côtés et j'espère avoir la chance de rebosser avec vous dans le futur. Merci pour tout et vivement nos prochains 5 à 7 !

Merci infiniment à toutes les **agences de subvention**, notamment la Fondation Cole qui me soutient depuis le début de mon doctorat. Merci à l'IRIC et au programme de biologie moléculaire pour leur coup de pouce financier. Merci aux Instituts de Recherche en Santé du Canada de m'avoir permis de partir en Australie, à l'autre bout du monde, pour présenter mes travaux de recherche. Un voyage qui aura marqué à jamais mon parcours et pour lequel je me sens privilégié ! Merci aux Études Supérieures et Postdoctorales de l'Université de Montréal et aux donateurs pour leur soutien financier et leur confiance, au cours et à la fin de mon doctorat. Merci à la Fédération des Associations Étudiantes du Campus de l'UdeM, la Faculté de Médecine, les Fonds d'Amélioration à la Vie Étudiante, Forces Avenir ainsi que le Vice-rectorat aux affaires étudiantes et aux études pour les récompenses et les prix en lien avec mon engagement social au sein de la communauté étudiante. Jongler entre la vie au laboratoire et la vie associative aura été tout un défi organisationnel pour lequel toutes ces reconnaissances m'ont continuellement encouragé à poursuivre mes efforts pour améliorer la vie étudiante.

Merci à tous les copains en dehors de l'IRIC mais surtout en dehors du domaine scientifique qui m'ont permis de m'aérer l'esprit, de penser et de parler d'autre chose que de science ! Bref, de décompresser ! Merci d'avoir été si présents pour moi ! Les *chalets lovers* (**Camille, Léandre, Marguerite, Mélissa, Jeff, Fatou, Alexia, Marie-Ève, Mathieu, Naomi, Hugo, Jennifer, Christine, Chris**), **Alex** et **Lana** (pour nos fondus en plein été !!), **Gab** et **Kéké** (pour nos soirées Villainous à n'en plus finir !!), **Catherine** (alias LA Catoche), la *gang de muséo* (**Violette, Jo, Olive, Yoyo, Justine, Paul, Val, Mathieu** ; merci pour les chalets passés et à venir, nos soirées nachos devant les Oscars, nos soirées cocktail & jeux de société). **Sean**, merci pour nos aprem' jeux & cafés et nos jasettes tout au long de notre jumelage à GFSGS. Merci aux copains de France avec qui, après toutes ces années et malgré les distances, les fous rires sont

toujours au rendez-vous : **Fanny, Timothé, Clément, Mathieu, Nazim, Patrick, Pauline**, sans oublier **Mathilde** (la Valooche !). J'ai hâte de vous revoir ou que vous passiez me voir par ici!

Évidemment que je me dois de partager la chance que j'ai d'avoir une belle-famille si bienveillante. Merci d'avoir accueilli le p'tit français ! Merci **Lise, Marc, Marie-Hélène, Robert, Arthur** et **Rebecca** pour nos soupers du dimanche qui rendent mes lundis moins difficiles. Merci d'être ma deuxième famille, de me permettre de m'enraciner progressivement au Québec, de me soutenir dans mes projets et de sortir des bulles à chaque petite victoire ! **Line** et **Isabelle**, merci pour vos bonnes ondes et un clin d'œil particulier à **Katherine** et **Sonia** pour nos larmes de rire et nos karaokés les soirs d'été (voire même les réveillons de Noël #*AllIWantForChristmasIsYou*). Quel bonheur de danser et de décompresser avec vous autres ! Vous êtes des joies de vivre et avez joué un rôle important dans mon cheminement doctoral.

Je n'aurais pas réussi à passer à travers toutes ces années de doctorat sans le soutien sans faille de ma famille. Merci à ma maman **Odile**, mon papa **Régis**, ma sœur **Delphine**, mon frère **Mathieu**, mon beau-frère **Julien**, ma belle-sœur **Graziella**, mes nièces **Élodie & Juliana**, mon neveu & filleul **Vincent** (et la chienne de la famille **Pepette** !). Vous avez été incroyablement présents dans tous mes projets, avec toujours les bons mots pour me soutenir. Grâce à vous, j'ai gardé ma bonne humeur. Je savais que, malgré la distance, vous étiez là. Comme je dis souvent : « Loin des yeux, près du cœur ». Vous ne vous rendez probablement pas compte de l'immense impact que vous avez eu dans mon cheminement doctoral mais surtout personnel. Merci d'être si bienveillant et présent pour moi. J'ai de la chance de vous avoir dans ma Vie.

À la Queen, mon meilleur ami, mon fiancé, celui qui aura bouleversé ma vie, me fait relativiser jour après jour et rend mes journées plus belles, **Marc-Olivier** (et la **Milka C. Milan**). Merci d'avoir toujours trouvé les bons mots pour me remonter le moral quand ça allait moins bien. Merci pour tes conseils toujours pertinents, ton oreille attentive et ta bonne humeur. Merci de me suivre dans mes projets les plus fous. Merci de me supporter chaque jour et de rire à mes blagues plus ou moins drôles. Merci d'avoir été à l'écoute lors de mes périodes de profonde remise en question mais surtout d'être là dans les meilleurs moments. Merci pour ton soutien sans faille et d'avoir toujours fait en sorte de me motiver pour que je ne lâche à aucun moment. Tu es l'un des piliers de ma Vie et j'ai hâte de la poursuivre à tes côtés.

*Merci pour tout.*

# **CHAPITRE 1**

## **INTRODUCTION**

# 1 Introduction

## 1.1 Organisation générale des chapitres

Le chapitre **introduction** présente l'hématopoïèse dans un contexte physiologique et les différents aspects cliniques et moléculaires impliqués dans la leucémie myéloïde aiguë. Le chapitre passe en revue les mutations à l'origine du développement leucémique et met l'accent sur les réarrangements chromosomiques impliquant le gène *KMT2A*. Les connaissances actuelles sur les mécanismes biologiques et les propriétés oncogéniques des protéines de fusion *KMT2A-MLLT3* et *KMT2A-MLLT4* sont détaillées.

Le **chapitre 2** présente un article publié en 2020 dans le journal *Haematologica*. Le papier vise à décrire les événements initiaux génétiques et épigénétiques impliqués dans la leucémogénèse causée par la translocation chromosomique de type *KMT2A-MLLT3*. Grâce à l'utilisation d'un modèle de leucémie humaine, il a été possible de caractériser l'ensemble de ces changements durant le développement de la maladie. *KMT2A-MLLT3* étant un régulateur épigénétique connu, nous avons caractérisé les changements dans la méthylation de l'ADN, certaines modifications d'histones ainsi que l'accessibilité de la chromatine. Nous avons observé un profil global d'hypométhylation de l'ADN dès les premiers stades de la transformation leucémique, avec un changement d'expression de certains gènes. Les données d'ATAC-seq ont révélé très peu de changements spécifiques à la leucémie en termes d'accessibilité de la chromatine. L'intégration de toutes ces analyses à l'échelle du génome a permis d'identifier de potentiels biomarqueurs. Par exemple, nous avons démontré que le gène codant pour l'adénylate cyclase *ADCY9* est essentiel pour assurer la viabilité des cellules leucémiques. Nos analyses suggèrent également un potentiel rôle du récepteur couplé aux protéines *CCR1* et son ligand spécifique *CCL23*. Nous avons ainsi découvert que la fusion *KMT2A-MLLT3* induit des changements minimes dans l'épigénome et profite de la machinerie transcriptionnelle 'normale' pour mener la leucémogénèse.

Le **chapitre 3** présente des données préliminaires qui visent à caractériser les différences génétiques et épigénétiques entre deux leucémies présentant un écart important en termes de probabilité de survie en clinique. Elles sont induites par deux fusions de gènes différentes : *KMT2A-MLLT3* et *KMT2A-MLLT4*. En adaptant le modèle du chapitre 2 à la translocation

chromosomique *KMT2A-MLLT4* et en utilisant les mêmes techniques expérimentales, il a été possible d'observer les différences entre les deux modèles. L'analyse des signatures transcriptomiques amènent à penser que la régulation protéique serait perturbée dans les leucémies causées par la fusion *KMT2A-MLLT4*. *PROM1* et *LPL* ont pu être identifiés comme étant deux potentiels biomarqueurs pour lesquels des analyses fonctionnelles sont nécessaires pour valider leur fonction.

Le **chapitre 4** présente une revue de littérature publiée en 2019 dans le journal *Experimental Hematology* et sert d'ouverture pour entamer la discussion. Le modèle utilisé tout au long de cette thèse fait partie de l'une des multiples façons d'étudier la leucémie pédiatrique qui sont résumées par l'article à travers une description des avantages et inconvénients des modèles existants.

Le **chapitre 5** discute des avancées dans la recherche contre la leucémie présentées dans cette thèse. L'objectif est de décrire les forces et les faiblesses du modèle utilisé et de proposer des façons de l'améliorer pour se rapprocher davantage de la réalité en clinique. Il est question également de revenir sur les mécanismes biologiques des deux fusions de gènes et de comprendre comment les biomarqueurs identifiés dans les analyses pourraient servir de potentielles cibles thérapeutiques.

## 1.2 Le système sanguin

### 1.2.1 L'hématopoïèse : un processus assurant l'homéostasie sanguine

L'hématopoïèse se définit comme étant le processus physiologique par lequel les composants cellulaires du sang sont continuellement renouvelés et constitués tout au long de la vie d'un organisme. Le sang est l'un des tissus qui se régénère le plus vite puisqu'à chaque seconde au cours de la vie, des millions de cellules sanguines sont produites et remplacées. Un humain adulte génère environ  $4 \text{ à } 5 \times 10^{11}$  de cellules hématopoïétiques par jour (Kaushansky 2006). La production continue de l'ensemble de ces types cellulaires requiert un système hautement régulé mais très réactif. Selon les types de cellules sanguines, leur durée de vie varie entre quelques heures à plusieurs années. Par exemple, les globules rouges ont une durée de vie d'environ 4 mois alors que les plaquettes et les lymphocytes ont respectivement une durée de vie de 8-10 jours ou jusqu'à quelques années. Par ailleurs, en situation d'urgence comme en anémie ou lors d'infection, le nombre de cellules sanguines peut augmenter ou diminuer rapidement et se rétablir après la récupération. La régulation très fine de l'hématopoïèse permet de maintenir le nombre adéquat de cellules sanguines et ainsi d'assurer l'homéostasie sanguine (Rieger and Schroeder 2012).

Le système sanguin est composé de diverses populations de cellules hautement spécialisées avec des fonctions uniques. Les leucocytes (ou globules blancs) représentent l'ensemble des cellules spécialisées dans l'immunité acquise et innée. Ce groupe comprend les granulocytes (incluant les neutrophiles, basophiles, éosinophiles et mastocytes), les lymphocytes (cellules T, B et NK (*Natural Killer*)), les monocytes/macrophages et les cellules dendritiques. Les érythrocytes (ou globules rouges) s'assurent du transport de l'O<sub>2</sub> et du CO<sub>2</sub> à travers l'organisme tandis que les mégacaryocytes génèrent les plaquettes, indispensables pour la cicatrisation des plaies et la coagulation du sang. (Fey 2007, Jagannathan-Bogdan and Zon 2013).

Pendant l'embryogenèse, l'hématopoïèse se déroule successivement dans le sac vitellin extra-embryonnaire, l'allantoïde et le placenta, la région intra-embryonnaire AGM (Aorte-Gonade-Mésonephros) et le placenta, puis le foie fœtal et finalement la moelle osseuse. Cette dernière est le site principal de l'hématopoïèse après la naissance (Arber, Orazi et al. 2016).



## 1.2.2 À la base de la hiérarchie : les cellules souches hématopoïétiques

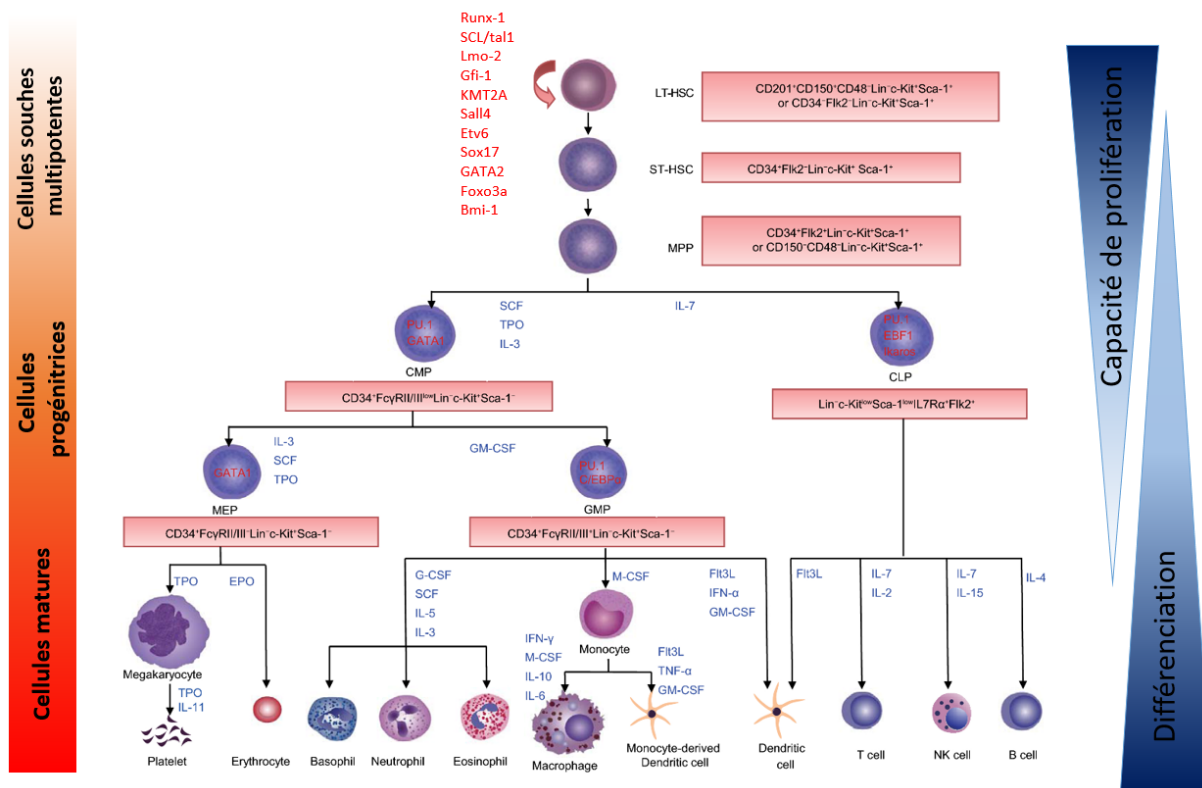
Tous les types cellulaires sanguins proviennent des cellules souches hématopoïétiques (CSH). Elles sont caractérisées par leur capacité unique à s'auto-renouveler et à produire divers progéniteurs qui se différencient en cellules sanguines matures (Metafuni, Chiusolo et al. 2018).

Pour assurer l'homéostasie hématopoïétique tout au long de la vie de l'organisme, une balance entre différenciation et auto-renouvellement est nécessaire et doit être finement régulée. C'est un processus essentiel pour maintenir une quantité adaptée de cellules sanguines différentes, à tout moment (état d'homéostasie sanguine). C'est la raison pour laquelle le destin des CSH est soumis à une série d'événements intracellulaires qui entraîneront un état de quiescence ou prolifération, de différenciation ou auto-renouvellement, de survie ou mort cellulaire. Une hématopoïèse considérée comme physiologique assure un équilibre constant, même en situation particulière (blessure ou infection). Les perturbations dans le système hiérarchique de l'hématopoïèse aboutissent à des troubles hématologiques. D'un côté, un épuisement de la réserve en CSH pourrait se produire dans le cas où le processus de différenciation est excessif ou l'auto-renouvellement est insuffisant. D'un autre côté, une différenciation insuffisante ou un auto-renouvellement excessif peut conduire à la formation de maladies sanguines (Rieger and Schroeder 2012).

Le modèle hiérarchique des cellules CSH se compose d'une série de progéniteurs multipotents, oligopotents et unipotents qui se différencient vers des cellules hématopoïétiques de plus en plus matures (**Figure 1.1**), tout en perdant leur potentiel d'autorenouvellement. À noter que les cellules totipotentes ont la capacité de produire tous les types de cellules de l'organisme, y compris les cellules souches embryonnaires. Ces dernières sont un exemple de cellules pluripotentes qui ont le potentiel de produire tous les types cellulaires. Les cellules multipotentes, comme les CSH, peuvent produire un nombre limité de types cellulaires dans une lignée particulière.

Les CSH génèrent des progéniteurs primitifs multipotents (MPP) ayant le plein potentiel de donner les différents types cellulaires mais présentent une capacité d'auto-renouvellement plus faible que les CSH (Osawa, Hanada et al. 1996, Benveniste, Frelin et al. 2010). Elles permettent la formation de progéniteurs oligopotents (Doulatov, Notta et al. 2010). Les progéniteurs communs de cellules myéloïdes (CMPs) se différencient en progéniteurs granulocytes/macrophages (GMP) ainsi qu'en progéniteurs mégacaryocytes/érythroïdes (MEP) qui pourront ensuite se différencier en hématies et plaquettes. Les progéniteurs lymphoïdes communs (CLP) se différencient en précurseurs de cellules B, T et NK (Murre 2007). Ces étapes successives de maturation s'achèvent par la création et la libération des cellules terminales dans le sang périphérique (**Figure 1.1**).

Des études récentes sur les étapes de différenciation *in vitro* ont remis en question le modèle classique de différenciation hiérarchique des progéniteurs hématopoïétiques. Les progéniteurs, limités à une ou deux lignées cellulaires, pourraient contourner les progéniteurs multipotents et seraient générés directement à partir des CSH (Yamamoto, Morita et al. 2013, Notta, Zandi et al. 2016, Haas, Trumpp et al. 2018). Malgré des études qui suggèrent une révision de l'organisation hiérarchique de l'hématopoïèse humaine, le modèle actuel est celui qui est largement utilisé pour étudier les mécanismes moléculaires des CSH et des progéniteurs (Pinho and Frenette 2019).



**Figure 1.1 – Processus de l’hématopoïèse.**

Toutes les cellules hématopoïétiques dérivent d’une population de cellules souches hématopoïétiques (HSC ou CSH) qui se divisent en deux groupes : celles au potentiel de repopulation à court terme (ST) ou à long terme (LT). Elles diffèrent par leur capacité de prolifération plus ou moins limitée.

Les principaux facteurs de transcription et cytokines sont indiqués respectivement en rouge et en bleu. Les encadrés rouges indiquent les différents marqueurs de surface des cellules sanguines qui ont été identifiés chez la souris.

MPP : progéniteurs multipotents ; CLP : progéniteurs lymphoïdes communs ; CMP : progéniteurs myéloïdes communs ; MEP : progéniteurs mégacaryocytes-érythrocytes ; GMP : progéniteurs granulocytes-monocytes.

*Le graphique est adapté de l’étude (Cheng, Zheng et al. 2020) et est reproduit avec leur permission.*

### 1.2.3 Un processus finement régulé

Au sein du microenvironnement médullaire, cohabitent des cellules hématopoïétiques à différentes étapes de différenciation, des cellules stromales mésenchymateuses, des ostéoblastes, des cellules endothéliales, des adipocytes et des neurones périphériques (Mendez-Ferrer, Bonnet et al. 2020). Cette « niche de cellules souches » fait référence à un réseau multicellulaire complexe au sein duquel des interactions physiques et moléculaires entre cellules sont essentielles pour permettre la maintenance et la différenciation des CSH (Pinho and Frenette 2019).

La régulation de l'hématopoïèse est assurée par la combinaison de facteurs intrinsèques aux cellules (régulateurs transcriptionnels, épigénétiques et voies métaboliques) et des facteurs extrinsèques (comme les hormones et les signaux locaux du microenvironnement de la moelle osseuse). Afin d'assurer l'homéostasie sanguine, de contrôler la différenciation des cellules sanguines ainsi que de maintenir le stock des cellules souches hématopoïétiques, des cytokines et des facteurs de transcription permettent de contrôler finement l'hématopoïèse (Cheng, Zheng et al. 2020).

#### 1.2.3.1 Cytokines

Les cytokines constituent une large famille de ligands extracellulaires ayant la capacité de stimuler des réponses biologiques à divers types cellulaires en se fixant et en activant des récepteurs aux cytokines. Celles qui sont spécifiques au système hématopoïétique sont les interleukines (IL), les facteurs CSF (Colony-Stimulating Factors), les interférons, l'érythropoïétine (EPO) et la thrombopoïétine (TPO) (Robb 2007). La **Figure 1.1** illustre les principales cytokines intervenant dans l'hématopoïèse.

Les cytokines activent la prolifération et la différenciation de cellules hématopoïétiques et diffèrent selon l'état de maturation de celles-ci. Certaines sont spécifiques d'une lignée cellulaire tandis que d'autres affectent de multiples lignées. Par exemple, l'EPO régule les niveaux de progéniteurs érythrocytaires alors que l'IL-3 stimule la croissance de la plupart des lignées cellulaires. Certaines cytokines ont donc une action plus large, comme le GM-CSF, les facteurs de cellules souches (SCF) et la tyrosine kinase 3 de type Fms (FLT3), IL-2, IL-3 et IL-

7, qui agissent à la fois sur des progéniteurs mais également des lignées plus spécifiques (Metcalf 1998).

### 1.2.3.2 Facteurs de transcription

Les progéniteurs hématopoïétiques se différencient également via l'activation de programmes d'expression génique spécifiques et régulés par des facteurs de transcription. Ceux-ci varient selon la lignée cellulaire. À chaque étape de différenciation, les gènes associés à la lignée s'expriment tandis que les autres gènes liés aux autres voies restent inactifs. Les principaux facteurs de transcription intervenant dans l'hématopoïèse sont représentés dans la **Figure 1.1**.

Plusieurs facteurs jouent un rôle clé dans le développement myéloïde. Par exemple, PU.1/SPI1 est un facteur essentiel dans la lignée myéloïde (Graf and Enver 2009). CEBP- $\alpha$ , CEBP- $\beta$  et CEBP- $\epsilon$  agissent dans l'engagement vers les granulocytes, macrophages et monocytes (Zhang, Zhang et al. 1997, Verbeek, Wachter et al. 2001, Tavor, Vuong et al. 2002, Dahl, Walsh et al. 2003, Laiosa, Stadtfeld et al. 2006). IRF8 est également impliqué dans la lignée myéloïde, notamment les cellules dendritiques et les monocytes (Becker, Michael et al. 2012, Kurotaki, Yamamoto et al. 2014). D'autres facteurs comme GATA-1, KLF1 ou encore GFI1B sont essentiels pour l'engagement dans les lignées érythrocytaires et mégacaryocytaires (Iwasaki, Mizuno et al. 2003). Corréler les mécanismes de régulation génique aux programmes de différenciation dans les populations myéloïdes a récemment été clarifié de manière plus précise grâce à des approches de séquençage de cellule unique (*single-cell*) (Paul, Arkin et al. 2015).

En ce qui concerne la régulation des CSH, certains facteurs de transcription sont impliqués dans des translocations chromosomiques ou des mutations somatiques à l'origine de cancers hématopoïétiques, comme des leucémies lymphoïdes aiguës de type T (SCL/TAL1 et LMO-2) ou des leucémies myéloïdes ou lymphoïdes (KMT2A, RUNX1, ETV6) (Orkin and Zon 2008).

## 1.3 Épigenétique

L'épigenétique correspond à l'étude des mécanismes héréditaires et réversibles qui influencent l'expression des gènes sans causer de modifications dans la séquence de l'ADN. Les changements épigénétiques jouent un rôle essentiel dans la régulation de l'expression des gènes. Par conséquent, des modifications épigénétiques aberrantes peuvent impacter la prolifération, la différenciation ou la survie des cellules et être associées à des types de cancers spécifiques (Berdasco and Esteller 2010). Ces changements incluent la méthylation de l'ADN, les modifications des histones et des ARNs non-codants. Nous nous concentrerons sur les deux premiers mécanismes dans le restant de cette thèse.

### 1.3.1 Méthylation de l'ADN

La méthylation de l'ADN est un mécanisme épigénétique qui implique le transfert d'un groupement méthyl, provenant du métabolite S-AdénosylMéthionine (SAM), au niveau de la position du 5<sup>ème</sup> carbone des résidus cytosines pour les transformer en 5-méthylCytosine (5mC). La réaction est catalysée par la famille des ADN méthyltransférases (DNMT). Les enzymes DNMT3A et DNMT3B permettent l'établissement de nouvelles méthylations : ce sont les DNMT *de novo*. À l'inverse, la protéine DNMT1 fonctionne pendant la réplication de l'ADN en assurant la copie du modèle de méthylation de l'ADN du brin parental sur le brin nouvellement synthétisé. La méthylation de l'ADN, notamment au niveau des îlots CpG, permet la régulation de l'expression des gènes en favorisant le recrutement de protéines impliquées dans la répression ou bien en empêchant la fixation de facteurs de transcription à l'ADN (Moore, Le et al. 2013).

L'approche la plus fréquemment utilisée pour étudier la méthylation des cytosines est la technique de séquençage de l'ADN après traitement au bisulfite (methyl-seq). Ce dernier convertit des cytosines non méthylées en uracile, sans affecter les cytosines méthylées. La comparaison entre le séquençage de l'ADN traité et la condition contrôle non traitée permet de localiser les régions méthylées à travers le génome et de quantifier le pourcentage de cytosines méthylées (Harris, Wang et al. 2010).

### 1.3.2 Modifications des histones

Les protéines histones sont l'un des éléments informatifs de l'épigénétique. Environ 146 pb d'ADN entourent chaque octamère d'histones pour former un nucléosome, sous-unité de base de la chromatine. Chaque nucléosome est composé d'un tétramère H3/H4 et de deux dimères H2A/H2B (Van Holde, Allen et al. 1980, Luger, Mader et al. 1997). Les unités nucléosomales sont compactées et condensées pour former la chromatine. Les extrémités N-terminales des histones H3 et H4 peuvent subir des modifications post-traductionnelles. Celles-ci représentent l'un des mécanismes épigénétiques que les organismes multicellulaires utilisent pour garantir l'expression spatio-temporelle des gènes impliqués dans la différenciation et le développement (Strahl and Allis 2000, Milne, Dou et al. 2005, Kouzarides 2007). Ces modifications sont appelées "marques", peuvent être de différentes formes (phosphorylation (P), acétylation (Ac), méthylation (Me, avec un groupement mono-, di- ou tri-méthyle), et ubiquitination (Ub)). Elles peuvent être déposées, interprétées ou effacées respectivement par des protéines appelées *readers*, *writers*, *erasers*. Par exemple, les groupements méthyles situés sur les lysines sont ajoutés ou retirés respectivement par des enzymes de type histone lysine méthyltransférase (KMT) ou de type histone lysine déméthylase (KDM). Nous discuterons plus en détail de la famille des KMT2 au point 1.5.1.2. Par ailleurs, le « code histone » vise à établir un lien entre l'activité transcriptionnelle et les marques d'histones. Les marques d'histones acétylées (H3K9ac et H3K27ac) sont associées une activation du gène alors que les modifications de type méthylation peuvent être corrélées à l'activation (H3K4 ou H3K79) ou à la répression des gènes (H3K9 ou H3K27). Elles peuvent aussi permettre de délimiter des régions fonctionnelles du génome. Par exemple, H3K4me1 et H3K27Ac marquent les régions amplificatrices (ou *enhancer*) ; H3K4me3 est caractéristique des promoteurs actifs ; H3K79me2 et H3K36me3 marquent les régions codantes des gènes actifs (Allis and Jenuwein 2016).

Pour étudier ces modifications, il est possible de réaliser une approche d'immunoprécipitation de chromatine suivie d'une étape de séquençage haut-débit (ChIP-seq) (Furey 2012). Cette technique se base sur l'utilisation d'anticorps reconnaissant spécifiquement les marques d'histones. Les données de séquençage de l'ADN immunoprécipité permettent d'identifier les profils des marques d'histones au niveau de l'ensemble des gènes et d'y associer des données d'expression via des approches de RNA-seq. À noter que la méthode d'ATAC-seq

permet de connaître l'état d'ouverture de la chromatine et ainsi de déterminer si un gène est potentiellement prêt – ou non – à être transcrit (Buenrostro, Wu et al. 2015). En quelques mots, cette récente technique se base sur l'utilisation d'une transposase capable de cliver et d'ajouter des adaptateurs aux extrémités des fragments d'ADN situés dans les régions ouvertes du génome. Après purification et amplification par PCR, une étape de séquençage permet de déduire les régions où la chromatine est ouverte ainsi que les potentiels sites de fixation de facteurs de transcription.



## **1.4 La leucémie myéloïde aiguë**

### **1.4.1 Définition**

La leucémie est une forme de cancer qui affecte les cellules du système hématopoïétique. Elle se développe non seulement à partir des CSH qui prolifèrent de manière incontrôlée et donnent des cellules souches leucémiques (CSL), mais également des progéniteurs qui perdent leur capacité à se différencier et deviennent des cellules blastiques. Ces cellules sont capables d'infiltrer la moelle osseuse, le sang périphérique ainsi que d'autres tissus, et devenir progressivement plus nombreuses que les cellules fonctionnelles (Dohner, Weisdorf et al. 2015).

Selon la voie de différenciation concernée et la vitesse de développement du cancer, la leucémie peut être classifiée en différentes catégories : myéloïde ou lymphoïde, aiguë (quelques semaines) ou chronique (quelques années).

### **1.4.2 Manifestations cliniques et méthodes de diagnostic**

Les symptômes que développent les patients leucémiques sont variables mais causés principalement par l'insuffisance hématopoïétique comme une neutropénie (taux bas de granulocytes neutrophiles), une anémie (nombre insuffisant d'érythrocytes) ou une thrombocytopenie (déficit de plaquettes). Cela se traduit généralement par de la fièvre et une faiblesse généralisée (et moins fréquemment par une coagulation intravasculaire disséminée ou un syndrome de leucostase) et apparaissent quelques semaines ou quelques mois avant de poser un diagnostic chez les patients atteints de leucémie myéloïde aiguë (LMA).

L'évaluation au microscope du sang périphérique et de la moelle osseuse est la première étape pour diagnostiquer une leucémie aiguë. Le diagnostic de la LMA est réalisé lorsqu'il y a au moins 20% de blastes myéloïdes retrouvés dans le sang ou la moelle osseuse (Arber, Orazi et al. 2016). Dans de rares cas, c'est la détection de réarrangements chromosomiques qui permet le diagnostic de LMA dans les patients ayant un taux inférieur de blastes. La cytométrie en flux est utilisée pour décrire l'immunophénotype, c'est-à-dire le profil aberrant des marqueurs présents à la surface des cellules leucémiques. L'évaluation cytogénétique est nécessaire pour rechercher des anomalies génétiques récurrentes et ainsi permettre de préciser le sous-groupe auquel appartient le patient. L'analyse du caryotype (ou des techniques de séquençage du

génomique) permet la détection d'anomalies cytogénétiques. Les techniques de FISH (ou « fluorescence in situ hybridization ») ou de PCR (ou « Polymerase Chain Reaction ») sont aussi utilisées pour réaliser des tests rapides de détection d'altérations chromosomiques spécifiques.

### 1.4.3 Classification de la leucémie myéloïde aiguë

Il existe différentes classifications se basant sur la génétique, l'immunophénotype et la morphologie. Dans les années 1970, la classification FAB (French-American-British) définissait 8 sous-types de LMA (de M0 à M7) selon des critères principalement morphologiques, immunophénotypiques et de degré de différenciation (**Tableau 1.1**) (Bennett, Catovsky et al. 1976).

Sous-type classification FAB	Nom de la maladie
<b>M0</b>	LMA avec différenciation minimale
<b>M1</b>	LMA sans maturation
<b>M2</b>	LMA avec maturation
<b>M3</b>	Leucémie promyélocytaire aiguë
<b>M4</b>	Leucémie myélomonocytaire aiguë
<b>M5</b>	Leucémie monocytaire aiguë
<b>M6</b>	Leucémie érythroblastique aiguë
<b>M7</b>	Leucémie mégacaryocytaire aiguë

**Tableau 1.1 – Classification FAB des leucémies myéloïdes aiguës.**

Tableau adapté de (Bennett, Catovsky et al. 1976).

La classification FAB est aujourd'hui de moins en moins utilisée. C'est la classification de l'Organisation Mondiale de la Santé (OMS) qui devient la référence dans l'établissement des diagnostics. Elle tient compte de multiples critères pour affiner l'appartenance aux différents groupes. L'objectif est de se baser sur des caractéristiques cliniques, la morphologie, l'immunophénotype mais surtout l'information génétique pour décrire des groupes distincts et sans chevauchement entre eux (Arber, Orazi et al. 2016). Le **Tableau 1.2** présente les catégories majeures de la classification de l'OMS mise en jour en 2016.

Les LMA causées par la présence d'une translocation chromosomique de type *KMT2A-MLLT3* et faisant l'objet de cette thèse se retrouvent dans la première catégorie et correspondent souvent au sous-groupe M4/5 dans la classification FAB.

---

**LMA avec des anomalies cytogénétiques récurrentes**

---

LMA avec t(8;21)(q22;q22) ; *RUNX1-RUNX1T1*

LMA avec inv(16)(p13.1q22) ou t(16;16)(p13.1;q22) ; *CBFB-MYH11*

Leucémie aiguë promyélocytaire avec *PML-RARA*

**LMA avec t(9;11)(p22;q23) ; *MLLT3-KMT2A***

LMA avec t(6;9)(p23;q34) ; *DEK-NUP214*

LMA avec inv(3)(q21q26.2) ou t(3;3)(q21;q26.2) ; *GATA2, MECOM*

LMA (mégacaryoblastique) avec t(1;22)(p13;q13) ; *RBM15-MKLI*

LMA avec mutation *NPM1*

LMA avec mutation bi allélique de *CEBPA*

*Entrée provisoire : LMA avec BCR-ABL1*

*Entrée provisoire : LMA avec mutation RUNX1*

---

**LMA avec anomalies associées aux myélodysplasies**

---

**Néoplasies myéloïdes post-traitements**

---

**LMA, sans autre spécification :**

---

LMA avec différenciation minimale (FAB : LMA M0)

LMA sans maturation (FAB : LMA M1)

LMA avec maturation (FAB : LMA M2)

Leucémie aiguë myélomonocytaire (FAB : LMA M4)

Leucémie aiguë monoblastique/monocytaire (FAB : LMA M5)

Leucémie aiguë érythroïde pure (FAB : LMA M6)

Leucémie aiguë mégacaryoblastique (FAB : LMA M7)

Leucémie aiguë myéloblastique à composante basophile

Leucémie aiguë avec myélofibrose

---

**Sarcome granulocytaire**

---

**Proliférations myéloïdes associées au syndrome de Down**

---

Réaction leucémoïde transitoire

LMA associée au syndrome de Down

---

**Tableau 1.2 – Classification de l’OMS des leucémies myéloïdes aiguës.**

Tableau adapté de (Arber, Orazi et al. 2016).

### 1.4.4 Facteurs de pronostic des leucémies myéloïdes aiguës

Le pronostic de la LMA est déterminé par les informations relatives à la maladie (facteurs génétiques, nombre de globules blancs, antécédent de syndrome myélodysplasique) et au patient (âge, état général, comorbidité). Celles-ci prédisent respectivement la résistance au traitement standard et la probabilité de survie après un traitement. En 2010, la classification ELN (European LeukemiaNet) a été conçue pour standardiser la stratification des risques chez les patients adultes atteints de LMA en utilisant les données cytogénétiques (Dohner, Estey et al. 2010). Les mutations des gènes *NPM1*, *CEBPα* et *FLT3* (avec des duplications en tandem (ITD)) sont notamment des marqueurs moléculaires de risque bien établis dans les LMA. Les autres gènes *TP53*, *RUNX1*, *DNMT3A*, *IDH1/2* et *ASXL1* sont également utilisés pour l'établissement du pronostic mais leur rôle dans la prise de décision clinique n'a pas encore été bien défini (Roug, Hansen et al. 2014, Dohner, Weisdorf et al. 2015). Les patients sont répartis dans l'un des quatre groupes de risque : favorable, intermédiaire-I, intermédiaire-II et défavorable (Tableau 1.3).

Groupe de risque	Caractéristiques génétiques
<b>Favorable</b>	t(8;21)(q22;q22); <i>RUNX1-RUNX1T1</i>
	inv(16)(p13.1q22) ou t(16;16)(p13.1;q22); <i>CBFB-MYH11</i>
	Mutation de <i>NPM1</i> sans <i>FLT3-ITD</i> (caryotype normal)
	Mutations bialléliques de <i>CEBPα</i> (caryotype normal)
<b>Intermédiaire-I</b>	Mutations de <i>NPM1</i> et <i>FLT3-ITD</i> (caryotype normal)
	<i>NPM1</i> wild-type et <i>FLT3-ITD</i> (caryotype normal)
	<i>NPM1</i> wild-type sans <i>FLT3-ITD</i> (caryotype normal)
<b>Intermédiaire-II</b>	t(9;11)(p22;q23); <i>KMT2A-MLL3</i>
	Anomalies cytogénétiques ni favorables, ni défavorables
<b>Défavorable</b>	inv(3)(q21q26.2) ou t(3;3)(q21;q26.2); <i>GATA2</i> , <i>MECOM (EVII)</i>
	t(6;9)(p23;q34); <i>DEK-NUP214</i>
	t(v;11)(v;q23); <i>KMT2A</i> réarrangé (excepté t(9;11)(p22;q23))
	-5 ou del(5q)
	-7
	-17/anomalie(17p)
	Caryotype complexe (≥3 altérations ; excepté les patients avec t(9;11), t(15;17), t(8;21), inv(16) ou t(16;16))

**Tableau 1.3 – Classification European LeukemiaNet (ENL) des leucémies myéloïdes aiguës.**

Tableau adapté de (Dohner, Estey et al. 2010).

### 1.4.5 Le paysage génétique des leucémies myéloïdes aiguës

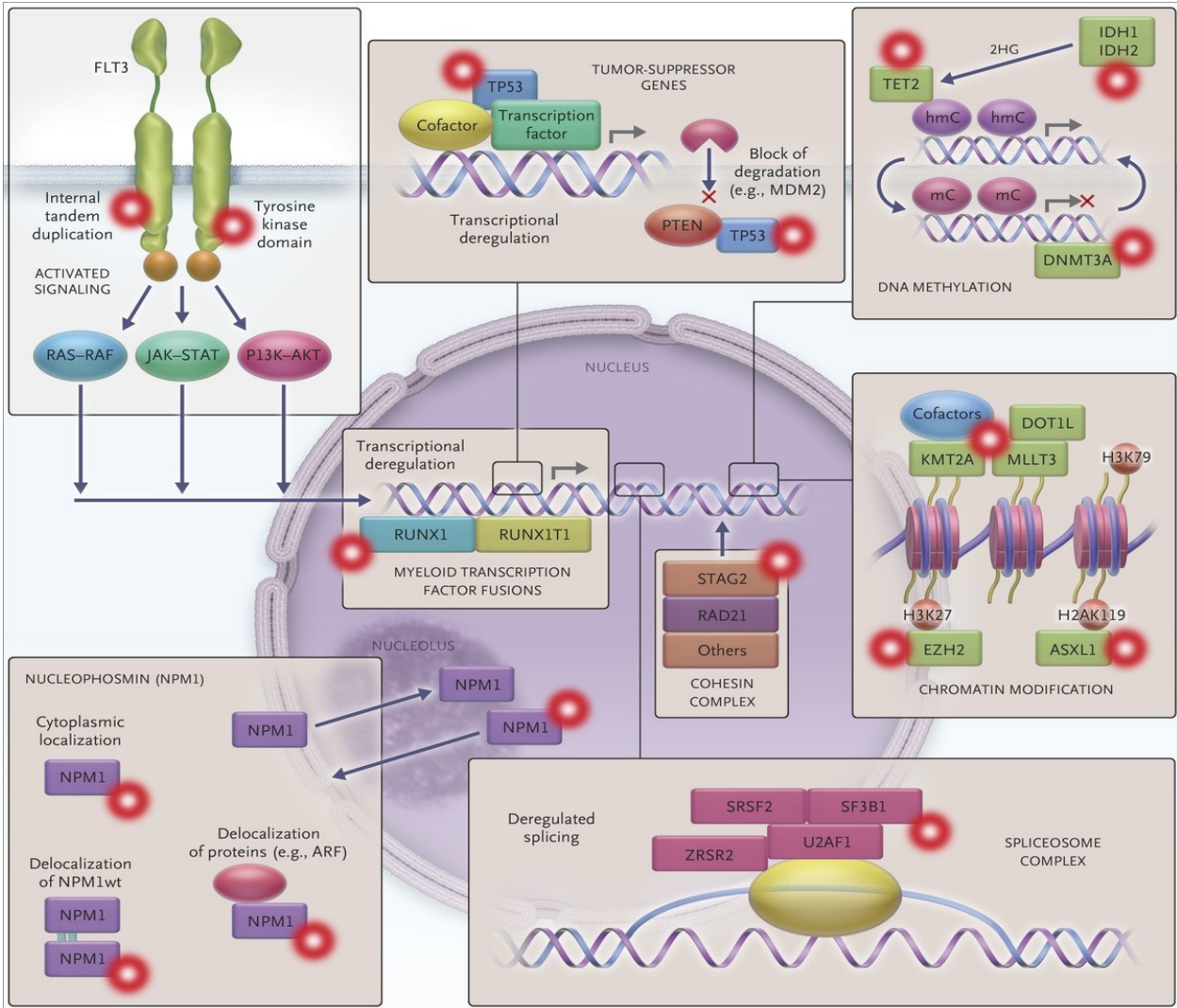
L'arrivée des techniques de séquençage nouvelle génération a permis de cartographier la fréquence des mutations chez les patients, l'architecture et l'évolution des clones, ainsi que le paysage épigénétique de la maladie.

Une étude récente (Bolouri, Farrar et al. 2018) a mis en lumière les différences génétiques liées à l'âge, entre patients adultes et pédiatriques atteints de LMA, et démontre la très forte hétérogénéité génétique de la maladie. Il a pu être observé que les translocations chromosomiques sont retrouvées plus fréquemment chez les jeunes patients. Les mutations ponctuelles, quant à elles, concernent certains gènes spécifiquement chez les enfants (comme *NRAS*, *KIT* ou *KRAS*) ou chez les adultes (*DNMT3A*, *NPM1*, *IDH1*, *IDH2*). Par ailleurs, les méthodes d'analyse à l'échelle de la cellule (*single-cell*) ont permis de révéler la complexité et l'imprévisibilité de l'évolution clonale de la maladie au cours du temps (Morita, Wang et al. 2020) ce qui démontre l'urgence de développer des médecines de précision.

L'analyse de 200 patients dans le cadre d'une étude menée par le réseau TCGA (*The Cancer Genome Atlas Research*) a permis de révéler l'existence des gènes fréquemment mutés qui permettent d'affiner le pronostic et d'aider à mieux adapter le traitement en clinique. Chaque patient possède en moyenne 13 mutations, dont 5 sont considérées comme étant des mutations pilotes (ou *driver mutation*) (Cancer Genome Atlas Research, Ley et al. 2013). Il est possible de les regrouper en 8 catégories fonctionnelles résumées sur la **Figure 1.2** (Dohner, Weisdorf et al. 2015, Saultz and Garzon 2016) :

- La mutation du gène *NPM1* entraîne une relocalisation cytoplasmique de la protéine NPM1 et de ses partenaires d'interaction ce qui stimule la prolifération des cellules myéloïdes et le développement leucémique.
- La translocation chromosomique t(8;21)(q22;q22); *RUNX1-RUNX1T1* ou les mutations touchant *RUNX1*, un gène codant pour un facteur de transcription myéloïde, impactent la transcription et la différenciation hématopoïétique.
- Les mutations du gène codant pour le récepteur à activité tyrosine kinase FLT3 facilitent la prolifération des cellules leucémiques via l'activation des voies de signalisation telles que RAS-RAF, JAK-STAT et PI3K-AKT.

- Les mutations touchant les gènes impliqués dans les spliceosomes (*SRSF2*, *SF3B1*, *U2AF1*, *ZRSR2*) affectent le processus d'épissage des ARN dans les cellules leucémiques.
- *STAG2* et *RAD21* sont deux gènes du complexe de cohésine pour lesquelles des mutations peuvent nuire à la régulation transcriptionnelle.
- Certaines mutations peuvent affecter directement la méthylation de l'ADN (*DNMT3A* et *TET2* ou encore *IDH1/2*), ce qui entraîne une dérégulation de l'expression d'autres gènes.
- Comme dans de nombreux autres types de cancers, les mutations touchant les gènes suppresseurs de tumeur tels que *p53* peuvent compromettre la régulation transcriptionnelle.
- La dernière catégorie concerne les gènes et fusions de gènes impliqués dans l'homéostasie épigénétique des cellules tels que la translocation chromosomique faisant intervenir le gène *KMT2A* qui fait l'objet du reste de cette thèse.



**Figure 1.2 – Résumé des huit catégories de gènes fréquemment mutés dans la leucémie myéloïde aiguë.**

Chaque point rouge correspond à un gène muté. Parmi tous les cas de LMA, il est possible de classer les différents gènes du plus au moins fréquemment mutés : *NPM1* (25-30%), *DNMT3A* (18-22%), *FLT3* (20%), *IDH1/2* (15-20%), *TET2* (9-23%), *RUNX1* (5-13%), *CEBPα* (6-10%), *ASXL1* (5-11%), *p53* (8-14%), *c-KIT* (5%).

Reproduit avec la permission de (Dohner, Weisdorf et al. 2015), copyright Massachusetts Medical Society.

## 1.5 Les LMA causées par une fusion de gènes de type *KMT2A*

### 1.5.1 Le gène *KMT2A* sauvage

Le gène *KMT2A* (ou *MLL* pour *Mixed Lineage Leukemia* ou *ALL-1*) a été identifié chez l'humain en raison de son association avec un point de cassure retrouvé fréquemment dans un sous-groupe de leucémies aiguës (Ziemin-van der Poel, McCabe et al. 1991, Djabali, Selleri et al. 1992, Gu, Nakamura et al. 1992, Tkachuk, Kohler et al. 1992). La plupart des mutations du gène *KMT2A* dans la leucémie sont des translocations chromosomiques qui entraînent la troncature du gène *KMT2A* et sa fusion avec un nombre de plus en plus croissant de gènes partenaires (Meyer, Burmeister et al. 2018).

Avant de discuter de la nature des fusions oncogéniques faisant intervenir *KMT2A*, cette partie détaille les connaissances actuelles de la protéine *KMT2A* considérée comme sauvage chez les mammifères (Crump and Milne 2019).

#### 1.5.1.1 Historique

Lorsque l'ADN complémentaire du gène *KMT2A* a été séquencé pour la première fois par différents groupes (Djabali, Selleri et al. 1992, Gu, Nakamura et al. 1992, Tkachuk, Kohler et al. 1992), une des premières caractéristiques mises en évidence a été son homologie avec le gène *trithorax*, retrouvé chez la drosophile (Mazo, Huang et al. 1990). *Trithorax* (ou *trx*) fait partie du groupe *trithorax* (*trxG*), identifié à l'origine comme régulateur des gènes homéotiques (ou Homéobox ou *Hox*) chez *Drosophila melanogaster*. Cet ensemble de gènes possède un rôle essentiel dans la structuration corporelle des organismes multicellulaires notamment dans le contrôle du développement de l'axe antéro-postérieur de la drosophile (Lewis 1978, Duncan and Montgomery 2002, Kassis, Kennison et al. 2017). Les protéines *TrxG* maintiennent l'expression des gènes *Hox* mais subissent l'activité répressive des protéines du groupe Polycomb (*PcG*) (Lewis 1978, Kassis, Kennison et al. 2017). Aucun de ces deux groupes n'est requis pour initier l'expression des gènes *Hox* mais ils sont indispensables pour maintenir les modèles d'expression une fois établis (Struhl and Akam 1985, Kassis, Kennison et al. 2017). Chez la souris, des travaux ont pu démontrer que l'orthologue de *KMT2A* chez les mammifères agit comme un membre du groupe *trxG*. En effet, les souris pour lesquelles le gène *KMT2A* est inactivé (« knock-out » (KO)) présentent une létalité embryonnaire et des défauts dans



l'organisation corporelle, causés par des modifications anormales du modèle d'expression des gènes *Hox* (Yu, Hess et al. 1995, Yu, Hanson et al. 1998). D'ailleurs, comme observé chez la drosophile, les modèles d'expression des gènes *Hox* chez les souris KO pour *KMT2A* ne sont pas affectés lors de l'initiation mais plutôt dans l'étape du maintien de l'expression et des stades ultérieurs de développement (Yu, Hess et al. 1995, Yu, Hanson et al. 1998, Beck, Faradji et al. 2010).

### 1.5.1.2 Structure de la famille MLL/KMT2

KMT2A est un membre de la famille des protéines méthyltransférases (MT) sur les lysines (K). C'est une famille très conservée chez les eucaryotes, particulièrement chez les mammifères pour lesquels 6 membres ont été caractérisés (Glaser, Schaft et al. 2006, Ruthenburg, Allis et al. 2007, Crump and Milne 2019). À noter que la nomenclature a été mise à jour depuis quelques années dans l'objectif d'illustrer plus spécifiquement les relations structurelles des différentes protéines. Alors que MLL était historiquement utilisé, il est aujourd'hui usuel de parler de KMT2. Cette famille est composée de trois paires de protéines structurellement très proches, chacune d'entre elles correspondant à une protéine retrouvée chez *Drosophila* (Ruthenburg, Allis et al. 2007, Shilatifard 2012). MLL (ou MLL1), MLL2, MLL3, MLL4, SETD1A et SETD1B correspondent respectivement à KMT2A, KMT2B, KMT2C, KMT2D, KMT2F et KMT2G. Bien que MLL5 (KMT2E) était originellement membre de la famille MLL, l'absence d'activité intrinsèque de type KMT a entraîné une divergence dans la classification (Emerling, Bonifas et al. 2002, Zhang, Novera et al. 2017). Le **Tableau 1.4** récapitule les différentes nomenclatures et correspondances entre espèces.

Le point commun entre les membres de la famille KMT2 est le fait qu'ils partagent tous un domaine SET, dont l'acronyme correspond aux premières lettres de protéines agissant sur la chromatine : **S**uppressor of variegation 3-9 (Su(var)3-9) ; **E**nhancer of Zeste (E(z)) et **T**rithorax. Ce domaine est responsable de la méthylation de la lysine 4 sur l'histone H3 (H3K4). Nous nous concentrons sur la protéine KMT2A pour le reste de la thèse.

Ancienne nomenclature	Nouvelle nomenclature	Correspondance chez <i>Drosophila</i>	Position dans le génome humain	Références
MLL/MLL1	KMT2A	Protéine trithorax	11q23	(Ziemin-van der Poel, McCabe et al. 1991)
MLL2	KMT2B		19q13	(FitzGerald and Diaz 1999, Huntsman, Chin et al. 1999)
MLL3	KMT2C	Protéine proche de trithorax ( <i>trithorax-related trr</i> ) (Sedkov, Benes et al. 1999, Shilatifard 2012)	7q36	(Tan and Chow 2001, Ruault, Brun et al. 2002)
MLL4	KMT2D		12q13	(Prasad, Zhadanov et al. 1997, Froimchuk, Jang et al. 2017)
SETD1A	KMT2F	Proche de la protéine Set1 retrouvée chez <i>Drosophila</i> (dSet1), homologue de Set1 chez <i>S. cerevisiae</i>	16p11	(Wysocka, Myers et al. 2003)
SETD1B	KMT2G		12q24	(Lee, Tate et al. 2007)

**Tableau 1.4 – Résumé de la nomenclature des membres de la famille de protéines KMT2/MLL.**

Adapté de (Crump and Milne 2019).

### 1.5.1.3 Fonction de la protéine KMT2A dans l'hématopoïèse

L'activité catalytique de KMT2A ne semble pas être essentielle à la survie ou au fonctionnement des cellules, comparativement à l'ensemble de la protéine KMT2A. En effet, l'insertion artificielle d'un codon stop induisant la perte de KMT2A est synonyme de mort embryonnaire précoce et d'anomalies hématopoïétiques sévères (Yu, Hess et al. 1995, Yagi, Deguchi et al. 1998) alors que la délétion génétique du domaine SET de KMT2A chez la souris n'amène que de légers défauts squelettiques, sans impact majeur sur le système sanguin (Terranova, Agherbi et al. 2006, Mishra, Zaffuto et al. 2014). Il a par ailleurs été démontré que la délétion de *KMT2A* dans des souris adultes entraîne une insuffisance médullaire mortelle causée notamment par : i) une diminution de la prolifération cellulaire et ii) une réponse réduite à l'entrée dans le cycle cellulaire (induite par les cytokines) des cellules progénitrices myéloïdes (Jude, Climer et al. 2007).

En plus de jouer un rôle clé dans l'hématopoïèse (Jude, Climer et al. 2007, McMahon, Hiew et al. 2007), la protéine sauvage KMT2A assure le maintien de l'activation des gènes essentiels tels que les gènes *Hox* lors de l'embryogenèse et la neurogenèse (Yu, Hess et al. 1995, Yu,

Hanson et al. 1998, Hanson, Hess et al. 1999, Lim, Huang et al. 2009). Certaines études génomiques par ChIP-seq ont révélé l'association de KMT2A avec un groupe de promoteurs actifs mais également avec des régions amplificatrices (ou *enhancers*) (Milne, Dou et al. 2005, Wang, Lin et al. 2009, Wang, Yang et al. 2011, Yang, Flynn et al. 2014). Par ailleurs, la perte de KMT2A n'entraîne pas de changements globaux dans la méthylation de H3K4 ce qui laisse à penser qu'il pourrait y avoir une compensation de la perte de KMT2A par d'autres membres de la famille (Wang, Lin et al. 2009). Finalement, les données actuelles soulèvent des questions quant à la fonction spécifique de la méthylation de H3K4 au cours du développement. La méthylation de H3K4 est associée à l'expression des gènes puisque de nombreuses études montrent que les perturbations de celle-ci à travers le génome conduisent à des changements dans l'expression de gènes. Sa contribution au niveau mécanistique demeure cependant incomprise (Krivtsov and Armstrong 2007, Wang, Lin et al. 2009).

À noter que le rôle de KMT2A dans la méthylation de substrats non-histoniens jouant un rôle important dans le processus de l'hématopoïèse n'a pas encore été exploré.

#### **1.5.1.4 Structure détaillée de la protéine KMT2A sauvage**

Le gène humain *KMT2A* est composé de 37 exons. Il code pour une large protéine nucléaire d'environ 500 kDa et dont le nombre d'acides aminés varie entre 3958 et 4005 dépendamment des épissages alternatifs qui concernent, par exemple, l'exon 2 dans 66% des transcrits ou les exons 15 et 16 (Rossler and Marschalek 2013).

De la partie N-terminale à la partie C-terminale, chacun des domaines permet de remplir une fonction particulière. Bien que les mécanismes exacts qui contrôlent l'association de KMT2A avec la chromatine ne soient pas totalement compris, la caractérisation de certains domaines de la protéine suggère que KMT2A forme un complexe multi-protéique agissant sur la régulation transcriptionnelle. Le résumé des différents domaines est présenté en **Figure 1.3**.

##### **1.5.1.4.1 Partie N-terminale de la protéine KMT2A**

La portion N-terminale de KMT2A interagit avec la protéine Menin (MEN1) (Hughes, Rozenblatt-Rosen et al. 2004) ainsi qu'avec le facteur de croissance dérivé de l'épithélium du cristallin (LEDGF) (Yokoyama and Cleary 2008). Ce sont des sous-unités nécessaires pour le recrutement des protéines KMT2A (sauvage ou fusion oncogénique) aux sites de fixation de

leurs gènes cibles (Milne, Hughes et al. 2005, Caslini, Yang et al. 2007, Yokoyama and Cleary 2008, Onodera, Yamashita et al. 2010, Thiel, Blessington et al. 2010, Huang, Gurung et al. 2012). LEDGF se fixe aux marques d'histones H3K36me2/3 via son domaine PWWP. Selon les contextes cellulaires, KMT2A interagit avec des facteurs de transcription tels que OCT4 (Ang, Tsai et al. 2011), E2F (Tyagi, Chabes et al. 2007), PAX7 (Kawabe, Wang et al. 2012) et NF-Y (Fossati, Dolfini et al. 2011) mais la fonction de ces interactions n'est toujours pas clairement démontrée.

Trois domaines de type « AT-crochet » (ou « AT-hook ») sont responsables de la fixation de KMT2A à des régions de l'ADN riches en Adénine et Thymines (Zelevnik-Le, Harden et al. 1994).

Deux domaines SNL1 et SNL2 sont responsables de la localisation nucléaire de KMT2A.

Un domaine CXXC assure la reconnaissance et la fixation à des îlots CpG non-méthylés ainsi que le complexe d'élongation associé à l'ARN polymérase (PAFc) (Birke, Schreiner et al. 2002, Allen, Grummitt et al. 2006, Cierpicki, Risner et al. 2010, Xu, Bian et al. 2011). Ce dernier assure le recrutement de KMT2A et des fusions oncogéniques de type KMT2A pour permettre l'activation des gènes cibles comme *HOXA9* (Milne, Kim et al. 2010, Muntean, Tan et al. 2010).

Un groupe de 4 domaines de type PHD (Plant Homology Domain) interagit avec la cyclophiline 33 (Cyp33) (Fair, Anderson et al. 2001, Xia, Anderson et al. 2003, Hom, Chang et al. 2010, Park, Osmer et al. 2010, Wang, Song et al. 2010), les marques d'histones H3K4me2 et H3K4me3 (Hom, Chang et al. 2010, Milne, Kim et al. 2010, Wang, Song et al. 2010) et la protéine ECS<sup>ASB</sup> du complexe ubiquitine ligase E3 (Wang, Muntean et al. 2012).

Un bromodomaine (BRD) assure la reconnaissance des groupements de type acétyle des lysines situées sur les queues d'histones (Wang, Song et al. 2010).

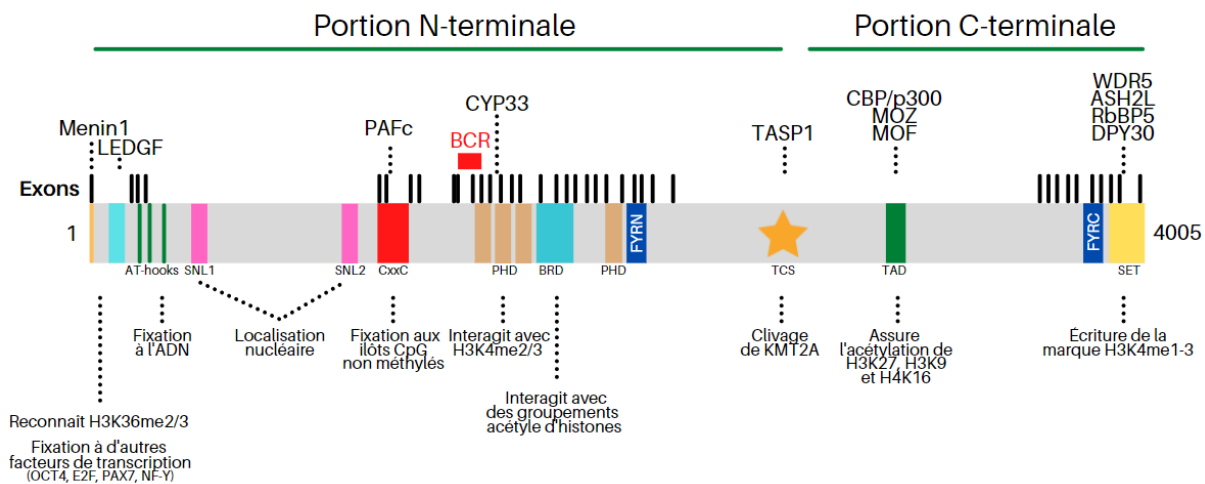
#### **1.5.1.4.2 Partie C-terminale de la protéine KMT2A**

La portion C-terminale possède deux domaines capables d'agir sur la chromatine.

Le premier est le domaine SET. C'est au début des années 2000 qu'il a été démontré que le domaine SET apporte la fonction de méthyltransférase pour la lysine 4 de l'histone H3 (H3K4) et plus précisément la catalyse des modifications de type monométhyle, diméthyle voire

triméthyle (Milne, Briggs et al. 2002, Nakamura, Mori et al. 2002, Wu, Lee et al. 2013). Les données montrent l'importance du domaine SET mais c'est l'ensemble du complexe KMT2A qui est requis pour permettre la méthylation de H3K4 (Rao and Dou 2015). La protéine de liaison au rétinoblastome (RbBP5), la sous-unité du complexe de l'histone méthyltransférase Set1/Ash2 (ASH2L) et les protéines WDR5 et DPY30 forment un complexe stable biochimiquement qui interagit avec KMT2A et permet l'activité catalytique du domaine SET (Dou, Milne et al. 2006, Patel, Dharmarajan et al. 2009).

Le deuxième domaine est le domaine transactivateur (TAD) par lequel KMT2A peut se lier physiquement à des histones acétyltransférases. CBP/p300, MOZ et MOF assurent l'acétylation de H3K27, H3K9 et H4K16 respectivement (Slany 2016). Il a d'ailleurs été observé que les promoteurs actifs présentent non seulement des marques d'histones H3K4me3 mais également des acétylations de la lysine 16 de l'histone 4 (H4K16) (Wang, Zang et al. 2008).



**Figure 1.3 – Schéma récapitulatif de la structure de la protéine KMT2A.**

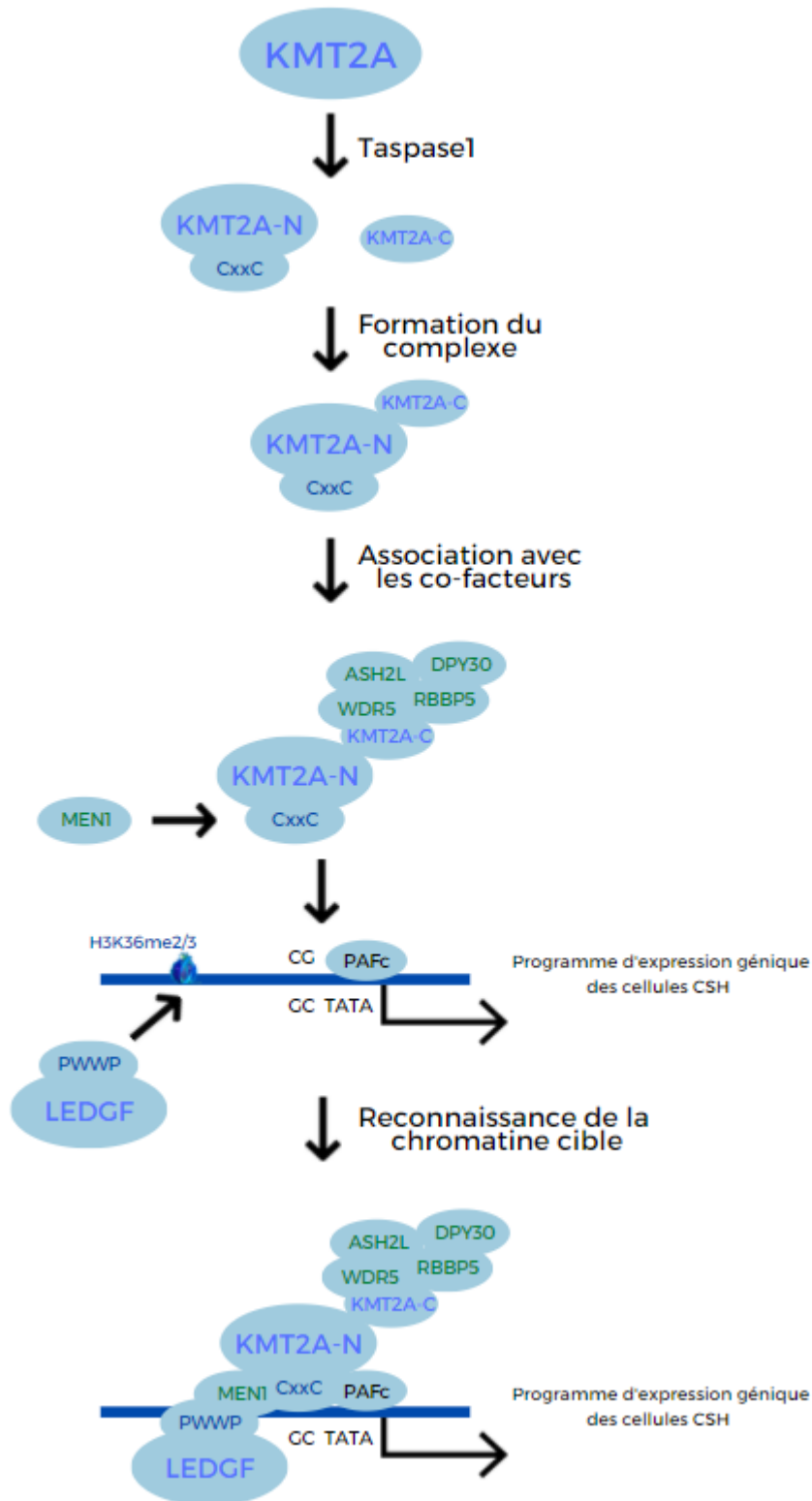
Les traits noirs épais délimitent les différents exons. Inspiré des revues de littérature : (Chan and Chen 2019) et (Ballabio and Milne 2012). *BCR* = *Breakpoint Cluster Region*

#### 1.5.1.4.3 Clivage de la protéine KMT2A

La protéine KMT2A peut être clivée par l'enzyme Taspase1 (TASP1). Cela génère un fragment N-terminal de 320 kDa (KMT2A-N) et un fragment C-terminal de 180 kDa (KMT2A-C) qui peuvent se dimériser et se réassembler dans le noyau pour former un complexe

moléculaire stable via l'interaction entre les domaines FYRN et FYRC (Nakamura, Mori et al. 2002, Yokoyama, Kitabayashi et al. 2002, Hsieh, Ernst et al. 2003, Yokoyama, Wang et al. 2004, Dou, Milne et al. 2005, Rao and Dou 2015). La **Figure 1.4** résume schématiquement la création du complexe.

L'ensemble des interactions de KMT2A avec les différentes protéines permet d'entrevoir les nombreuses possibilités et les différents rôles que peuvent jouer ces complexes dans la régulation de l'expression génique.



**Figure 1.4 – Schéma résumant la formation du complexe multiprotéique de KMT2A et la reconnaissance des cibles de KMT2A.**

Une partie des protéines est représentée par souci de simplification.

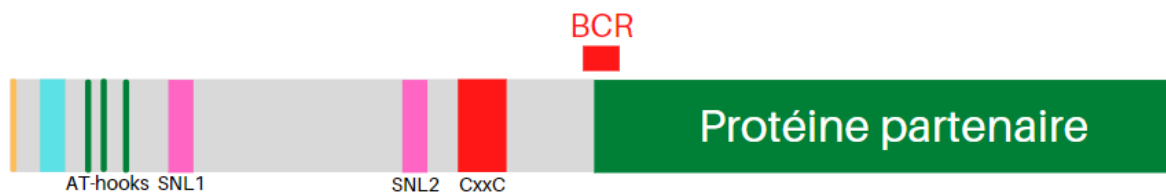
Figure inspirée de la revue (Takahashi and Yokoyama 2020)

## 1.5.2 Les translocations chromosomiques faisant intervenir le gène *KMT2A*

### 1.5.2.1 Origine moléculaire de la translocation

Des études menées sur des patients atteints de leucémies aiguës et porteurs de translocations chromosomiques touchant le bras long (q) du chromosome 11 en position 23 (11q23) ont permis d'identifier le gène *KMT2A* comme cible récurrente de ces translocations (Ziemin-van der Poel, McCabe et al. 1991, Tkachuk, Kohler et al. 1992, Rowley 1993). Plus de 98% des points de cassure du gène *KMT2A* sont localisés au niveau de la région BCR (*Breakpoint Cluster Region*) (voir **Figure 1.3**) d'une longueur d'environ 8.3 kb et située entre les exons 8 et 13 (Meyer, Burmeister et al. 2018). Cette séquence présente de nombreux sites de clivage de la topoisomérase II ainsi que des régions de rattachement à la matrice nucléaire. Ces translocations seraient causées par un défaut dans le mécanisme de réparation de l'ADN médié par le système de recombinaison non-homologue de type NHEJ (*Non-Homologous End-Joining*) (Lieber, Gu et al. 2010). En clinique, notamment dans le traitement de la LMA, l'interférence dans les processus cellulaires normaux via l'utilisation de traitements inhibiteurs de la topoisomérase II peut conduire à la formation de réarrangements chromosomiques (Strissel, Strick et al. 1998).

Les réarrangements chromosomiques aboutissent à la création d'une nouvelle protéine chimérique et sans décalage dans le cadre de lecture. Dans les cas des *KMT2A*-PTD (pour Duplications Partielles en Tandem), il est question d'une protéine *KMT2A* codée par la duplication des 5 à 12 premiers exons de *KMT2A* insérés dans les exons 11 et 12 (Krivtsov and Armstrong 2007). Dans les cas des fusions *KMT2A*, il s'agit d'une protéine codée par les 8 à 13 premiers exons de *KMT2A* et d'un nombre variable d'exons du gène partenaire de la fusion (**Figure 1.5**).



**Figure 1.5 – Schéma simplifié de la structure d'une protéine chimérique issue d'une translocation chromosomique de type *KMT2A*.**

À noter que la présence des domaines SNL forcent les protéines de fusion à rester dans le noyau.



## 1.5.2.2 Caractéristique des partenaires de fusion

### 1.5.2.2.1 Classification des partenaires de fusion

Il est possible de les classer en 5 différents groupes tel que décrit dans le **Tableau 1.5**. Dans cette thèse, nous nous concentrerons sur deux partenaires en particulier : MLLT3 et MLLT4. La première protéine appartient au premier groupe représentant la très large majorité (80%) des leucémies médiées par des fusions de type *KMT2A*. Il inclut des protéines nucléaires comme MLLT2, **MLLT3**, MLLT1, MLLT10 et ELL. Un deuxième groupe concerne environ 10% de ce type de leucémies et rassemble des protéines cytoplasmiques possédant des domaines d'oligomérisation qui sont essentiels pour assurer leur potentiel oncogénique : MLLT5, GAS7, EEN, **MLLT4**, MLLT7 (So and Cleary 2003, So, Lin et al. 2003). Les autres groupes sont indiqués dans le **Tableau 1.5** à titre d'information.

Fonction	Chromosome	Partenaire de fusion	Fréquence en clinique
<b>Protéines nucléaires se fixant à l'ADN</b>	4q21	MLLT2/AF4	> 80%
	9q23	MLLT3/AF9	
	19p13	MLLT1/ENL	
	10p12	MLLT10/AF10	
	19p13	ELL	
<b>Protéines cytoplasmiques porteuses d'un domaine d'oligomérisation de type <i>coiled-coil</i></b>	1q32	MLLT5/EPS15	> 10%
	17p13	GAS7	
	19p13	EEN	
	6q27	MLLT4/AF6	
	Xq13	MLLT7/AFX	
<b>Protéines cytoplasmiques de la famille des septines. Interagit avec le cytosquelette et joue un rôle dans la mitose</b>	Xq22	SEPT2	> 1%
	22q11	SEPT5	
	Xq24	SEPT6	
	17q25	SEPT9	
	4q21	SEPT11	
<b>Protéines nucléaires de type histone acétyltransférase</b>	16q13	CBP	> 1%
	22q13	P300	
<b>Duplication partielle en tandem (PTD) des exons 5-12</b>	11q23	N/A	> 1%

**Tableau 1.5 – Classification des différents partenaires de fusion retrouvés dans les leucémies médiées par le gène *KMT2A*.**

Adaptée de la revue (Krivtsov and Armstrong 2007).

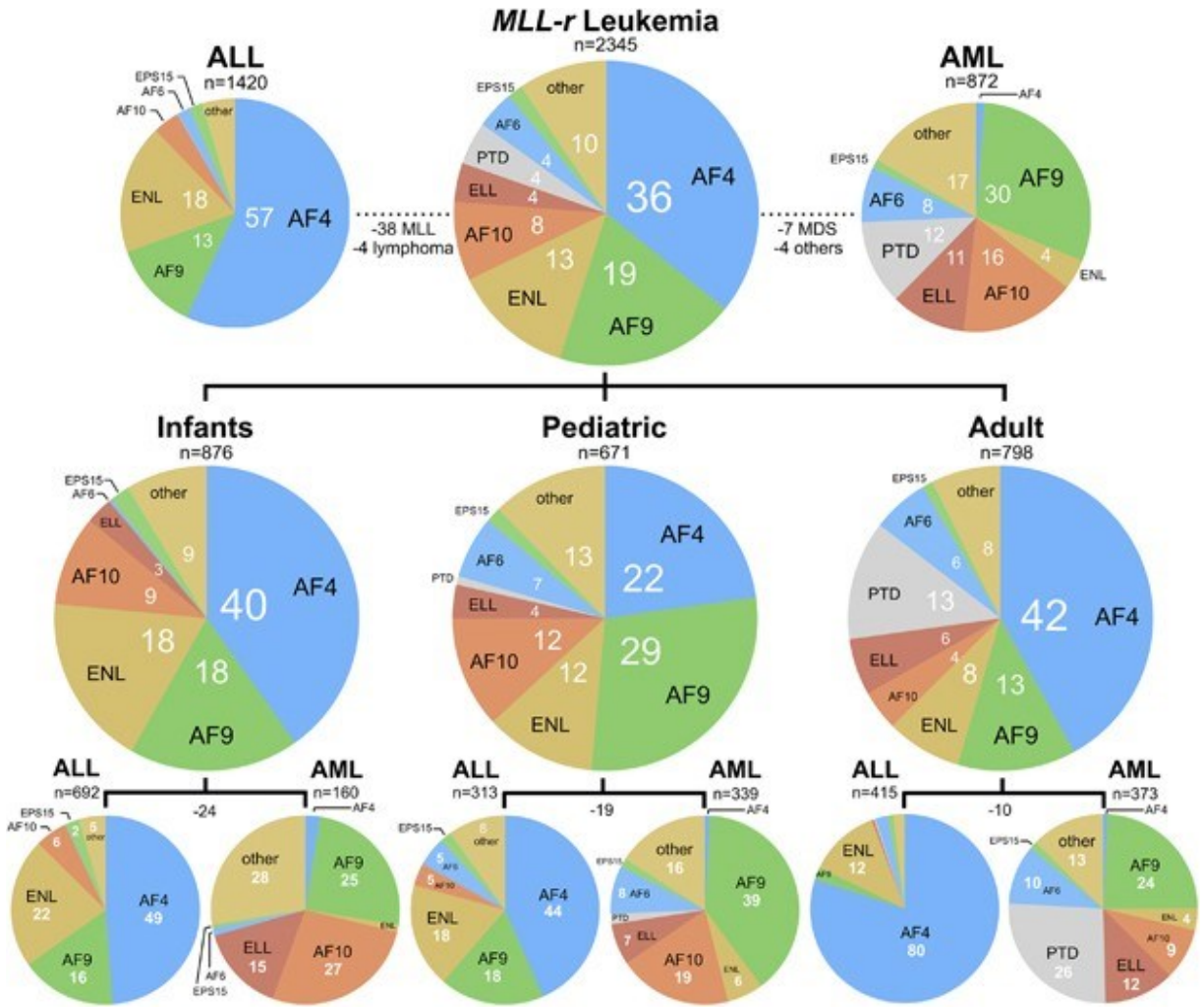
#### 1.5.2.2.2 Une fréquence variable en clinique

Les leucémies causées par une translocation chromosomique faisant intervenir le gène *KMT2A* représentent entre 5 et 10% de tous les types de leucémies aiguës, avec une forte proportion chez les patients pédiatriques (plus de 70% des patients âgés de moins de 12 mois) et sont à l'origine de la plupart des leucémies induites par un traitement de chimiothérapie. Ce type de fusion de gènes peut être responsable du développement de leucémie de type lymphoïde ou myéloïde. Un total de 135 réarrangements de ce type ont été identifiés en clinique. Quarante-deux d'entre eux ont été caractérisés et génèrent des protéines de fusion sans décalage du cadre de lecture. Selon une étude réalisée sur une cohorte de 2345 patients (Meyer, Burmeister et al. 2018), les neuf partenaires de fusion de *KMT2A* les plus fréquents (toutes leucémies confondues) sont *MLLT2/AF4* (~36%), *MLLT3/AF9* (~19%), *MLLT1/ENL* (~13%), *MLLT10/AF10* (~8%), *ELL* (~4%), *MLLT4/AF6* (~4%), *MLLT5/EPS15* (~2%) et *MLLT11/AF1Q* (~1%) (Figure 1.6).

#### 1.5.2.2.3 Partenaires de fusion et pronostics cliniques

Pour les patients pédiatriques non-porteurs de translocations de type *KMT2A*, la probabilité de survie s'élève à plus de 80%. À l'inverse, les enfants ayant un réarrangement chromosomique impliquant *KMT2A* voient leur chance de survie descendre à 30-50% (Chen, Sorensen et al. 1993, Bernt and Armstrong 2011). Cette observation souligne le besoin de décrire les mécanismes génétiques induites par ces fusions de gènes pour adapter et développer de nouvelles approches thérapeutiques.

D'autre part, le pronostic clinique varie significativement selon le type de partenaires de fusion. Par exemple, les translocations chromosomiques de type *KMT2A-MLLT11*, *KMT2A-MLLT3* et *KMT2A-MLLT4* sont respectivement associées à une probabilité de survie d'environ 92%, 50% et 11-25% (Balgobind, Raimondi et al. 2009). Ces données cliniques laissent à penser que les fusions oncogéniques font appel à des mécanismes différents de transformation pour activer des programmes génétiques particuliers qui mènent à des LMA plus ou moins agressives (Chen, Sun et al. 2016).

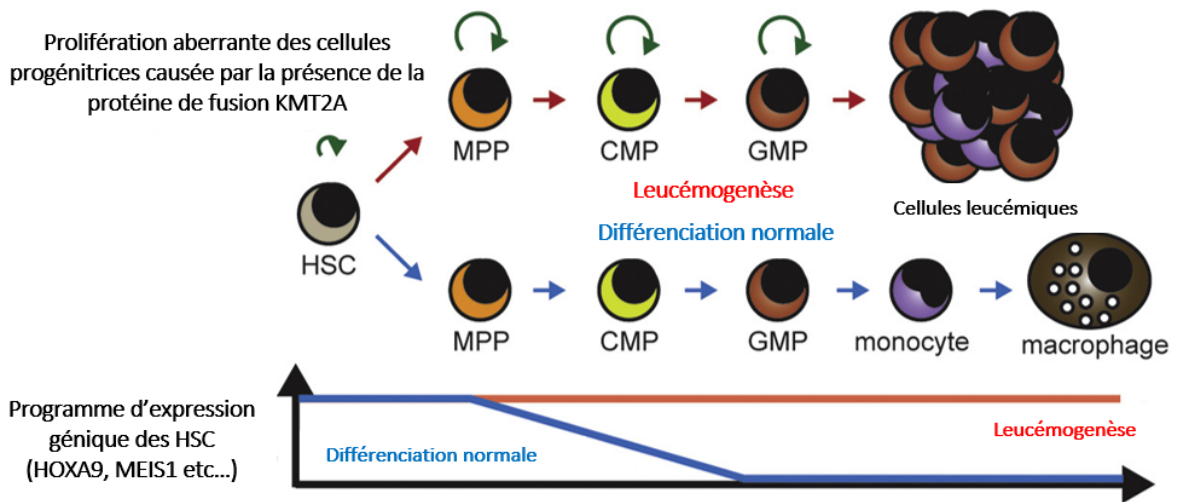


**Figure 1.6 – Classification des patients basée sur la classe d'âge et le type de leucémies.**

Les fréquences des partenaires de fusion retrouvés dans une cohorte de patients atteints d'une leucémie aiguë causée par une translocation chromosomique faisant intervenir le gène *KMT2A/MLL* (n=2345) sont indiquées en pourcentage. Cette cohorte peut être divisée en deux groupes de 1420 et 872 patients atteints respectivement d'une leucémie aiguë de type lymphoïde (ALL) ou myéloïde (AML). L'ensemble de cette cohorte peut être subdivisé selon la classe d'âge (enfants = nourrissons; pediatric = pédiatrique; adult = adulte). Les chiffres négatifs réfèrent au nombre de patients qui ne sont classifiés dans aucun des groupes de LMA ou LLA. Le graphique est tiré de l'étude (Meyer, Burmeister et al. 2018) et est reproduit avec leur permission.

#### 1.5.2.2.4 Un programme d'expression génique en commun

Les translocations chromosomiques médiées par *KMT2A* génèrent des fusions de gènes codant pour des protéines avec de nouvelles fonctions dérivées des protéines partenaires tout en conservant certaines fonctions de *KMT2A*. Par exemple, certains motifs sont conservés comme le motif CxxC et les domaines « AT-hook » qui sont essentiels pour permettre la transformation leucémique (Slany, Lavau et al. 1998). À l'inverse, le domaine SET à l'extrémité C-terminale (responsable de la méthylation de H3K4 sur la chromatine) n'est plus présent. Pourtant, les protéines de fusion sont toujours capables de réguler la transcription des gènes *HoxA* (*HoxA7*, *HoxA9* et *HoxA10*) et d'autres gènes comme *MEIS1* (Milne, Martin et al. 2005). Les fusions oncogéniques fonctionnent comme une machinerie transcriptionnelle constitutivement active, ce qui maintient l'expression de gènes dans les cellules progénitrices et les empêche de compléter leur processus de différenciation (**Figure 1.7**).



**Figure 1.7 – Programme d'expression génique des cellules souches hématopoïétiques (HSC) dans les cellules hématopoïétiques non-leucémiques (considérées comme normales) et les cellules leucémiques porteuses d'une translocation de type *KMT2A*.**

Pour rappel, MPP = Progéniteurs primitifs multipotents ; CMP = Progéniteurs communs des cellules myéloïdes ; GMP = Progéniteurs granulocytes/macrophages.

Figure adaptée de la revue (Takahashi and Yokoyama 2020) et reproduite avec leur permission.

Dans le cas des leucémies médiées par un réarrangement de type KMT2A, les études génomiques ont révélé une certaine stabilité avec seulement quelques modifications génétiques (Mullighan, Goorha et al. 2007, Radtke, Mullighan et al. 2009) : cela suggère que les leucémies causées par une fusion de type KMT2A sont principalement causées par une dérégulation au niveau épigénétique (Bernt and Armstrong 2011).

#### 1.5.2.2.5 Une mécanique qui varie selon les protéines partenaires

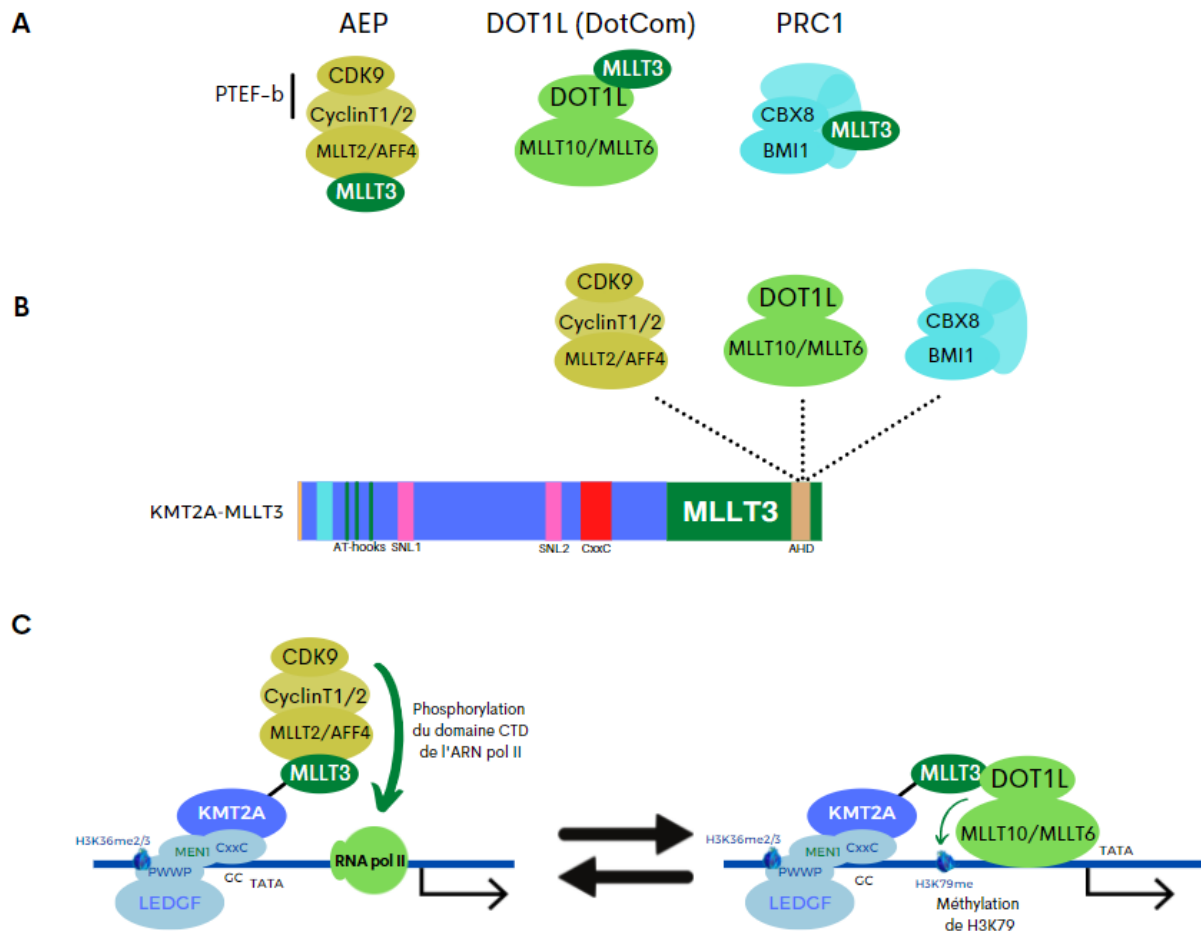
Pour quelle raison le pronostic vital chez les patients varie-t-il d'une fusion oncogénique à l'autre ? Ces aspects sont encore mal compris. Dans cette thèse, nous nous attardons sur deux translocations particulières : *KMT2A-MLLT3* (Figure 1.8) et *KMT2A-MLLT4* (Figure 1.9) qui fonctionnent de deux manières différentes.

##### 1.5.2.2.5.1 Le cas de la fusion *KMT2A-MLLT3*

Les protéines partenaires les plus fréquemment retrouvées se regroupent en deux familles de protéines : la famille MLLT2 regroupant les protéines MLLT2, AFF4, AFF3 et AFF2, et la famille MLLT1 composée de MLLT1 et MLLT3. Les membres de la famille MLLT2, les membres de la famille MLLT1 et le facteur d'élongation P-TEFb (*positive transcription elongation factor b*) forment un complexe multi-protéique appelé complexe AEP (Yokoyama, Lin et al. 2010). Ce dernier est impliqué dans divers processus biologiques en facilitant l'élongation transcriptionnelle, via P-TEFb (hétérodimère de CDK9 et cyclineT1/2) qui assure la phosphorylation du domaine C-terminal (CTD) de l'ARN polymérase II (Lin, Smith et al. 2010) (Figure 1.8).

Les protéines de fusion *KMT2A* forment une structure multiprotéique hybride avec le complexe AEP qui se positionne sur la chromatine cible (Yokoyama, Lin et al. 2010). *KMT2A-MLLT3* transforme les cellules hématopoïétiques notamment via le domaine AHD (« ANC1 homology domain ») de la protéine MLLT3. Il sert de site de fixation pour plusieurs protéines : i) de la famille MLLT2 ce qui permet le recrutement du complexe AEP (Okuda, Stanojevic et al. 2017) ; ii) DOT1L, responsable du dépôt de la marque activatrice H3K79me1/2/3 (Kuntimaddi, Achille et al. 2015) et essentielle dans la leucémogénèse médiée par *KMT2A-MLLT3* (Nguyen, Taranova et al. 2011) ; iii) le complexe PRC1 (complexe répressif polycomb 1) (Mueller, Bach et al. 2007). Ces associations semblent être exclusives ce qui laisse à penser que le domaine AHD pourrait échanger les protéines qui s'y fixent en fonction du contexte cellulaire (Biswas, Milne et al. 2011). Il semblerait que le complexe *KMT2A/AEP* déclenche l'activation transcriptionnelle, tandis que le complexe *KMT2A/DOT1L* maintient le profil d'expression génique, contribuant ainsi à l'expression du programme leucémique (Chen, Koche

et al. 2015). À noter que le rôle de l'association du complexe KMT2A/PRC1 n'est pas clairement établi.



**Figure 1.8 – Mécanismes moléculaires menant à l'activité oncogénique des protéines de fusion KMT2A-MLL3.**

(A) Détail des complexes protéiques capables d'interagir avec la protéine MLLT3 via son domaine AHD.

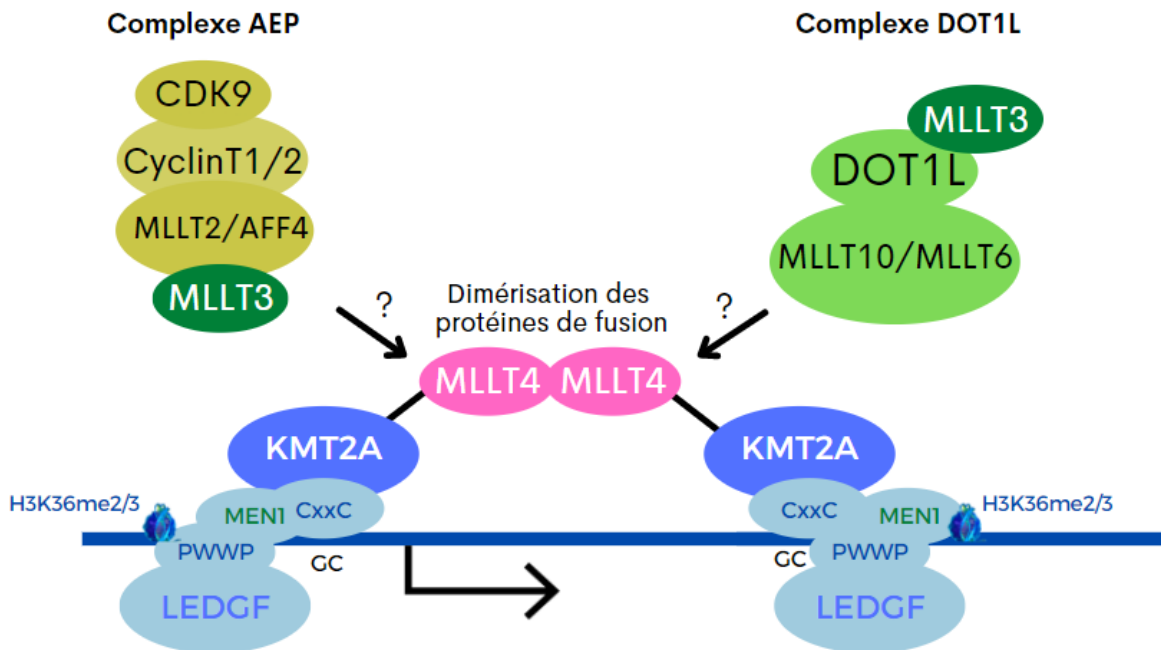
(B) Structure détaillée de la protéine KMT2A-MLL3. AHD = ANC1 homology domain.

(C) Modèle de formation des complexes (AEP, DOT1L) avec la protéine KMT2A-MLL3. Le complexe KMT2A/AEP permet la phosphorylation du domaine C-terminal de l'ARN polymérase II, tandis que le complexe KMT2A-DOT1L permet le maintien de l'activité transcriptionnelle.

Figure tirée et adaptée des revues (Takahashi and Yokoyama 2020) et (Krivtsov, Hoshii et al. 2017).

#### 1.5.2.2.5.2 Le cas de la fusion *KMT2A-MLLT4*

La particularité de la protéine partenaire MLLT4 est d'être située dans le cytoplasme. Son mécanisme d'action se base sur l'oligomérisation de la protéine de fusion et est essentiel pour permettre le développement d'une leucémie induite par la translocation *KMT2A-MLLT4*. MLLT4 ne faisant pas partie des complexes AEP et DOT1L, la protéine de fusion parvient à les recruter via un mécanisme encore méconnu (Yokoyama, Lin et al. 2010) (**Figure 1.9**). Nous en discuterons plus en détail dans le chapitre 3.



**Figure 1.9 – Mécanismes moléculaires menant à l'activité oncogénique des protéines de fusion KMT2A-MLLT4.**

La protéine semble recruter les complexes AEP et DOT1L mais par un mécanisme inconnu. Figure tirée et adaptée de la revue (Takahashi and Yokoyama 2020).

Pour information, les autres types de fusions oncogéniques utilisent différents mécanismes. Par exemple, alors que la protéine KMT2A-EEN interagit avec l'histone méthyltransférase de H4R3 pour modifier la chromatine et établir son programme leucémique, les protéines KMT2A-CBP et KMT2A-p300 agissent en tant qu'histone acétyltransférase (Slany 2016) sur la chromatine.

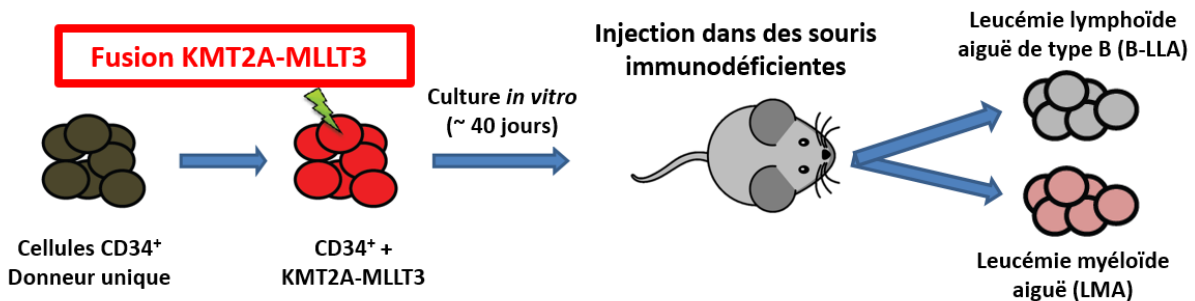


## 1.6 Hypothèse et objectifs de recherche

### 1.6.1 Problématique

L'impact biologique des translocations chromosomiques faisant intervenir le gène *KMT2A* est encore mal compris. L'objectif de cartographier les impacts de ces fusions de gènes se heurte à l'hétérogénéité génétique retrouvée dans le génome des patients ou des lignées cellulaires. En effet, il existe de nombreuses façons d'étudier les LMA (lignées cellulaires, modèles murins, xéno greffes dérivées de cellules de patients...) possédant chacune des avantages et des inconvénients qui sont détaillés dans le chapitre 4 à travers une revue de littérature. Cependant, les mutations secondaires récurrentes qui accompagnent la présence des aberrations chromosomiques compliquent la compréhension des conséquences directes de ces dernières.

Afin de contrer ces enjeux, un modèle leucémique humain chez la souris a été mis en place afin d'étudier spécifiquement les conséquences directes de la présence de fusions impliquant le gène *KMT2A* (Barabe, Kennedy et al. 2007). Le modèle utilisé consiste à induire de manière rétrovirale l'expression de la fusion *KMT2A-MLLT3* (*KM3*) dans des cellules souches hématopoïétiques et progénitrices ( $CD34^+$ ) d'un unique donneur sain. Ces cellules pré-leucémiques ( $CD34^+KMT2A-MLLT3$ ) sont ensuite injectées dans des souris immunodéficientes, qui développent une leucémie de type myéloïde ou lymphoïde après 24-30 semaines (Figure 1.10).



**Figure 1.10 – Représentation du modèle leucémique humain de type *KMT2A-MLLT3*.**

Des cellules sanguines  $CD34^+$  provenant d'un sang de cordon ombilical de donneur unique sont infectées par un rétrovirus portant la fusion oncogénique *KM3*. Ces cellules sont ensuite injectées dans une souris immunodéficiente pour produire une leucémie myéloïde aiguë ou une leucémie lymphoïde aiguë humaine après quelques semaines.

Nos précédentes analyses génomiques ont démontré que la fusion *KM3* seule est suffisante pour développer des leucémies (Barabe, Gil et al. 2017), sans qu'il soit nécessaire de recourir à des mutations secondaires récurrentes, comme cela peut être le cas chez les patients pédiatriques (Barabe, Kennedy et al. 2007). Notre modèle constitue donc un système expérimental pour étudier les mécanismes génétiques et épigénétiques impliqués dans la maladie et directement causés par la translocation chromosomique médiée par *KMT2A*. Il permet également d'avoir un aperçu unique des premiers stades de la transformation. Les avantages de ce modèle sont donc multiples: modéliser le phénotype et l'évolution de la maladie, analyser les différents événements génétiques et épigénétiques intervenant aux différentes étapes de la leucémogénèse et éliminer l'hétérogénéité génétique retrouvée chez les patients.

Les premières données préliminaires nous ont permis d'identifier de nouveaux biomarqueurs chez les patients porteurs de la fusion *KM3* (Barabe, Gil et al. 2017). Alors que  $\sim 1/3$  de ces gènes sont exprimés *in vitro* dans notre modèle, la majorité n'est exprimée qu'après injection dans les souris. Cette observation suggère que les cellules partiellement transformées par *KM3* nécessitent des signaux *in vivo* supplémentaires (par exemple à partir de la niche de la moelle osseuse) pour terminer leur transformation leucémique. Par ailleurs, *KM3* étant un régulateur épigénétique, il est probable que des modifications épigénétiques impactent la régulation de l'expression des gènes dans le modèle.

## 1.6.2 Hypothèse générale et objectifs

L'hypothèse générale du projet est la suivante : le modèle leucémique nous permet d'identifier des gènes régulés de manière épigénétique et qui pourraient être essentiels dans le processus de leucémogénèse médié par des fusions faisant intervenir le gène *KMT2A*.

Pour apporter des éléments de réponse, cette thèse est divisée en deux objectifs principaux :

- Objectif 1 : Caractériser les changements génétiques et épigénétiques à chacun des stades de la leucémogénèse causée par la fusion *KM3*.
- Objectif 2 : Déterminer les différences génétiques et épigénétiques entre les leucémies médiées par les fusions *KMT2A-MLLT3* (*KM3*) et *KMT2A-MLLT4* (*KM4*) pour comprendre pourquoi le pronostic vital diffère autant en clinique.

## 1.7 Références

- Allen, M. D., C. G. Grummitt, C. Hilcenko, S. Y. Min, L. M. Tonkin, C. M. Johnson, S. M. Freund, M. Bycroft and A. J. Warren (2006). "Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase." *EMBO J* **25**(19): 4503-4512.
- Allis, C. D. and T. Jenuwein (2016). "The molecular hallmarks of epigenetic control." *Nat Rev Genet* **17**(8): 487-500.
- Ang, Y. S., S. Y. Tsai, D. F. Lee, J. Monk, J. Su, K. Ratnakumar, J. Ding, Y. Ge, H. Darr, B. Chang, J. Wang, M. Rendl, E. Bernstein, C. Schaniel and I. R. Lemischka (2011). "Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network." *Cell* **145**(2): 183-197.
- Arber, D. A., A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. Le Beau, C. D. Bloomfield, M. Cazzola and J. W. Vardiman (2016). "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia." *Blood* **127**(20): 2391-2405.
- Balgobind, B. V., S. C. Raimondi, J. Harbott, M. Zimmermann, T. A. Alonzo, A. Auvrignon, H. B. Beverloo, M. Chang, U. Creutzig, M. N. Dworzak, E. Forestier, B. Gibson, H. Hasle, C. J. Harrison, N. A. Heerema, G. J. Kaspers, A. Leszl, N. Litvinko, L. L. Nigro, A. Morimoto, C. Perot, R. Pieters, D. Reinhardt, J. E. Rubnitz, F. O. Smith, J. Stary, I. Stasevich, S. Strehl, T. Taga, D. Tomizawa, D. Webb, Z. Zemanova, C. M. Zwaan and M. M. van den Heuvel-Eibrink (2009). "Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study." *Blood* **114**(12): 2489-2496.
- Ballabio, E. and T. A. Milne (2012). "Molecular and Epigenetic Mechanisms of MLL in Human Leukemogenesis." *Cancers (Basel)* **4**(3): 904-944.
- Barabe, F., L. Gil, M. Celton, A. Bergeron, V. Lamontagne, E. Roques, K. Lagace, A. Forest, R. Johnson, L. Pecheux, J. Simard, J. Pelloux, A. Bellemare-Pelletier, E. Gagnon, J. Hebert, S. Cellot and B. T. Wilhelm (2017). "Modeling human MLL-AF9 translocated acute myeloid leukemia from single donors reveals RET as a potential therapeutic target." *Leukemia* **31**(5): 1166-1176.
- Barabe, F., J. A. Kennedy, K. J. Hope and J. E. Dick (2007). "Modeling the initiation and progression of human acute leukemia in mice." *Science* **316**(5824): 600-604.
- Beck, S., F. Faradji, H. Brock and F. Peronnet (2010). "Maintenance of Hox gene expression patterns." *Adv Exp Med Biol* **689**: 41-62.
- Becker, A. M., D. G. Michael, A. T. Satpathy, R. Sciammas, H. Singh and D. Bhattacharya (2012). "IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors." *Blood* **119**(9): 2003-2012.
- Bennett, J. M., D. Catovsky, M. T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick and C. Sultan (1976). "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group." *Br J Haematol* **33**(4): 451-458.

Benveniste, P., C. Frelin, S. Janmohamed, M. Barbara, R. Herrington, D. Hyam and N. N. Iscove (2010). "Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential." Cell Stem Cell **6**(1): 48-58.

Berdasco, M. and M. Esteller (2010). "Aberrant epigenetic landscape in cancer: how cellular identity goes awry." Dev Cell **19**(5): 698-711.

Bernt, K. M. and S. A. Armstrong (2011). "Targeting epigenetic programs in MLL-rearranged leukemias." Hematology Am Soc Hematol Educ Program **2011**: 354-360.

Birke, M., S. Schreiner, M. P. Garcia-Cuellar, K. Mahr, F. Titgemeyer and R. K. Slany (2002). "The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation." Nucleic Acids Res **30**(4): 958-965.

Biswas, D., T. A. Milne, V. Basrur, J. Kim, K. S. Elenitoba-Johnson, C. D. Allis and R. G. Roeder (2011). "Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes." Proc Natl Acad Sci U S A **108**(38): 15751-15756.

Bolouri, H., J. E. Farrar, T. Triche, Jr., R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvil, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard and S. Meshinchi (2018). "The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions." Nat Med **24**(1): 103-112.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21 29 21-21 29 29.

Cancer Genome Atlas Research, N., T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing and G. Eley (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." N Engl J Med **368**(22): 2059-2074.

Caslini, C., Z. Yang, M. El-Osta, T. A. Milne, R. K. Slany and J. L. Hess (2007). "Interaction of MLL amino terminal sequences with menin is required for transformation." Cancer Res **67**(15): 7275-7283.

- Chan, A. K. N. and C. W. Chen (2019). "Rewiring the Epigenetic Networks in MLL-Rearranged Leukemias: Epigenetic Dysregulation and Pharmacological Interventions." Front Cell Dev Biol **7**: 81.
- Chen, C. S., P. H. Sorensen, P. H. Domer, G. H. Reaman, S. J. Korsmeyer, N. A. Heerema, G. D. Hammond and J. H. Kersey (1993). "Molecular rearrangements on chromosome 11q23 predominate in infant acute lymphoblastic leukemia and are associated with specific biologic variables and poor outcome." Blood **81**(9): 2386-2393.
- Chen, C. W., R. P. Koche, A. U. Sinha, A. J. Deshpande, N. Zhu, R. Eng, J. G. Doench, H. Xu, S. H. Chu, J. Qi, X. Wang, C. Delaney, K. M. Bernt, D. E. Root, W. C. Hahn, J. E. Bradner and S. A. Armstrong (2015). "DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia." Nat Med **21**(4): 335-343.
- Chen, L., Y. Sun, J. Wang, H. Jiang and A. G. Muntean (2016). "Differential regulation of the c-Myc/Lin28 axis discriminates subclasses of rearranged MLL leukemia." Oncotarget **7**(18): 25208-25223.
- Cheng, H., Z. Zheng and T. Cheng (2020). "New paradigms on hematopoietic stem cell differentiation." Protein Cell **11**(1): 34-44.
- Cierpicki, T., L. E. Risner, J. Grembecka, S. M. Lukasik, R. Popovic, M. Omonkowska, D. D. Shultis, N. J. Zeleznik-Le and J. H. Bushweller (2010). "Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia." Nat Struct Mol Biol **17**(1): 62-68.
- Crump, N. T. and T. A. Milne (2019). "Why are so many MLL lysine methyltransferases required for normal mammalian development?" Cell Mol Life Sci **76**(15): 2885-2898.
- Dahl, R., J. C. Walsh, D. Lancki, P. Laslo, S. R. Iyer, H. Singh and M. C. Simon (2003). "Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor." Nat Immunol **4**(10): 1029-1036.
- Djabali, M., L. Selleri, P. Parry, M. Bower, B. D. Young and G. A. Evans (1992). "A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias." Nat Genet **2**(2): 113-118.
- Dohner, H., E. H. Estey, S. Amadori, F. R. Appelbaum, T. Buchner, A. K. Burnett, H. Dombret, P. Fenau, D. Grimwade, R. A. Larson, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. A. Sanz, J. Sierra, M. S. Tallman, B. Lowenberg, C. D. Bloomfield and L. European (2010). "Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet." Blood **115**(3): 453-474.
- Dohner, H., D. J. Weisdorf and C. D. Bloomfield (2015). "Acute Myeloid Leukemia." N Engl J Med **373**(12): 1136-1152.
- Dou, Y., T. A. Milne, A. J. Ruthenburg, S. Lee, J. W. Lee, G. L. Verdine, C. D. Allis and R. G. Roeder (2006). "Regulation of MLL1 H3K4 methyltransferase activity by its core components." Nat Struct Mol Biol **13**(8): 713-719.

Dou, Y., T. A. Milne, A. J. Tackett, E. R. Smith, A. Fukuda, J. Wysocka, C. D. Allis, B. T. Chait, J. L. Hess and R. G. Roeder (2005). "Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF." *Cell* **121**(6): 873-885.

Doulatov, S., F. Notta, K. Eppert, L. T. Nguyen, P. S. Ohashi and J. E. Dick (2010). "Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development." *Nat Immunol* **11**(7): 585-593.

Duncan, I. and G. Montgomery (2002). "E. B. Lewis and the bithorax complex: part II. From cis-trans test to the genetic control of development." *Genetics* **161**(1): 1-10.

Emerling, B. M., J. Bonifas, C. P. Kratz, S. Donovan, B. R. Taylor, E. D. Green, M. M. Le Beau and K. M. Shannon (2002). "MLL5, a homolog of *Drosophila trithorax* located within a segment of chromosome band 7q22 implicated in myeloid leukemia." *Oncogene* **21**(31): 4849-4854.

Fair, K., M. Anderson, E. Bulanova, H. Mi, M. Tropschug and M. O. Diaz (2001). "Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells." *Mol Cell Biol* **21**(10): 3589-3597.

Fey, M. F. (2007). "Normal and malignant hematopoiesis." *Ann Oncol* **18 Suppl 1**: i9-i13.

FitzGerald, K. T. and M. O. Diaz (1999). "MLL2: A new mammalian member of the *trx*/MLL family of genes." *Genomics* **59**(2): 187-192.

Fossati, A., D. Dolfini, G. Donati and R. Mantovani (2011). "NF-Y recruits Ash2L to impart H3K4 trimethylation on CCAAT promoters." *PLoS One* **6**(3): e17220.

Froimchuk, E., Y. Jang and K. Ge (2017). "Histone H3 lysine 4 methyltransferase KMT2D." *Gene* **627**: 337-342.

Furey, T. S. (2012). "ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions." *Nat Rev Genet* **13**(12): 840-852.

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S. L. Ang and A. F. Stewart (2006). "Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development." *Development* **133**(8): 1423-1432.

Graf, T. and T. Enver (2009). "Forcing cells to change lineages." *Nature* **462**(7273): 587-594.

Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C. M. Croce and E. Canaani (1992). "The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila trithorax*, to the AF-4 gene." *Cell* **71**(4): 701-708.

Haas, S., A. Trumpp and M. D. Milsom (2018). "Causes and Consequences of Hematopoietic Stem Cell Heterogeneity." *Cell Stem Cell* **22**(5): 627-638.

Hanson, R. D., J. L. Hess, B. D. Yu, P. Ernst, M. van Lohuizen, A. Berns, N. M. van der Lugt, C. S. Shashikant, F. H. Ruddle, M. Seto and S. J. Korsmeyer (1999). "Mammalian *Trithorax* and polycomb-group homologues are antagonistic regulators of homeotic development." *Proc Natl Acad Sci U S A* **96**(25): 14372-14377.

Harris, R. A., T. Wang, C. Coarfa, R. P. Nagarajan, C. Hong, S. L. Downey, B. E. Johnson, S. D. Fouse, A. Delaney, Y. Zhao, A. Olshen, T. Ballinger, X. Zhou, K. J. Forsberg, J. Gu, L. Echipare, H. O'Geen, R. Lister, M. Pelizzola, Y. Xi, C. B. Epstein, B. E. Bernstein, R. D.

- Hawkins, B. Ren, W. Y. Chung, H. Gu, C. Bock, A. Gnirke, M. Q. Zhang, D. Haussler, J. R. Ecker, W. Li, P. J. Farnham, R. A. Waterland, A. Meissner, M. A. Marra, M. Hirst, A. Milosavljevic and J. F. Costello (2010). "Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications." Nat Biotechnol **28**(10): 1097-1105.
- Hom, R. A., P. Y. Chang, S. Roy, C. A. Musselman, K. C. Glass, A. I. Selezneva, O. Gozani, R. F. Ismagilov, M. L. Cleary and T. G. Kutateladze (2010). "Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain." J Mol Biol **400**(2): 145-154.
- Hsieh, J. J., P. Ernst, H. Erdjument-Bromage, P. Tempst and S. J. Korsmeyer (2003). "Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization." Mol Cell Biol **23**(1): 186-194.
- Huang, J., B. Gurung, B. Wan, S. Matkar, N. A. Veniaminova, K. Wan, J. L. Merchant, X. Hua and M. Lei (2012). "The same pocket in menin binds both MLL and JUND but has opposite effects on transcription." Nature **482**(7386): 542-546.
- Hughes, C. M., O. Rozenblatt-Rosen, T. A. Milne, T. D. Copeland, S. S. Levine, J. C. Lee, D. N. Hayes, K. S. Shanmugam, A. Bhattacharjee, C. A. Biondi, G. F. Kay, N. K. Hayward, J. L. Hess and M. Meyerson (2004). "Menin associates with a trithorax family histone methyltransferase complex and with the *hoxc8* locus." Mol Cell **13**(4): 587-597.
- Huntsman, D. G., S. F. Chin, M. Muleris, S. J. Batley, V. P. Collins, L. M. Wiedemann, S. Aparicio and C. Caldas (1999). "MLL2, the second human homolog of the *Drosophila* trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines." Oncogene **18**(56): 7975-7984.
- Iwasaki, H., S. Mizuno, R. A. Wells, A. B. Cantor, S. Watanabe and K. Akashi (2003). "GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages." Immunity **19**(3): 451-462.
- Jagannathan-Bogdan, M. and L. I. Zon (2013). "Hematopoiesis." Development **140**(12): 2463-2467.
- Jude, C. D., L. Climer, D. Xu, E. Artinger, J. K. Fisher and P. Ernst (2007). "Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors." Cell Stem Cell **1**(3): 324-337.
- Kassis, J. A., J. A. Kennison and J. W. Tamkun (2017). "Polycomb and Trithorax Group Genes in *Drosophila*." Genetics **206**(4): 1699-1725.
- Kaushansky, K. (2006). "Lineage-specific hematopoietic growth factors." N Engl J Med **354**(19): 2034-2045.
- Kawabe, Y., Y. X. Wang, I. W. McKinnell, M. T. Bedford and M. A. Rudnicki (2012). "Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions." Cell Stem Cell **11**(3): 333-345.
- Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
- Krivtsov, A. V. and S. A. Armstrong (2007). "MLL translocations, histone modifications and leukaemia stem-cell development." Nat Rev Cancer **7**(11): 823-833.

Krivtsov, A. V., T. Hoshii and S. A. Armstrong (2017). "Mixed-Lineage Leukemia Fusions and Chromatin in Leukemia." Cold Spring Harb Perspect Med 7(11).

Kuntimaddi, A., N. J. Achille, J. Thorpe, A. A. Lokken, R. Singh, C. S. Hemenway, M. Adli, N. J. Zeleznik-Le and J. H. Bushweller (2015). "Degree of recruitment of DOT1L to MLL-AF9 defines level of H3K79 Di- and tri-methylation on target genes and transformation potential." Cell Rep 11(5): 808-820.

Kurotaki, D., M. Yamamoto, A. Nishiyama, K. Uno, T. Ban, M. Ichino, H. Sasaki, S. Matsunaga, M. Yoshinari, A. Ryo, M. Nakazawa, K. Ozato and T. Tamura (2014). "IRF8 inhibits C/EBPalpha activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils." Nat Commun 5: 4978.

Laiosa, C. V., M. Stadtfeld, H. Xie, L. de Andres-Aguayo and T. Graf (2006). "Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors." Immunity 25(5): 731-744.

Lee, J. H., C. M. Tate, J. S. You and D. G. Skalnik (2007). "Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex." J Biol Chem 282(18): 13419-13428.

Lewis, E. B. (1978). "A gene complex controlling segmentation in Drosophila." Nature 276(5688): 565-570.

Lieber, M. R., J. Gu, H. Lu, N. Shimazaki and A. G. Tsai (2010). "Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans." Subcell Biochem 50: 279-296.

Lim, D. A., Y. C. Huang, T. Swigut, A. L. Mirick, J. M. Garcia-Verdugo, J. Wysocka, P. Ernst and A. Alvarez-Buylla (2009). "Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells." Nature 458(7237): 529-533.

Lin, C., E. R. Smith, H. Takahashi, K. C. Lai, S. Martin-Brown, L. Florens, M. P. Washburn, J. W. Conaway, R. C. Conaway and A. Shilatifard (2010). "AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia." Mol Cell 37(3): 429-437.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution." Nature 389(6648): 251-260.

Mazo, A. M., D. H. Huang, B. A. Mozer and I. B. Dawid (1990). "The trithorax gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains." Proc Natl Acad Sci U S A 87(6): 2112-2116.

McMahon, K. A., S. Y. Hiew, S. Hadjur, H. Veiga-Fernandes, U. Menzel, A. J. Price, D. Kioussis, O. Williams and H. J. Brady (2007). "Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal." Cell Stem Cell 1(3): 338-345.

Mendez-Ferrer, S., D. Bonnet, D. P. Steensma, R. P. Hasserjian, I. M. Ghobrial, J. G. Gribben, M. Andreeff and D. S. Krause (2020). "Bone marrow niches in haematological malignancies." Nat Rev Cancer 20(5): 285-298.

Metafuni, E., P. Chiusolo, L. Laurenti, F. Sora, S. Giammarco, A. Bacigalupo, G. Leone and S. Sica (2018). "Allogeneic Hematopoietic Stem Cell Transplantation In Therapy-Related Myeloid



Neoplasms (t-MN) of the Adult: Monocentric Observational Study and Review of the Literature." Mediterr J Hematol Infect Dis **10**(1): e2018005.

Metcalf, D. (1998). "Regulatory mechanisms controlling hematopoiesis: principles and problems." Stem Cells **16 Suppl 1**: 3-11.

Meyer, C., T. Burmeister, D. Groger, G. Tsaur, L. Fechina, A. Renneville, R. Sutton, N. C. Venn, M. Emerenciano, M. S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal, V. H. J. van der Velden, E. Delabesse, T. S. Park, S. H. Oh, M. L. M. Silva, T. Lund-Aho, V. Juvonen, A. S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenkova, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. Zur Stadt, J. Lentjes, G. Gohring, A. Kustanovich, O. Aleinikova, B. W. Schafer, S. Kubetzko, H. O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J. M. Cayuela, E. Clappier, C. N. Alonso, C. M. Zwaan, M. M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Moricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O. A. Haas, R. Panzer-Grumayer, L. Sedek, T. Szczepanski, A. Caye, L. Suarez, H. Cave and R. Marschalek (2018). "The MLL recombinome of acute leukemias in 2017." Leukemia **32**(2): 273-284.

Milne, T. A., S. D. Briggs, H. W. Brock, M. E. Martin, D. Gibbs, C. D. Allis and J. L. Hess (2002). "MLL targets SET domain methyltransferase activity to Hox gene promoters." Mol Cell **10**(5): 1107-1117.

Milne, T. A., Y. Dou, M. E. Martin, H. W. Brock, R. G. Roeder and J. L. Hess (2005). "MLL associates specifically with a subset of transcriptionally active target genes." Proc Natl Acad Sci U S A **102**(41): 14765-14770.

Milne, T. A., C. M. Hughes, R. Lloyd, Z. Yang, O. Rozenblatt-Rosen, Y. Dou, R. W. Schnepf, C. Krankel, V. A. Livolsi, D. Gibbs, X. Hua, R. G. Roeder, M. Meyerson and J. L. Hess (2005). "Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors." Proc Natl Acad Sci U S A **102**(3): 749-754.

Milne, T. A., J. Kim, G. G. Wang, S. C. Stadler, V. Basrur, S. J. Whitcomb, Z. Wang, A. J. Ruthenburg, K. S. Elenitoba-Johnson, R. G. Roeder and C. D. Allis (2010). "Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis." Mol Cell **38**(6): 853-863.

Milne, T. A., M. E. Martin, H. W. Brock, R. K. Slany and J. L. Hess (2005). "Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications." Cancer Res **65**(24): 11367-11374.

Mishra, B. P., K. M. Zaffuto, E. L. Artinger, T. Org, H. K. Mikkola, C. Cheng, M. Djabali and P. Ernst (2014). "The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis." Cell Rep **7**(4): 1239-1247.

Moore, L. D., T. Le and G. Fan (2013). "DNA methylation and its basic function." Neuropsychopharmacology **38**(1): 23-38.

Morita, K., F. Wang, K. Jahn, T. Hu, T. Tanaka, Y. Sasaki, J. Kuipers, S. Loghavi, S. A. Wang, Y. Yan, K. Furudate, J. Matthews, L. Little, C. Gumbs, J. Zhang, X. Song, E. Thompson, K. P. Patel, C. E. Bueso-Ramos, C. D. DiNardo, F. Ravandi, E. Jabbour, M. Andreeff, J. Cortes, K.

Bhalla, G. Garcia-Manero, H. Kantarjian, M. Konopleva, D. Nakada, N. Navin, N. Beerenwinkel, P. A. Futreal and K. Takahashi (2020). "Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics." Nat Commun **11**(1): 5327.

Mueller, D., C. Bach, D. Zeisig, M. P. Garcia-Cuellar, S. Monroe, A. Sreekumar, R. Zhou, A. Nesvizhskii, A. Chinnaiyan, J. L. Hess and R. K. Slany (2007). "A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification." Blood **110**(13): 4445-4454.

Mullighan, C. G., S. Goorha, I. Radtke, C. B. Miller, E. Coustan-Smith, J. D. Dalton, K. Girtman, S. Mathew, J. Ma, S. B. Pounds, X. Su, C. H. Pui, M. V. Relling, W. E. Evans, S. A. Shurtleff and J. R. Downing (2007). "Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia." Nature **446**(7137): 758-764.

Muntean, A. G., J. Tan, K. Sitwala, Y. Huang, J. Bronstein, J. A. Connelly, V. Basrur, K. S. Elenitoba-Johnson and J. L. Hess (2010). "The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis." Cancer Cell **17**(6): 609-621.

Murre, C. (2007). "Defining the pathways of early adult hematopoiesis." Cell Stem Cell **1**(4): 357-358.

Nakamura, T., T. Mori, S. Tada, W. Krajewski, T. Rozovskaia, R. Wassell, G. Dubois, A. Mazo, C. M. Croce and E. Canaani (2002). "ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation." Mol Cell **10**(5): 1119-1128.

Nguyen, A. T., O. Taranova, J. He and Y. Zhang (2011). "DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis." Blood **117**(25): 6912-6922.

Notta, F., S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J. McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror and J. E. Dick (2016). "Distinct routes of lineage development reshape the human blood hierarchy across ontogeny." Science **351**(6269): aab2116.

Okuda, H., B. Stanojevic, A. Kanai, T. Kawamura, S. Takahashi, H. Matsui, A. Takaori-Kondo and A. Yokoyama (2017). "Cooperative gene activation by AF4 and DOT1L drives MLL-rearranged leukemia." J Clin Invest **127**(5): 1918-1931.

Onodera, A., M. Yamashita, Y. Endo, M. Kuwahara, S. Tofukuji, H. Hosokawa, A. Kanai, Y. Suzuki and T. Nakayama (2010). "STAT6-mediated displacement of polycomb by trithorax complex establishes long-term maintenance of GATA3 expression in T helper type 2 cells." J Exp Med **207**(11): 2493-2506.

Orkin, S. H. and L. I. Zon (2008). "Hematopoiesis: an evolving paradigm for stem cell biology." Cell **132**(4): 631-644.

Osawa, M., K. Hanada, H. Hamada and H. Nakauchi (1996). "Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell." Science **273**(5272): 242-245.

- Park, S., U. Osmer, G. Raman, R. H. Schwantes, M. O. Diaz and J. H. Bushweller (2010). "The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression." Biochemistry **49**(31): 6576-6586.
- Patel, A., V. Dharmarajan, V. E. Vought and M. S. Cosgrove (2009). "On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex." J Biol Chem **284**(36): 24242-24256.
- Paul, F., Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay and I. Amit (2015). "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors." Cell **163**(7): 1663-1677.
- Pinho, S. and P. S. Frenette (2019). "Haematopoietic stem cell activity and interactions with the niche." Nat Rev Mol Cell Biol **20**(5): 303-320.
- Prasad, R., A. B. Zhadanov, Y. Sedkov, F. Bullrich, T. Druck, R. Rallapalli, T. Yano, H. Alder, C. M. Croce, K. Huebner, A. Mazo and E. Canaani (1997). "Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to *Drosophila trithorax*." Oncogene **15**(5): 549-560.
- Radtke, I., C. G. Mullighan, M. Ishii, X. Su, J. Cheng, J. Ma, R. Ganti, Z. Cai, S. Goorha, S. B. Pounds, X. Cao, C. Obert, J. Armstrong, J. Zhang, G. Song, R. C. Ribeiro, J. E. Rubnitz, S. C. Raimondi, S. A. Shurtleff and J. R. Downing (2009). "Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia." Proc Natl Acad Sci U S A **106**(31): 12944-12949.
- Rao, R. C. and Y. Dou (2015). "Hijacked in cancer: the KMT2 (MLL) family of methyltransferases." Nat Rev Cancer **15**(6): 334-346.
- Rieger, M. A. and T. Schroeder (2012). "Hematopoiesis." Cold Spring Harb Perspect Biol **4**(12).
- Robb, L. (2007). "Cytokine receptors and hematopoietic differentiation." Oncogene **26**(47): 6715-6723.
- Rosler, T. and R. Marschalek (2013). "An alternative splice process renders the MLL protein either into a transcriptional activator or repressor." Pharmazie **68**(7): 601-607.
- Roug, A. S., M. C. Hansen, L. Norderby and P. Hokland (2014). "Diagnosing and following adult patients with acute myeloid leukaemia in the genomic age." Br J Haematol **167**(2): 162-176.
- Rowley, J. D. (1993). "Rearrangements involving chromosome band 11Q23 in acute leukaemia." Semin Cancer Biol **4**(6): 377-385.
- Ruault, M., M. E. Brun, M. Ventura, G. Roizes and A. De Sario (2002). "MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia." Gene **284**(1-2): 73-81.
- Ruthenburg, A. J., C. D. Allis and J. Wysocka (2007). "Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark." Mol Cell **25**(1): 15-30.
- Saultz, J. N. and R. Garzon (2016). "Acute Myeloid Leukemia: A Concise Review." J Clin Med **5**(3).

- Sedkov, Y., J. J. Benes, J. R. Berger, K. M. Riker, S. Tillib, R. S. Jones and A. Mazo (1999). "Molecular genetic analysis of the *Drosophila* trithorax-related gene which encodes a novel SET domain protein." *Mech Dev* **82**(1-2): 171-179.
- Shilatifard, A. (2012). "The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis." *Annu Rev Biochem* **81**: 65-95.
- Slany, R. K. (2016). "The molecular mechanics of mixed lineage leukemia." *Oncogene* **35**(40): 5215-5223.
- Slany, R. K., C. Lavau and M. L. Cleary (1998). "The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX." *Mol Cell Biol* **18**(1): 122-129.
- So, C. W. and M. L. Cleary (2003). "Common mechanism for oncogenic activation of MLL by forkhead family proteins." *Blood* **101**(2): 633-639.
- So, C. W., M. Lin, P. M. Ayton, E. H. Chen and M. L. Cleary (2003). "Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias." *Cancer Cell* **4**(2): 99-110.
- Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." *Nature* **403**(6765): 41-45.
- Strissel, P. L., R. Strick, J. D. Rowley and N. J. Zeleznik-Le (1998). "An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region." *Blood* **92**(10): 3793-3803.
- Struhl, G. and M. Akam (1985). "Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*." *EMBO J* **4**(12): 3259-3264.
- Takahashi, S. and A. Yokoyama (2020). "The molecular functions of common and atypical MLL fusion protein complexes." *Biochim Biophys Acta Gene Regul Mech* **1863**(7): 194548.
- Tan, Y. C. and V. T. Chow (2001). "Novel human HALR (MLL3) gene encodes a protein homologous to ALR and to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated with leukemia and developmental defects." *Cancer Detect Prev* **25**(5): 454-469.
- Tavor, S., P. T. Vuong, D. J. Park, A. F. Gombart, A. H. Cohen and H. P. Koeffler (2002). "Macrophage functional maturation and cytokine production are impaired in C/EBP epsilon-deficient mice." *Blood* **99**(5): 1794-1801.
- Terranova, R., H. Agherbi, A. Boned, S. Meresse and M. Djabali (2006). "Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll." *Proc Natl Acad Sci U S A* **103**(17): 6629-6634.
- Thiel, A. T., P. Blessington, T. Zou, D. Feather, X. Wu, J. Yan, H. Zhang, Z. Liu, P. Ernst, G. A. Koretzky and X. Hua (2010). "MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele." *Cancer Cell* **17**(2): 148-159.
- Tkachuk, D. C., S. Kohler and M. L. Cleary (1992). "Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias." *Cell* **71**(4): 691-700.

Tyagi, S., A. L. Chabes, J. Wysocka and W. Herr (2007). "E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases." Mol Cell **27**(1): 107-119.

Van Holde, K. E., J. R. Allen, K. Tatchell, W. O. Weischet and D. Lohr (1980). "DNA-histone interactions in nucleosomes." Biophys J **32**(1): 271-282.

Verbeek, W., M. Wachter, J. Lekstrom-Himes and H. P. Koeffler (2001). "C/EBPepsilon  $\pm$  mice: increased rate of myeloid proliferation and apoptosis." Leukemia **15**(1): 103-111.

Wang, J., A. G. Muntean and J. L. Hess (2012). "ECSASB2 mediates MLL degradation during hematopoietic differentiation." Blood **119**(5): 1151-1161.

Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms and H. Y. Chang (2011). "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." Nature **472**(7341): 120-124.

Wang, P., C. Lin, E. R. Smith, H. Guo, B. W. Sanderson, M. Wu, M. Gogol, T. Alexander, C. Seidel, L. M. Wiedemann, K. Ge, R. Krumlauf and A. Shilatifard (2009). "Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II." Mol Cell Biol **29**(22): 6074-6085.

Wang, Z., J. Song, T. A. Milne, G. G. Wang, H. Li, C. D. Allis and D. J. Patel (2010). "Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression." Cell **141**(7): 1183-1194.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang and K. Zhao (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Wu, L., S. Y. Lee, B. Zhou, U. T. Nguyen, T. W. Muir, S. Tan and Y. Dou (2013). "ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes." Mol Cell **49**(6): 1108-1120.

Wysocka, J., M. P. Myers, C. D. Laherty, R. N. Eisenman and W. Herr (2003). "Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1." Genes Dev **17**(7): 896-911.

Xia, Z. B., M. Anderson, M. O. Diaz and N. J. Zeleznik-Le (2003). "MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein." Proc Natl Acad Sci U S A **100**(14): 8342-8347.

Xu, C., C. Bian, R. Lam, A. Dong and J. Min (2011). "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain." Nat Commun **2**: 227.

Yagi, H., K. Deguchi, A. Aono, Y. Tani, T. Kishimoto and T. Komori (1998). "Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice." Blood **92**(1): 108-117.

Yamamoto, R., Y. Morita, J. Oohara, S. Hamanaka, M. Onodera, K. L. Rudolph, H. Ema and H. Nakauchi (2013). "Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells." Cell **154**(5): 1112-1126.

- Yang, Y. W., R. A. Flynn, Y. Chen, K. Qu, B. Wan, K. C. Wang, M. Lei and H. Y. Chang (2014). "Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency." Elife **3**: e02046.
- Yokoyama, A. and M. L. Cleary (2008). "Menin critically links MLL proteins with LEDGF on cancer-associated target genes." Cancer Cell **14**(1): 36-46.
- Yokoyama, A., I. Kitabayashi, P. M. Ayton, M. L. Cleary and M. Ohki (2002). "Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties." Blood **100**(10): 3710-3718.
- Yokoyama, A., M. Lin, A. Naresh, I. Kitabayashi and M. L. Cleary (2010). "A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription." Cancer Cell **17**(2): 198-212.
- Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D. J. Aufiero, I. Kitabayashi, W. Herr and M. L. Cleary (2004). "Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression." Mol Cell Biol **24**(13): 5639-5649.
- Yu, B. D., R. D. Hanson, J. L. Hess, S. E. Horning and S. J. Korsmeyer (1998). "MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis." Proc Natl Acad Sci U S A **95**(18): 10632-10636.
- Yu, B. D., J. L. Hess, S. E. Horning, G. A. Brown and S. J. Korsmeyer (1995). "Altered Hox expression and segmental identity in Mll-mutant mice." Nature **378**(6556): 505-508.
- Zelevnik-Le, N. J., A. M. Harden and J. D. Rowley (1994). "11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene." Proc Natl Acad Sci U S A **91**(22): 10610-10614.
- Zhang, D. E., P. Zhang, N. D. Wang, C. J. Hetherington, G. J. Darlington and D. G. Tenen (1997). "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice." Proc Natl Acad Sci U S A **94**(2): 569-574.
- Zhang, X., W. Novera, Y. Zhang and L. W. Deng (2017). "MLL5 (KMT2E): structure, function, and clinical relevance." Cell Mol Life Sci **74**(13): 2333-2344.
- Ziemin-van der Poel, S., N. R. McCabe, H. J. Gill, R. Espinosa, 3rd, Y. Patel, A. Harden, P. Rubinelli, S. D. Smith, M. M. LeBeau, J. D. Rowley and et al. (1991). "Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias." Proc Natl Acad Sci U S A **88**(23): 10735-10739.

# **CHAPITRE 2**

## **ÉTUDE DES CHANGEMENTS ÉPIGÉNÉTIQUES DANS LE MODÈLE DE LEUCÉMIES HUMAINES DE TYPE *KMT2A-MLLT3***

## 2 Étude des changements épigénétiques dans le modèle de leucémies humaines de type *KMT2A-MLLT3*

### 2.1 Préface

Le but de l'étude est de décrire les événements initiaux génétiques et épigénétiques impliqués dans la leucémogénèse causée par la translocation chromosomique de type *KM3*, en utilisant le modèle de leucémie humaine décrit dans le chapitre 1. Nous avons étudié les changements dans la méthylation de l'ADN, les modifications des histones et l'accessibilité de la chromatine à chaque étape de notre système modèle et les avons corrélés avec les changements d'expression. Nous avons observé un profil d'hypométhylation prononcé dans les premiers stades de la transformation leucémique après l'ajout de la fusion oncogénique, ainsi qu'une perte d'expression des gènes associés aux cellules souches. Les données d'ATAC-seq ont montré qu'il y avait relativement peu de changements spécifiques à la leucémie et que la grande majorité correspondait à des régions de chromatine ouvertes et à des *clusters* de facteurs de transcription précédemment observés dans d'autres types de cellules sanguines. L'intégration de l'expression génique et des changements épigénétiques a révélé que le gène codant pour l'adénylate cyclase *ADCY9* est essentiel dans la leucémie myéloïde aiguë médiée par la translocation chromosomique de type *KM3*. Par ailleurs, les données suggèrent également un potentiel rôle du récepteur *CCR1* et son ligand spécifique *CCL23*. Nous avons ainsi découvert que la fusion *KM3* induit des changements minimes dans l'épigénome et profite de la machinerie transcriptionnelle 'normale' pour mener la leucémogénèse.

### 2.2 Contributions

J'ai effectué le design expérimental, analysé les données et rédigé le manuscrit conjointement avec Brian T Wilhelm. J'ai effectué les expériences de ChIP-seq et d'ATAC-seq, analysé les données et généré les figures, avec l'aide de Karine Lagacé. Magalie Celton a réalisé l'expérience de Methyl-seq et a participé à l'écriture du manuscrit. Safia Safa-Tahar-Henni a aidé à la création de figures issues de l'analyse d'ATAC-seq. Karine Lagacé, Élodie Roques, Eva Bresson et Anne Bergeron ont contribué aux expériences de biologie moléculaire et de culture cellulaire. Eva Bresson, Anne Bergeron et Frédéric Barabé ont généré le modèle



leucémique. Soheil Meshinchi a permis l'accès aux données cliniques des échantillons de patients. Josée Hebert et Sonia Cellot ont caractérisé et mis en banque les échantillons de patients utilisés dans l'étude. Tous les auteurs ont discuté des résultats, de la structure de la publication et ont révisé le manuscrit.

### **2.3 Référence de la publication**

Thomas Milan, Magalie Celton, Karine Lagacé, Élodie Roques, Safia Safa-Tahar-Henni, Eva Bresson, Anne Bergeron, Josée Hebert, Soheil Meshinchi, Sonia Cellot, Frédéric Barabé, Brian T Wilhelm. "Epigenetic changes in human model KMT2A leukemias highlight early events during leukemogenesis." *Haematologica* (2020). PMID: 33375773.

## 2.4 Article

### **Epigenetic changes in human model KMT2A leukemias highlight early events during leukemogenesis**

**Thomas Milan**<sup>1</sup>, Magalie Celton<sup>1</sup>, Karine Lagacé<sup>1</sup>, Élodie Roques<sup>1</sup>, Safia Safa-Tahar-Henni<sup>1</sup>, Eva Bresson<sup>2,3,4</sup>, Anne Bergeron<sup>2,3,4</sup>, Josée Hebert<sup>5,6</sup>, Soheil Meshinchi<sup>7</sup>, Sonia Cellot<sup>8</sup>, Frédéric Barabé<sup>2,3,4</sup>, Brian T Wilhelm<sup>1,6</sup>

<sup>1</sup>Laboratory for High Throughput Biology, Institute for Research in Immunology and Cancer, Montréal, QC, Canada

<sup>2</sup>Centre de recherche en infectiologie du CHUL, Centre de recherche du CHU de Québec – Université Laval, Québec City, QC, Canada

<sup>3</sup>CHU de Québec – Université Laval – Hôpital Enfant-Jésus; Québec City, QC, Canada;

<sup>4</sup>Department of Medicine, Université Laval, Québec City, QC, Canada

<sup>5</sup>Division of Hematology-Oncology and Leukemia Cell Bank of Quebec, Maisonneuve-Rosemont Hospital, Montréal, QC, Canada

<sup>6</sup>Department of Medicine, Université de Montréal, Montréal, QC, Canada

<sup>7</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

<sup>8</sup>Department of pediatrics, Division of Hematology, Ste-Justine Hospital, Montréal, QC, Canada

\*Correspondence to:

Brian T. Wilhelm

Institute for Research in Immunology and Cancer

Université de Montréal

P.O. Box 6128, Station Centre-Ville

Montreal, Québec, Canada H3C 3J7

Phone : +1 514 343.6111 ext 0923

Fax : +1 514 343.7780

E-mail : [brian.wilhelm@umontreal.ca](mailto:brian.wilhelm@umontreal.ca)

### 2.4.1 Abstract

Chromosomal translocations involving the *KMT2A* gene are among the most common genetic alterations found in pediatric acute myeloid leukemias although the molecular mechanisms that initiate the disease remain incompletely defined. To elucidate these initiating events we used a human model system of acute myeloid leukemia driven by the *KMT2A-MLLT3* (KM3) fusion. More specifically, we investigated changes in DNA methylation, histone modifications, and chromatin accessibility at each stage of our model system and correlated these with expression changes. We observed the development of a pronounced hypomethylation phenotype in the early stages of leukemic transformation after KM3 addition along with loss of expression of stem-cell-associated genes and skewed expression of other genes, such as *SI00A8/9*, implicated in leukemogenesis. In addition, early increases in the expression of the lysine demethylase KDM4B was functionally linked to these expression changes as well as other key transcription factors. Remarkably, our ATAC-sequencing data showed that there were relatively few leukemia-specific changes and that the vast majority corresponded to open chromatin regions and transcription factor clusters previously observed in other cell types. Integration of the gene expression and epigenetic changes revealed that the adenylate cyclase gene *ADCY9* is an essential gene in KM3-acute myeloid leukemia, and suggested the potential for autocrine signaling through the chemokine receptor CCR1 and CCL23 ligand. Collectively, our results suggest that KM3 induces minimal changes in the epigenome while co-opting the normal transcriptional machinery to drive leukemogenesis.

## 2.4.2 Introduction

Leukemias are a class of blood disorders characterised by the uncontrolled proliferation of hematopoietic stem and progenitor cells (HSPC) that have acquired a block in the normal process of differentiation (Andersson, Ma et al. 2015, De Kouchkovsky and Abdul-Hay 2016, Saultz and Garzon 2016). Recent large-scale studies have highlighted the age-related genetic differences between pediatric and adult acute myeloid leukemia (AML) and have also revealed the high level of genetic heterogeneity in this disease (Bolouri, Farrar et al. 2018). Not only are chromosomal translocations more frequent in younger patients, but recurrent mutations in specific genes implicated in the disease show a bias in either pediatric patients (e.g. *NRAS*, *KIT*, *KRAS*) or adults (*DNMT3A*, *NPM1*, *IDH1/2*) (Cancer Genome Atlas Research, Ley et al. 2013). As a result, defining the combined role of these mutations in each patient, especially in pediatric patients in whom predisposition variants may play a greater role, is extremely challenging and increases the complexity of identifying novel potential targeted therapeutics. Studies underscore the fact that even though the mutational burden in AML is one of the lowest of any cancer (Bolouri, Farrar et al. 2018) there is still tremendous heterogeneity between leukemic patients which creates challenges when trying to define the genetic determinates of the disease.

As noted, chromosomal fusions are common in pediatric AML, particularly those involving the lysine specific methyl transferase 2A gene (*KMT2A*; also known as Mixed Lineage Leukemia (*MLL*) gene). *KMT2A* is rearranged in approximately 10% of all leukemias (Winters and Bernt 2017) but the frequency in infant acute lymphocytic leukemia (ALL, >70% (Sam, Kersey et al. 2012)) and AML (>35% (Sam, Kersey et al. 2012)) is much higher. The *KMT2A* gene encodes a large (500 kDa) and complex 38-exon protein (Gu, Nakamura et al. 1992, Tkachuk, Kohler et al. 1992) that can be cleaved by taspase-1 into two separate fragments (KMT2A-C and KMT2A-N) (Hsieh, Ernst et al. 2003). The conserved SET domain located at the C-terminus of the protein is responsible for the methylation of the histone H3 at lysine 4 (H3K4). Previous studies have shown that KMT2A is part of a large macromolecular complex composed of different proteins (Slany 2009) that function to improve the stability of KMT2A allowing the complex to regulate the transcriptional activation of *HOX* genes. The *KMT2A* gene can be fused in-frame with more than 120 different partners, creating fusion proteins that are typically associated with poor prognosis leukemias (Winters and Bernt 2017). One of the most

common fusion partners is the *MLLT3* gene (also known as *AF9*), which is found in 30% and 13% of *KMT2A*-rearranged AML or ALL (Meyer, Burmeister et al. 2018), respectively; this fusion is associated with an intermediate risk. The *KMT2A*-*MLLT3* (KM3) fusion protein is part of the DOT1L complex (DOTCOM) and leads to aberrant expression of specific target genes marked by H3K79 methylation (Bernt, Zhu et al. 2011). Despite the identification of several key gene targets required for transformation, the complete molecular mechanisms utilized by this oncogenic fusion are still unclear.

The strong oncogenic potential of *KMT2A* fusions makes them ideal drivers for *in vivo* experimental model systems to explore molecular mechanisms involved in leukemogenesis (Milne 2017). To specifically overcome the problems of patients' genetic heterogeneity and scarcity of samples, we have developed a single-donor, human model leukemia system using healthy CD34<sup>+</sup> cord blood HSPC (Barabe, Gil et al. 2017). This adaptation of a prior human model (Barabe, Kennedy et al. 2007) uses a retrovirus to deliver the human KM3 fusion gene which we have shown is sufficient to generate a human leukemia (Barabe, Gil et al. 2017). Transduced cells are cultured *in vitro* for 30-40 days prior to injection into immune-deficient mice which develop a leukemia 24-30 weeks later. Because we can sequence the transcriptome/genome/exome of the initial CD34<sup>+</sup> cells used to generate the model AML, the genetic background (e.g. single nucleotide variants) of the initial donor is well defined, allowing the potential role of acquired mutations to be assessed precisely. In addition to these advantages, this model also allows the study of genetic mechanisms involved in the initiation of the disease, when the fusion is first introduced but before the complete transformation of the cells. Our previous genomic analyses demonstrated that the oncogenic fusion alone is sufficient for the development of leukemias, without the requirement for any recurrent secondary mutations (Barabe, Gil et al. 2017), despite the presence of these in pediatric patients with *KMT2A*-mutated AML. We have also used these data to identify a number of genes that represent novel biomarkers in patients with *KMT2A* fusions (Lagace, Barabe et al. 2017). Interestingly, while approximately one-third of these biomarker genes were expressed in our model leukemia system shortly after cells were transduced with the KM3 fusion, the majority were only expressed after xenotransplantation. This observation, coupled with the limited potential for *in vitro* growth of these cells, suggests that cells partially transformed (or "primed") by KM3 may require

additional *in vivo* signals (e.g., from the bone marrow niche) to complete their leukemic transformation. The single-donor model leukemias therefore not only recapitulate the behavior and phenotype of the disease, but also provide an experimental system to study genetic and epigenetic mechanisms involved in the disease, and a unique insight into early stages of transformation.

In the present study, we leverage these advantages to perform a detailed epigenetic analysis of the various stages of the model and, through correlation with model and expression data from patients with leukemia, define the epigenetic changes driven by the oncogenic fusion that contribute to leukemia development. Analysis of changes in the patterns of DNA methylation confirmed previous observations of profound hypomethylation in *KMT2A*-translocated AML. Interestingly, however, B-cell ALL driven by the same fusion show a smaller number of differentially methylated cytosines relative to HSPC methylation, which are predominately hypermethylated. Despite the generally poor global correlation between DNA methylation and gene expression, we identified *ADCY9*, a member of the adenylyl cyclase family, as a gene essential for KM3-AML growth that also exhibits coordinated changes in DNA methylation and gene expression. We further characterized changes in chromatin accessibility through an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-sequencing) along with alterations in histone modifications including H3K4me3 and H3K79me2. Interestingly, our ATAC-sequencing analysis revealed that the vast majority of regions of open chromatin are shared with normal HSPC and monocytes, with very few being specific to the leukemic cells. Even those regions that are unique to the leukemia show a high degree of overlap with regions of open chromatin seen in other cell types and are also characterized by the presence of known clusters of transcription factor binding sites. Integration of the epigenetic data with expression data highlighted a role for the histone demethylase KDM4B during the initial stages of leukemic transformation through its impact on *S100A8/A9* expression levels, whose relative ratios have been demonstrated to be critical for blocking differentiation of leukemic cells (Laouedj, Tardif et al. 2017). Collectively, these results suggest the leukemic transformation of normal HSPC by KM3 involves a minimal epigenetic shift that also implicates co-option of normal transcriptional networks.

## 2.4.3 Results

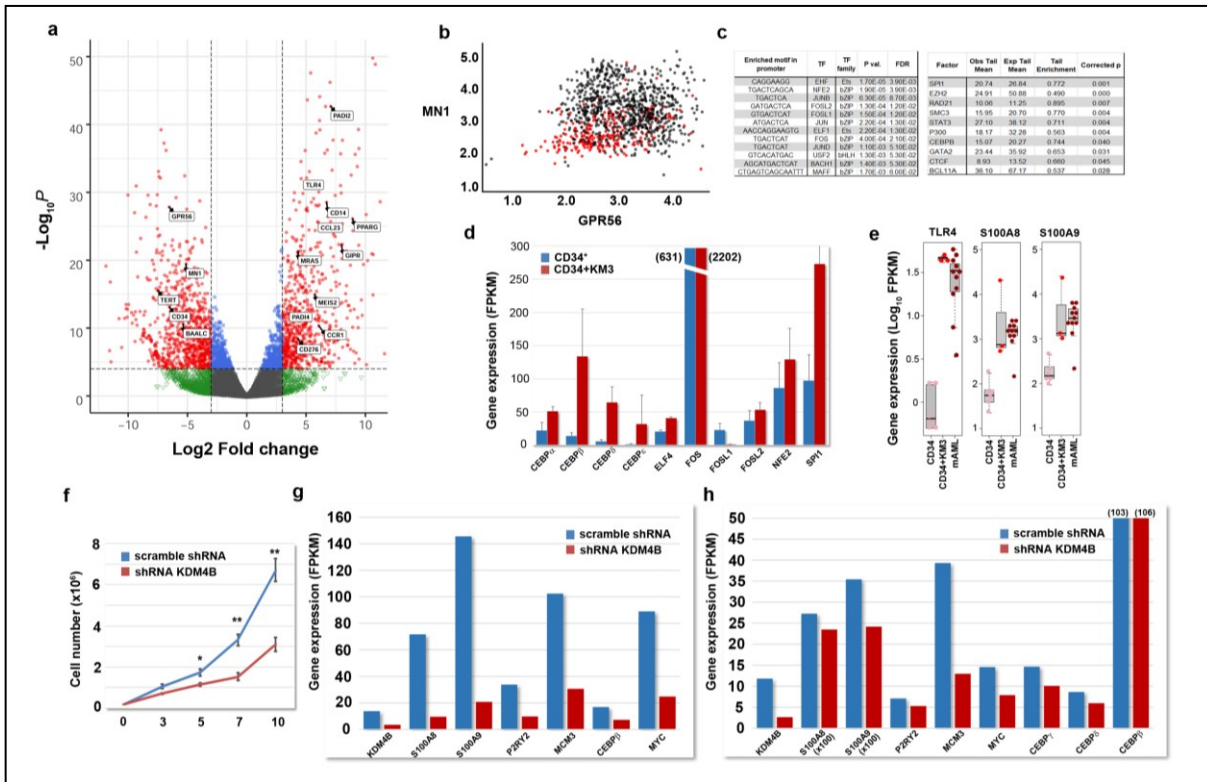
### 2.4.3.1 Transcriptional consequences of KM3 expression

To characterize the short-term transcriptional changes induced by the KM3 fusion, we analyzed the RNA-sequencing data from the initial unmanipulated cord blood cells and these cells after their transduction with the KM3 fusion. As previously described (Barabe, Kennedy et al. 2007), untransduced cells are largely lost in culture after about 1 month of *in vitro* culture and the remaining cells show induction of genes associated with a myeloid phenotype (e.g., *CCR1*, *CD14*; **Figure 2.1a**). Interestingly, transduced cells also display the loss of normal or leukemia stem cell markers (e.g. CD34 and GPR56 respectively) and the loss of expression of other genes previously associated with AML induction in other subtypes (e.g.. *MNI* (Sharma, Jyotsana et al. 2019)). This observation is supported by expression data from patients with *KMT2A*-translocated AML in the TARGET study (Bolouri, Farrar et al. 2018), which showed an equivalent loss of *MNI* and *GPR56* (Bolouri, Farrar et al. 2018) (**Figure 2.1b**) and a similar loss of expression in studies of *GPR56* seen in adult AML (Pabst, Bergeron et al. 2016). A global view of the expression changes through gene set enrichment analysis confirmed the loss of stem cell phenotype and the gain in expression of HOXA9-MEIS1 targets consistent with leukemic transformation (**Figure 2.2**).

To examine the regulation underlying the transcriptional changes seen, we studied the transcription factor binding motifs upstream of the differentially expressed genes. Analysis of enriched motifs using several different software (including HOMER (Heinz, Benner et al. 2010), MAGICTRICKS (Roopra 2019), ShinyGO (Ge and Jung 2018)) produced differing lists of potential transcription factors (**Figure 2.1c**), but factors in the CEBP and FOS families, along with NFE2, were highlighted consistently, at differing ranks. Analysis of RNA-sequencing data confirmed that the expression of all members of the CEBP and FOS families increased during *in vitro* culture (**Figure 2.1d**), along with transient changes in ETS, MAF, and JUN families (**Figure 2.3**). We also observed an increase in the expression of *S100A9* and *S100A8* (**Figure 2.1e**), which have been described as direct targets of CEBP $\beta$  (Endoh, Chung et al. 2009). It has been recently demonstrated that the ratio of these calcium-binding proteins is critical for blocking AML differentiation (Laouedj, Tardif et al. 2017), with S100A8 opposing the

differentiation promoted by S100A9, which acts through *TLR4*. Interestingly, and in agreement with these findings, while the expression of both *S100A8* and *S100A9* increase when KM3 was added to CD34<sup>+</sup> cells, there was a larger increase in *S100A8*, changing the ratio of S100A8:S100A9 from 1.1:1 to 3.2:1, favoring inhibition of differentiation. Other CEBP members, including CEBP $\beta$ , have also previously been implicated in the process of normal myeloid differentiation (Tsukada, Yoshida et al. 2011), which suggests that transcription factors normally present are co-opted to activate genes that are part of a leukemic transcriptional program. At the same time, these factors and others upregulated are also involved in immune and inflammatory pathways, suggesting that there may be additional transcriptional programs that are activated and superimposed on the myeloid differentiation that we observe. To determine whether changes in the expression of chromatin-modifying enzymes might be correlated with our RNA-sequencing data, we looked for enzymes whose expression increased at least 2-fold during *in vitro* culture and whose expression was maintained in both the model and patients' AML (**Table 2.1**). This analysis highlighted a single gene with this profile, *KDM4B*, which has been implicated in murine AML, albeit with functional redundancy (Agger, Miyagi et al. 2016). The shRNA-mediated knockdown of *KDM4B* in THP-1 cells showed a significant impact on growth (**Figure 2.1f**) and RNA-sequencing from these cells revealed a profound downregulation of *S100A8* and *S100A9*, *CEBP $\beta$* , *P2RY2* (a KM3-AML biomarker (Lagace, Barabe et al. 2017)), and other cell cycle regulators (**Figure 2.1g**). We next performed the same experiment but with CD34<sup>+</sup> cells transduced with the KM3 fusion to validate the changes in our model system. These data showed very similar gene expression changes in these genes with the exception of *CEBP $\beta$* , although other CEBP family members were downregulated (**Figure 2.1h**). Together, these data support the hypothesis that the early epigenetic changes, at least partially mediated through KDM4B, lead to activation of *S100A8/9* as well as other genes required for sustained proliferation. Given the similarity of their DNA binding consensus sequences, defining the precise contribution of different CEBP family members in this process will require further focused studies.





**Figure 2.1 – Gene expression changes in the *KMT2A-MLL3* leukemia model.**

**(A)** Volcano plot of differentially expressed genes (red dots are genes with  $p$ -value  $< 10^{-5}$ , and a  $\log_2$  fold change  $> 3$ ) from RNA-sequencing data of  $CD34^+$  cells and  $CD34+KM3$  cells. Highlighted genes include known markers for stem cells, myeloid cells or previously reported biomarkers for  $KM3$ -AML.

**(B)** Scatterplot of *MNI* and *GPR56* expression genes in approximately 1,000 cases of pediatric AML in the TARGET cohort. Red dots represent patients with *KMT2A*-rearranged AML.

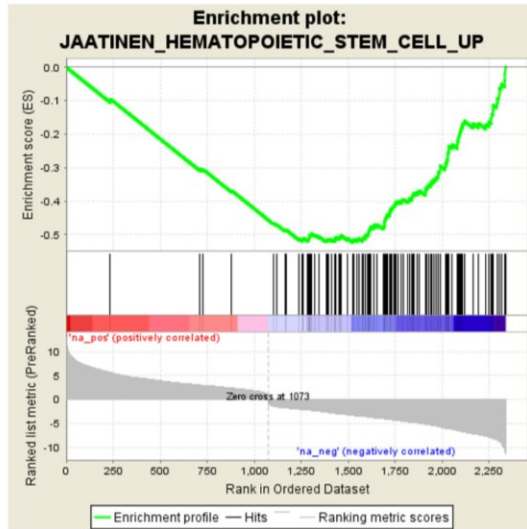
**(C)** Summary of transcription factor binding motifs found to be enriched in promoters of upregulated genes in **(A)** with either ShinyGO v0.61 (Ge and Jung 2018) or MAGICTRICKS (Roopra 2019).

**(D)** Gene expression levels of transcription factor families highlighted in **(C)** in either  $CD34^+$  cells (blue) or  $CD34+KM3$  cells (red).

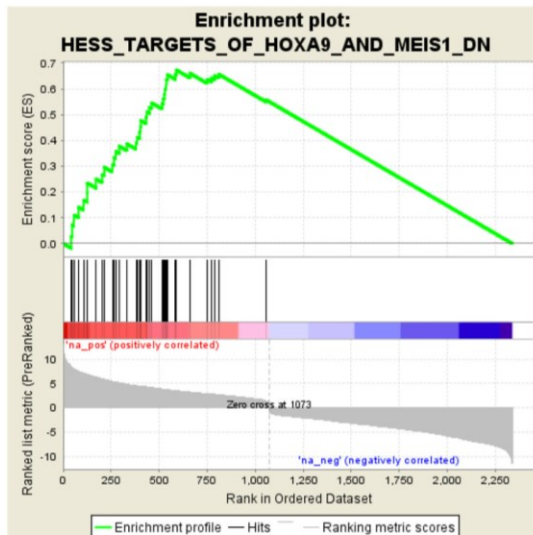
**(E)** High levels of expression of *TLR4*, *S100A8* and *S100A9* are seen after introduction of the *KMT2A-MLL3* fusion and are maintained in the model AML.

**(F)** Loss of *KDM4B* expression through shRNA knockdown in the *KMT2A-MLL3*<sup>+</sup> cell line (THP-1) induces growth inhibition (\* =  $p < 0.005$ , \*\* =  $p < 0.001$ ) and **(G)** reduced expression of *S100A8/A9*, *CEBP $\beta$*  and other cell-cycle related genes.

**(H)** shRNA knockdown of *KDM4B* in  $CD34^+$  cells transduced with the  $KM3$  fusion show gene expression decreases very similar to those in THP-1 cells.

**a**

GeneSet	JAATINEN_HEMATOPOIETIC_STEM_CELL_UP
Enrichment Score (ES)	-0.5234885
Normalized Enrichment Score (NES)	-3.1416063
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

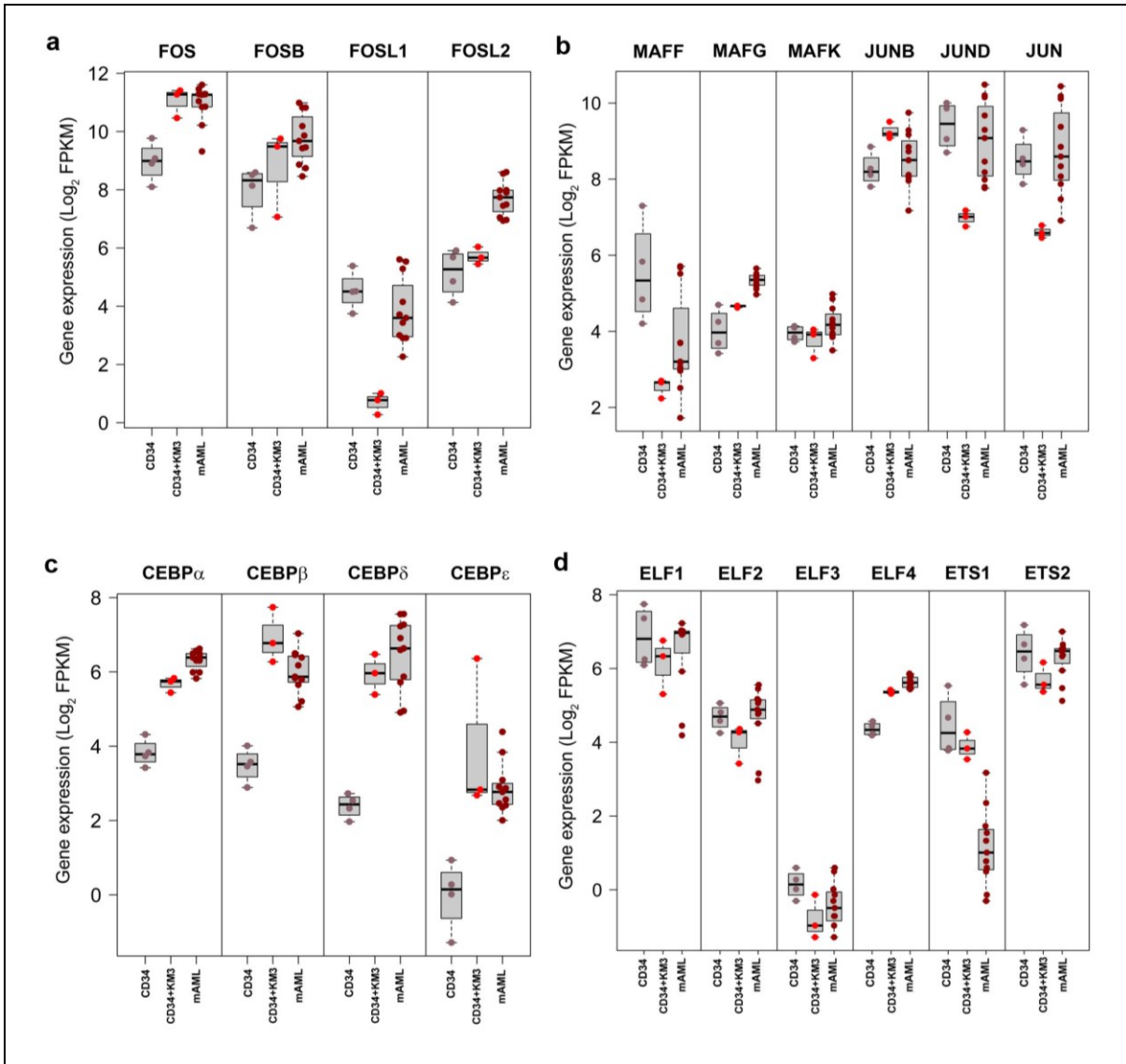
**b**

GeneSet	HESS_TARGETS_OF_HOXA9_AND_MEIS1_DN
Enrichment Score (ES)	0.67227644
Normalized Enrichment Score (NES)	3.4345574
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

**Figure 2.2 – Gene set enrichment analysis of *KMT2A-MLL3* expressing CD34<sup>+</sup> cells.**

After performing differential gene expression analysis using RNA-seq data from human CD34<sup>+</sup> and CD34<sup>+</sup> cells transduced with the *KMT2A-MLL3* gene fusion, a ranked list of significantly altered genes was used for gene set enrichment analysis (GSEA) to match to published signatures.

(A) Genes whose expression was lost clearly matched HSC signatures while genes that increased in expression (B) matched very closely published targets of HOXA9 and MEIS1.



**Figure 2.3 – Expression of various transcription factor families in model leukemia stages.**

Box plots of gene expression data ( $\log_2$  of FPKM values) from three different stages in model leukemia system are shown for several transcription factor families. Transcription factor families include (A) FOS and FOSL, (B) MAF and JUN, (C) CEBP and (D) ETS.

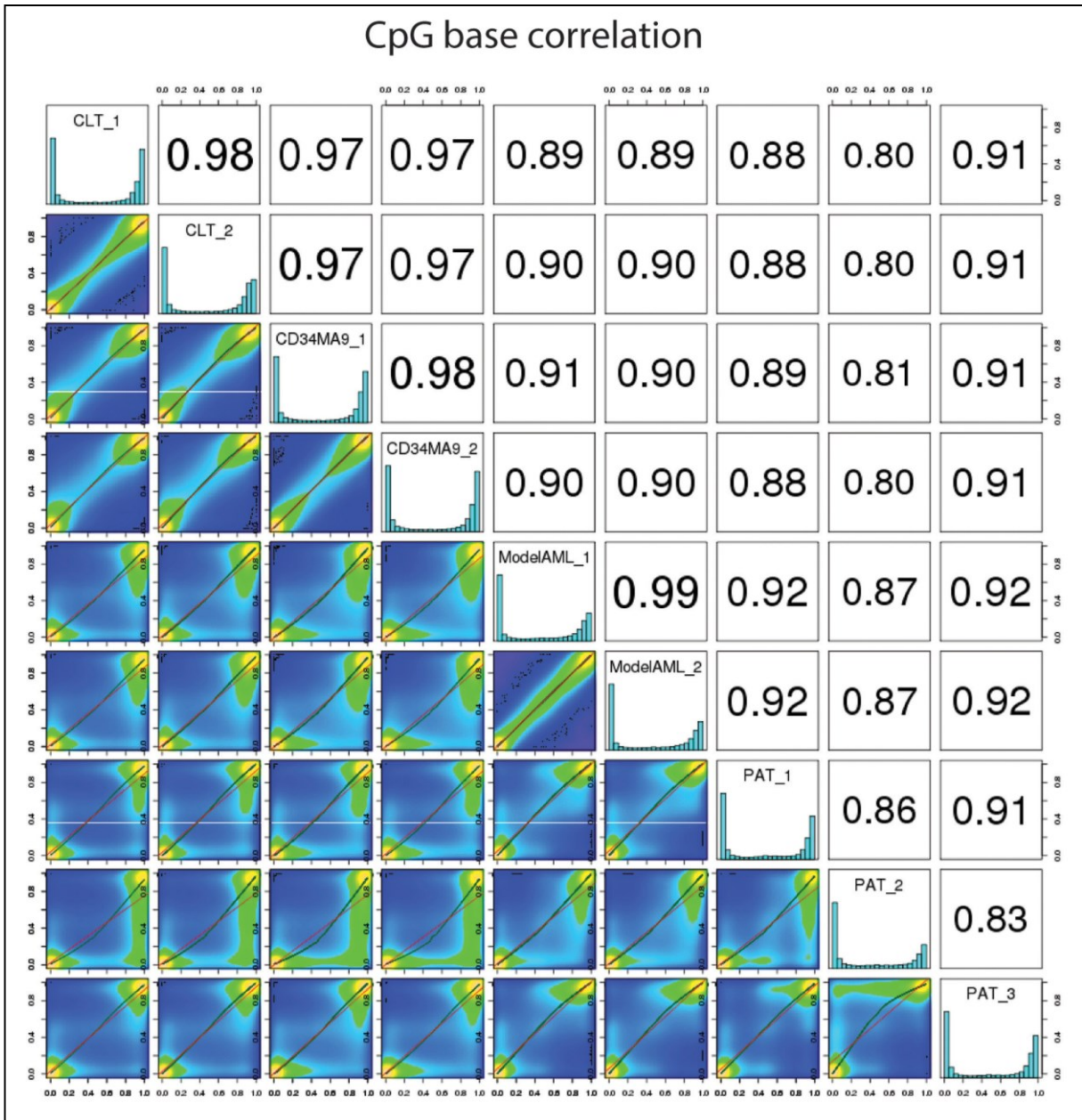
FPKM : Fragments per kilobase of exon per million mapped reads.

Gene	CD34	CD34+MA9	model AML	patient AML	mononuclear
KDM1A	46.471	18.755	22.589	24.626	15.197
KDM1B	12.039	16.921	11.720	17.662	10.199
KDM2A	37.539	20.886	36.493	37.906	42.284
KDM2B	16.744	13.540	21.444	17.248	10.929
KDM3A	17.257	28.231	14.068	22.170	23.684
KDM3B	20.863	23.689	19.920	17.813	17.052
KDM4A	14.711	17.619	18.516	19.248	5.202
KDM4B	11.361	22.256	8.988	9.353	13.776
KDM4C	6.737	8.488	6.843	7.315	11.092
KDM4D	0.286	0.138	0.041	0.076	0.062
KDM4DL	0.160	0.000	0.016	0.026	0.018
KDM5A	9.035	10.903	8.611	10.799	8.188
KDM5B	33.772	10.942	6.980	12.961	9.377
KDM5C	24.483	19.851	19.540	24.656	18.954
KDM5D	17.998	35.667	11.306	0.003	9.699
KDM6A	8.392	7.614	8.840	13.332	11.476
KDM6B	18.730	16.113	32.582	18.426	30.926
DNMT1	35.691	20.617	50.089	29.453	18.929
DNMT3A	20.330	8.424	6.115	6.457	10.768
DNMT3B	17.454	1.187	1.685	1.100	1.082
DNMT3L	0.225	0.046	0.029	0.094	0.138
TET1	0.723	0.173	0.369	0.406	0.378
TET2	4.916	8.426	12.149	10.080	5.934
TET3	9.040	15.189	12.756	9.543	13.917
HDAC1	77.530	45.049	52.593	57.414	45.955
HDAC10	21.643	18.454	25.500	30.143	19.384
HDAC11	2.707	0.846	0.987	1.490	0.992
HDAC2	17.853	10.142	12.326	13.356	6.523
HDAC3	34.843	28.302	42.277	36.867	23.097
HDAC4	2.440	7.822	12.053	12.984	3.985
HDAC5	18.406	10.859	18.155	17.600	16.219
HDAC6	28.259	14.481	22.566	24.029	16.753
HDAC7	57.189	11.657	18.272	18.399	44.518
HDAC8	5.560	4.548	4.139	3.364	1.891
HDAC9	2.568	0.928	4.653	6.201	2.599
SUV39H1	6.791	4.060	9.971	5.327	3.284
SUV39H2	3.267	1.604	1.370	2.067	1.090
JMJD1C	17.121	8.540	61.059	58.157	22.144
JMJD4	6.837	4.110	3.535	4.776	1.887
JMJD5	2.456	1.309	1.729	1.594	1.630
JMJD6	8.376	7.191	18.882	15.544	12.219
JMJD7	8.551	7.777	10.008	7.202	13.590
JMJD8	21.773	12.187	22.350	25.104	13.826

Table 2.1 – Gene expression changes (FPKM) in chromatin modifying enzymes.

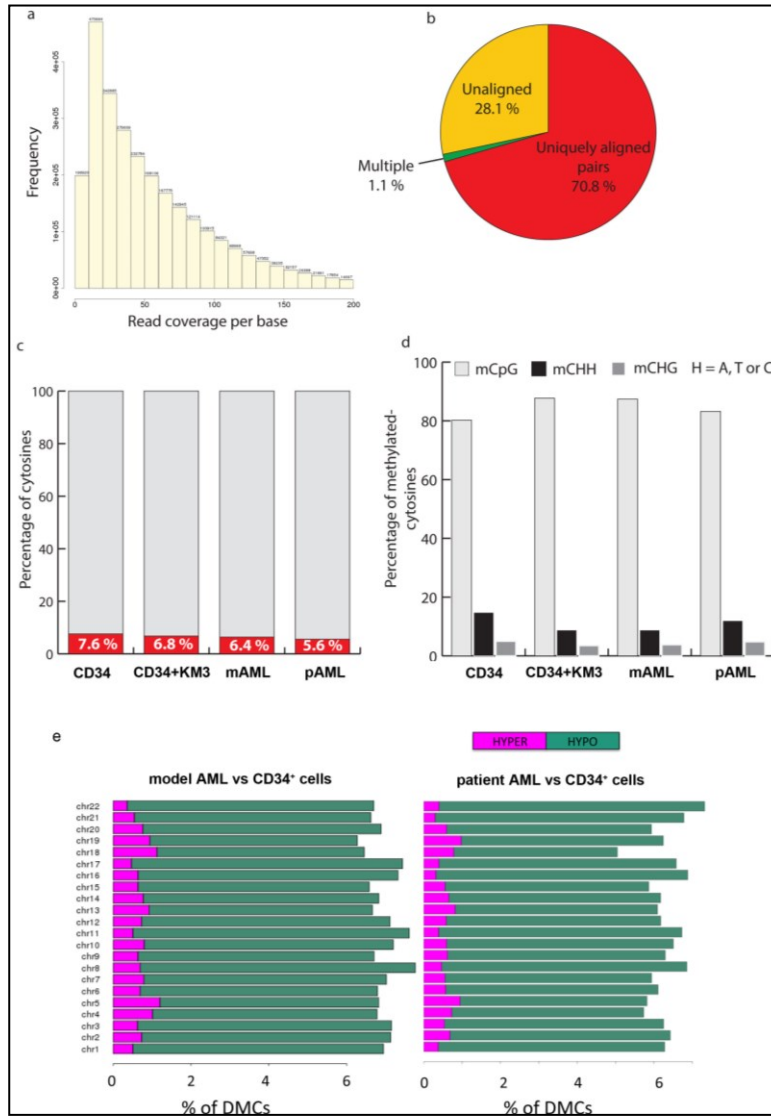
#### **2.4.3.2 Sequential changes in DNA methylation are not globally correlated with expression changes.**

Given the known role of KM3 as an epigenetic regulator and the demonstrated importance of DNA methylation changes in other subtypes of AML (Ley, Ding et al. 2010, Sandoval, Huang et al. 2019), we next sought to investigate changes in DNA methylation in our model leukemias and genetically matched patients' samples. Using methyl-sequencing, capture-based, bisulfite sequencing (Lee, Luo et al. 2013), we assessed DNA methylation levels genome-wide at 84Mbp of target sequence regions including CpG islands, promoters and published differentially methylated regions. Biological duplicates at each step of our model leukemia system were compared to bisulfite-treated DNA from three adult KM3 AML patients. In line with expression data from our model leukemias (Barabe, Gil et al. 2017), the methylation data between replicates were highly consistent (**Figure 2.4**) and were also consistent with published data regarding global methylation levels (Akalin, Garrett-Bakelman et al. 2012) (**Figure 2.5**).



**Figure 2.4 – Intra- and inter-individual reproducibility of methyl-sequencing replicates.**

Scatterplots for pairwise comparisons of percentage levels of methylation in different samples used (CLT\_1/2 = CD34+ controls, CD34MA9\_1/2 = CD34+KM3, PAT\_1/2 = patient samples 1 and 2). Numbers on upper right corners denote pair-wise Pearson correlation scores. The histograms on the diagonal are percentage methylation histograms.



**Figure 2.5 – Descriptive statistics of methylation data.**

**(A)** Histogram of read coverage per cytosine for the KM3-AML model (one example of the 2 biological replicates).

**(B)** Pie chart of the alignment statistics for the KM3-AML model (one of the 2 biological replicates).

**(C)** Percentage of methylated (red) and unmethylated (grey) cytosines (compare to all cytosine loci in the genome) for the control CD34+, CD34+KM3 and model AML cells.

**(D)** Composition of all methylated cytosine loci (mCpG:grey; mCHH: black; mCHG: dark grey) in the control CD34+, CD34+KM3 and model AML cells.

**(E)** Percentage of differentially methylated cytosines (DMCs) in the 22 chromosomes, in model AML (left panel) or in patient AML (right panel), compared to CD34+ cells (pink : hypermethylation; green : hypomethylation).

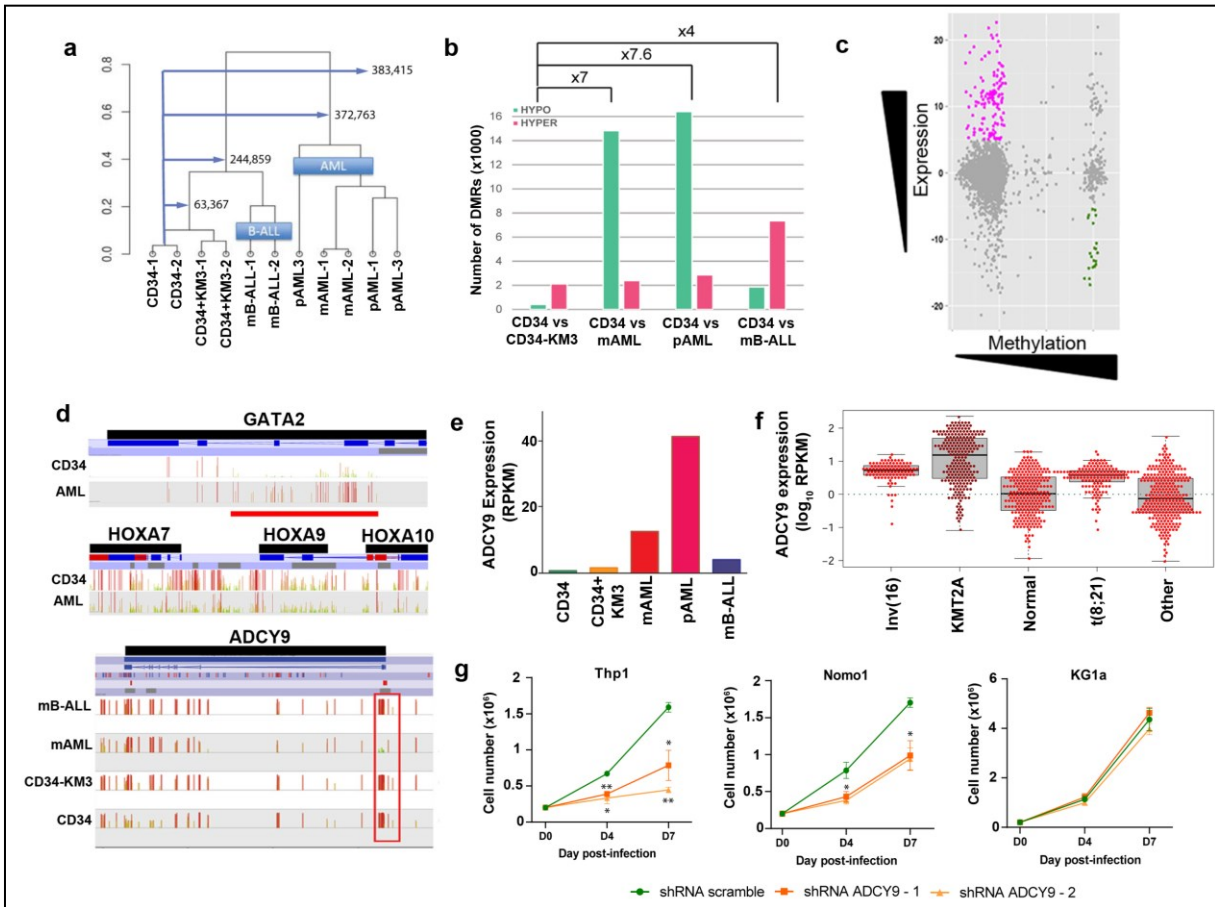
Hierarchical clustering of differentially methylated cytosines (DMCs; CpG coverage  $\geq 8$ ,  $q$ -value  $\leq 0.1$  and methylation differences  $\geq 40\%$ ) demonstrated that relatively small numbers of changes occurred after introduction of the KM3 fusion gene and *in vitro* culture (**Figure 2.6a**). The number of genes with DMC present varied from 7,645 (CD34<sup>+</sup> vs. CD34+KM3) to 13,947 (CD34<sup>+</sup> vs. model AML), however a large percentage of these genes (40% and 20%, respectively) contained only a single DMC. Interestingly, relative to the initial CD34<sup>+</sup> cells with the KM3 fusion, there was a striking increase ( $>6$  fold) in the number of DMC detected in the final model leukemias and patients' samples. This suggests that the initial epigenetic changes induced by the fusion gene are relatively limited, with the majority of changes occurring only during *in vivo* growth. Along with the previous stepwise changes in gene expression patterns we have reported (Barabe, Gil et al. 2017), these results again support the importance of microenvironmental cues received *in vivo*. With respect to the direction of change, in agreement with recent observations in patients with *KMT2A*-AML (Akalin, Garrett-Bakelman et al. 2012), DMC observed in both model and patient KM3-AML were profoundly biased ( $\sim 89\%$ ) towards demethylation (**Figure 2.6b**). In opposition to this trend, model leukemias generated from the same donor cells but which developed a lymphoid phenotype (B-cell ALL) had primarily hypermethylated cytosines, and fewer DMC compared to the AML samples. The majority ( $>75\%$ ) of DMC were located in gene bodies and particularly in promoter regions and putative functional elements at various stages of the model AML and in patients (**Figure 2.7**, **Figure 2.8**).

To clarify the importance of the localization of these changes and their potential impact, we next examined the correlation between DNA methylation changes and gene expression. Despite the general view regarding the impact of DNA methylation changes in the promoter regions and transcription levels, we observed only a weak global correlation between the two (**Figure 2.6c**). While we could find examples of expected changes (e.g., in previously described *KMT2A* target genes such as *GATA2*, *HOXA7* and *HOXA9* (Celton, Forest et al. 2014, Jung, Dai et al. 2015) (**Figure 2.6d**)), overall, there was a very poor global correlation. This observation mirrors those made in recent comprehensive single-cell studies of DNA methylation changes during normal hematopoiesis in humans (Farlik, Halbritter et al. 2016) as well as previous bulk studies (Hodges, Molaro et al. 2011). We were able to identify new examples of genes with evidence for epigenetic regulation such as the atypical adenylate cyclase gene *ADCY9*. The promoter



region of *ADCY9* exhibits hypomethylation specifically in the model AML compared to normal CD34<sup>+</sup> cells and model B-cell ALL (**Figure 2.6e**) and the gene is only expressed in *KMT2A*-rearranged model and patients' leukemias (**Figure 2.6f**). *ADCY9* is a poorly characterized and distantly related member of the adenylate cyclase gene family and no specific inhibitors have yet been described, although it has recently been identified as a fusion partner of the *KMT2D* gene in a case of ALL (Mankuzhy, Walling et al. 2019). We therefore perform shRNA knockdowns in several KM3-AML cell lines to determine whether *ADCY9* expression was relevant to these leukemias. We found that all the KM3-AML lines showed impaired proliferation with multiple shRNA in contrast to no effect in a non-KM3-AML control line (**Figure 2.6g**) demonstrating that *ADCY9* is required in these cells.

Beyond regulating specific genes, we next wanted to examine whether global methylation changes might have an impact on gene expression through their localization within the DNA binding sites of transcription factors. To do this, we selected all the DMC with coverage that exhibited consistently higher ( $\geq 40\%$ ) methylation levels in CD34<sup>+</sup> cells than in either CD34<sup>+</sup>KM3 or model leukemias. Transcription factors in the ENCODE ChIP-sequencing dataset (covering 338 factors in 130 cell lines) that overlapped the positions of these cytosines were then used to generate a network (using GOnet (Pomaznoy, Ha et al. 2018)) based on shared gene ontology terms (**Figure 2.9**). This network contains a large collection of factors that: i) are highly expressed in normal bone marrow, ii) exhibit a large loss of DNA methylation directly in their binding sites after KM3 addition, and iii) are enriched for activity in regulating hematopoietic and myeloid differentiation. These results indicate that despite the poor overall correlation between DNA methylation and the expression of individual genes, the changes that occur are globally relevant for leukemic transformation.



**Figure 2.6 – DNA methylation changes in model leukemias.**

(A) Hierarchical clustering of differentially methylated cytosines (DMC) identified by methylsequencing analysis of model leukemias at different stages and KM3 patients' samples (pAML). Numbers and arrows represent the number of DMC relative to the initial CD34<sup>+</sup> cells.

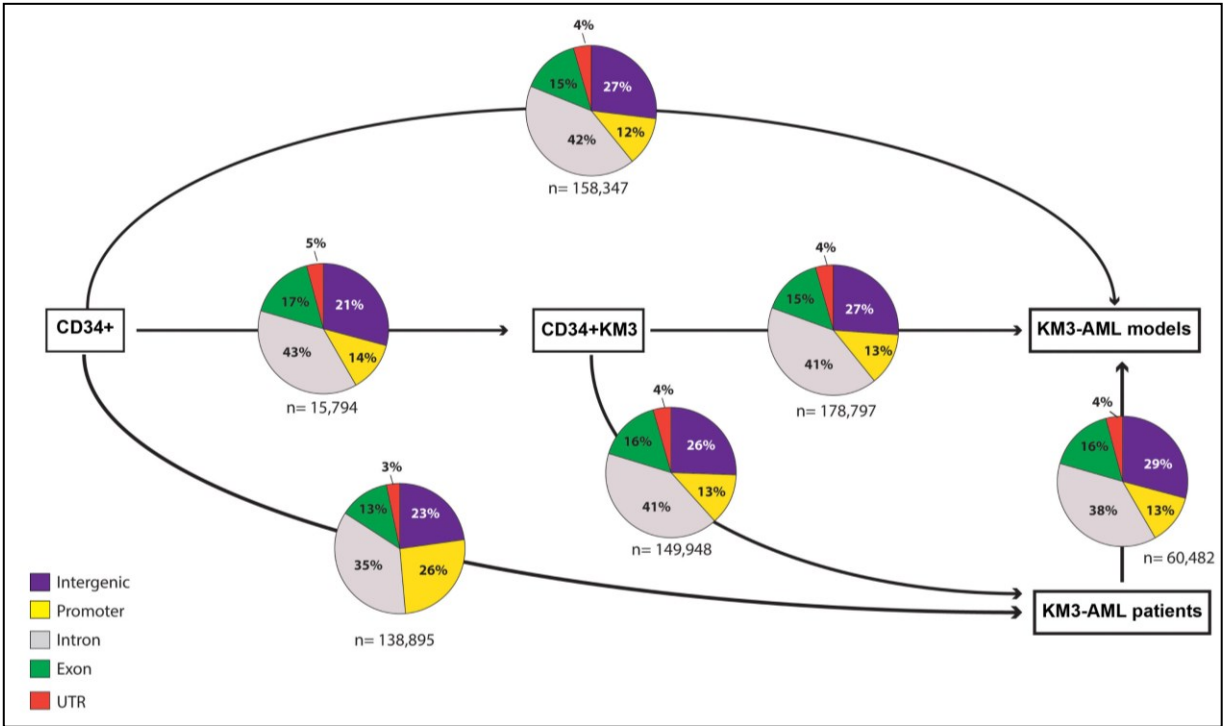
(B) The KM3 fusion induces a strong hypomethylation phenotype in both patient and model acute myeloid leukemias (pAML and mAML, respectively), whereas model B-cell acute lymphoblastic leukemias (mB-ALL) predominately exhibit hypermethylation.

(C) As noted in other global studies, the overall correlation between DNA methylation and expression is relatively limited, likely due in part to the complexity of the localization and size of the changes.

(D) Differentially methylated regions around genes known to be suppressed (*GATA2*) or activated (*HOXA9*) in AML can be seen comparing CD34<sup>+</sup> cells versus AML patients' cells.

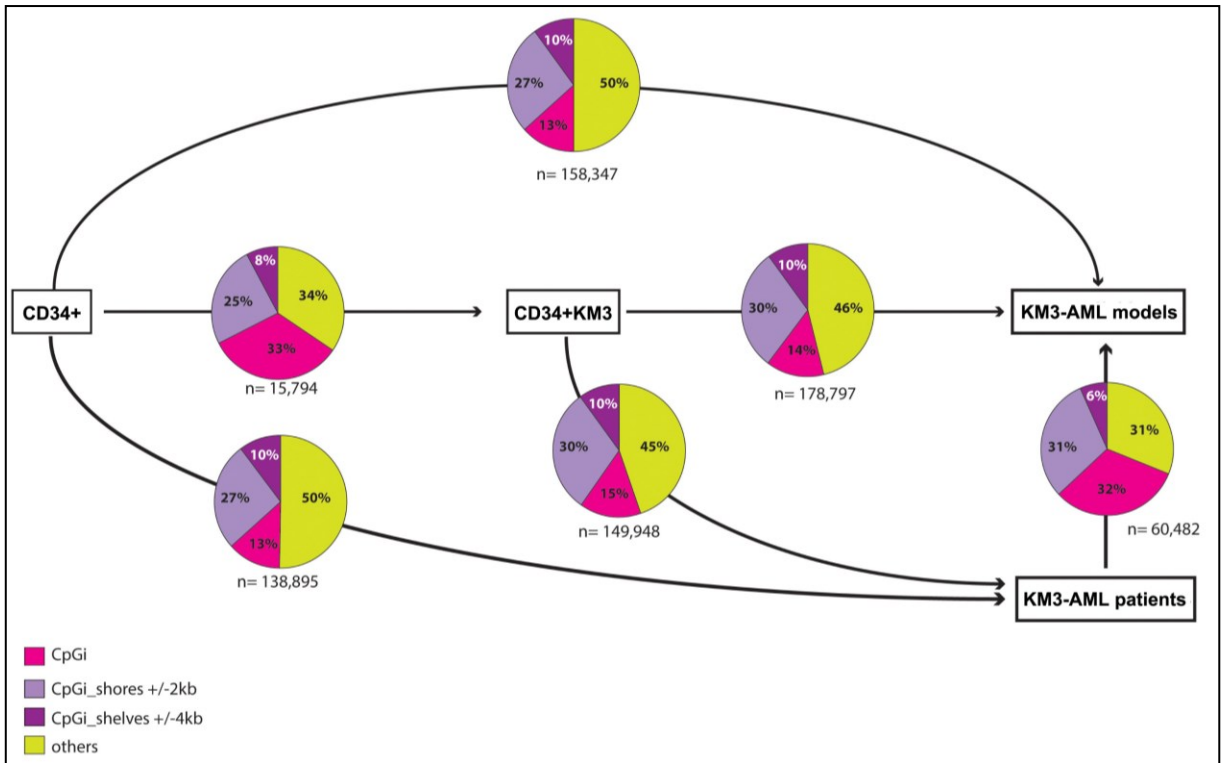
(E, F) *ADCY9* exhibits hypomethylation in its promoter region specifically in the model AML, compared to CD34<sup>+</sup> cells, which is correlated with gene expression levels, specifically in *KMT2A-MLL3* AML models as well as patients (E) and external data from TARGET (F).

(G) Knockdown of *ADCY9* using multiple shRNA shows a significant impact specifically on KM3-AML cell proliferation but not on AML cells lacking the fusion (KG1a; \* =  $p < 0.005$ , \*\* =  $p < 0.001$ ).



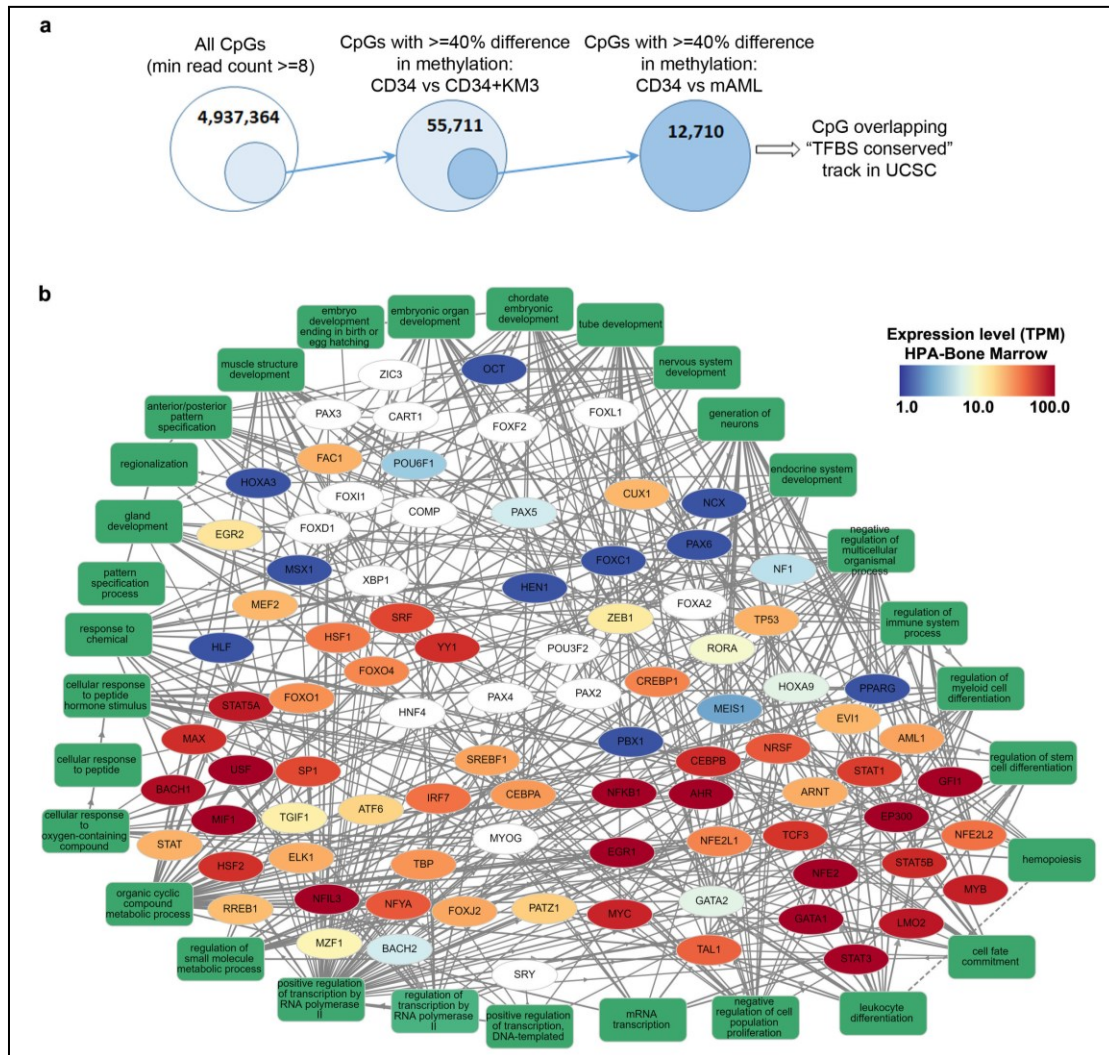
**Figure 2.7 – Distribution in functional genomic elements of DMCs associated to the development of KM3 translocated model and patient AMLs.**

Differentially methylated cytosines (DMCs) from each pairwise comparison are indicated under each pie chart. Gene annotation is taken from RefSeq (GRCH37/hg19) at UCSC genome browser. Promoter regions were defined as the 1,000 bp upstream of the annotated transcription start site.



**Figure 2.8 – Distribution in and around CpG islands of DMCs associated to the development of KM3 translocated model and patient AMLs.**

Differentially methylated cytosines (DMCs) from each pairwise comparison are indicated under each pie chart. CpG islands (CpGi) are defined from RefSeq (GRCH37/hg19) downloaded from the UCSC genome browser. The CpGi shores and CpGi shelves sequences up to 2,000 bp and 4,000 bp distant, respectively.



**Figure 2.9 – DNA methylation in the transcription factor binding sites.**

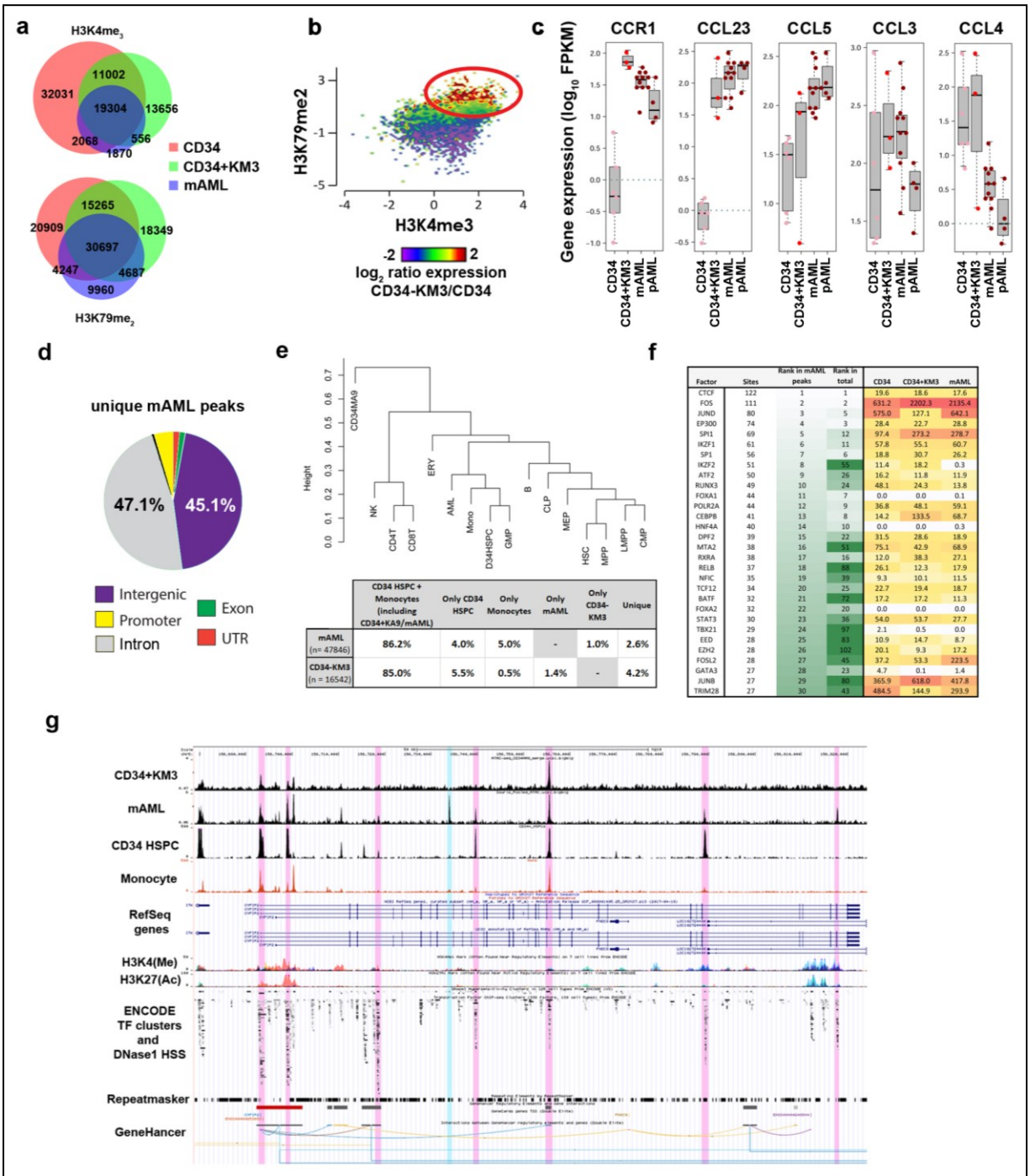
**(A)** To identify consistently differentially methylated cytosines (DMCs), positions were first limited to those in the genome that had sufficient coverage to assess changes (>8 reads). These positions were then filtered for those where there was >40% loss of methylation in both CD34+KM3 and model AMLs (12,710 final positions). These DMCs positions were then compared to the conserved transcription factor binding site track (“TFBS conserved”) in the UCSC human genome browser to identify transcription factors (TFs) whose accessibility might be impacted by methylation changes.

**(B)** The interactive GOnet website was then used to take the TFs from (A) and build a network based on shared hierarchical gene ontology terms (*q*-value threshold  $\leq 0.0001$ ). The gene nodes in the network were then automatically coloured by their expression in normal bone marrow using data from the Human Protein Atlas. The bottom right portion of the graph shows an enrichment in GO-terms associated with hematopoietic/stem cell fate and myeloid differentiation involving TFs that are all highly expressed in bone marrow.

### 2.4.3.3 Epigenetic analysis reveals potential CCR1-CCL23 autocrine signaling in KM3-acute myeloid leukemia.

Given that the KM3 fusion is a well-known epigenetic regulator, we next asked whether the presence of the gene fusion had a specific impact on the chromatin organization in our model leukemias. To investigate this, we performed ChIP-sequencing on relevant histone marks (H3K4me3, H3K79me2 (Slany 2009)) and ATAC-sequencing on samples of each step of our leukemia model (CD34<sup>+</sup>, CD34+KM3, KM3-AML). Enrichment of the different histone marks correlated with gene expression levels (**Figure 2.11**) and expected spatial distribution, and the majority of peaks seen in the model leukemias were common to both the initial and transduced cells (**Figure 2.10a**). In addition, we analyzed the combined expression and epigenetic data to compare the initial CD34<sup>+</sup> cells and CD34+KM3 cells to highlight critical early events potentially implicated in the disease. Using the EaSeq package (Lerdrup, Johansen et al. 2016) we identified a set of genes (n=331 ; **Table 2.2**) with coordinated epigenetic and expression changes (**Figure 2.10b**). Among the genes with the most striking changes was the C-C motif chemokine receptor 1 (*CCR1*) gene, a G-protein coupled receptor family member (GPCR). *CCR1* is expressed in different types of blood cells, can bind a number of different ligands and is induced during normal monocyte development (Schaller, Kallal et al. 2008). This gene was of specific interest because our previous work had shown that CCL23, a ligand only recognized by CCR1, is also a biomarker for *KMT2A*-translocated AML (Barabe, Gil et al. 2017). RNA-sequencing data from our models and genetically matched patients' samples confirmed that the expression of both *CCR1* and *CCL23* is increased dramatically upon the integration of the KM3 fusion, while neither is expressed in normal CD34<sup>+</sup> cells (**Figure 2.10c**). External RNA-sequencing data from the BloodSpot database (**Figure 2.12a**) or TARGET cohort (Bolouri, Farrar et al. 2018) also confirm the *KMT2A*-specific expression of *CCL23* (**Figure 2.12b**) and there is a small but significant difference in survival of pediatric AML patients with high levels of *CCL23* (**Figure 2.12c**). Although previous studies have highlighted the importance of *CCR1* in different cancers (e.g. breast (Shin, Lee et al. 2017), prostate (Kato, Fujita et al. 2013)) no studies have looked at its role in AML. Although we saw an inhibition of growth specifically in KM3-AML cell lines using a potent inhibitor of CCR1 (data not shown), this inhibition was only observed at levels (10-20  $\mu$ M) much higher than the reported half maximal inhibitory

concentration ( $IC_{50}$ ) (9 nM) for the inhibitor (Naya, Sagara et al. 2001). The very high expression levels of both the receptor and ligand may explain this observation; however, specific genetic models will be required to define the biological importance of this potential autocrine-signaling axis.



**Figure 2.10 – Epigenetic analysis changes in *KMT2A-MLLT3* acute myeloid leukemia.**

(A) Venn diagram of peaks determined by MACS2 in different stages of the model system (CD34<sup>+</sup>, CD34+KM3, KM3 AML model) for H3K4Me 3 and H3K79Me2 (n=3 for chromatin immunoprecipitation (ChIP)-sequencing of each mark).



**(B)** A scatterplot of ChIP-sequencing enrichments for each histone mark in an area +/- 2,000 pb (H3K4me3) or -1,000 to +3,000 pb (H3K79me2) around the transcription start site for each gene is shown, with individual genes colored by the  $\log_2$  ratio ( $CD34^+ / CD34+KM3$ ) of expression for each gene.

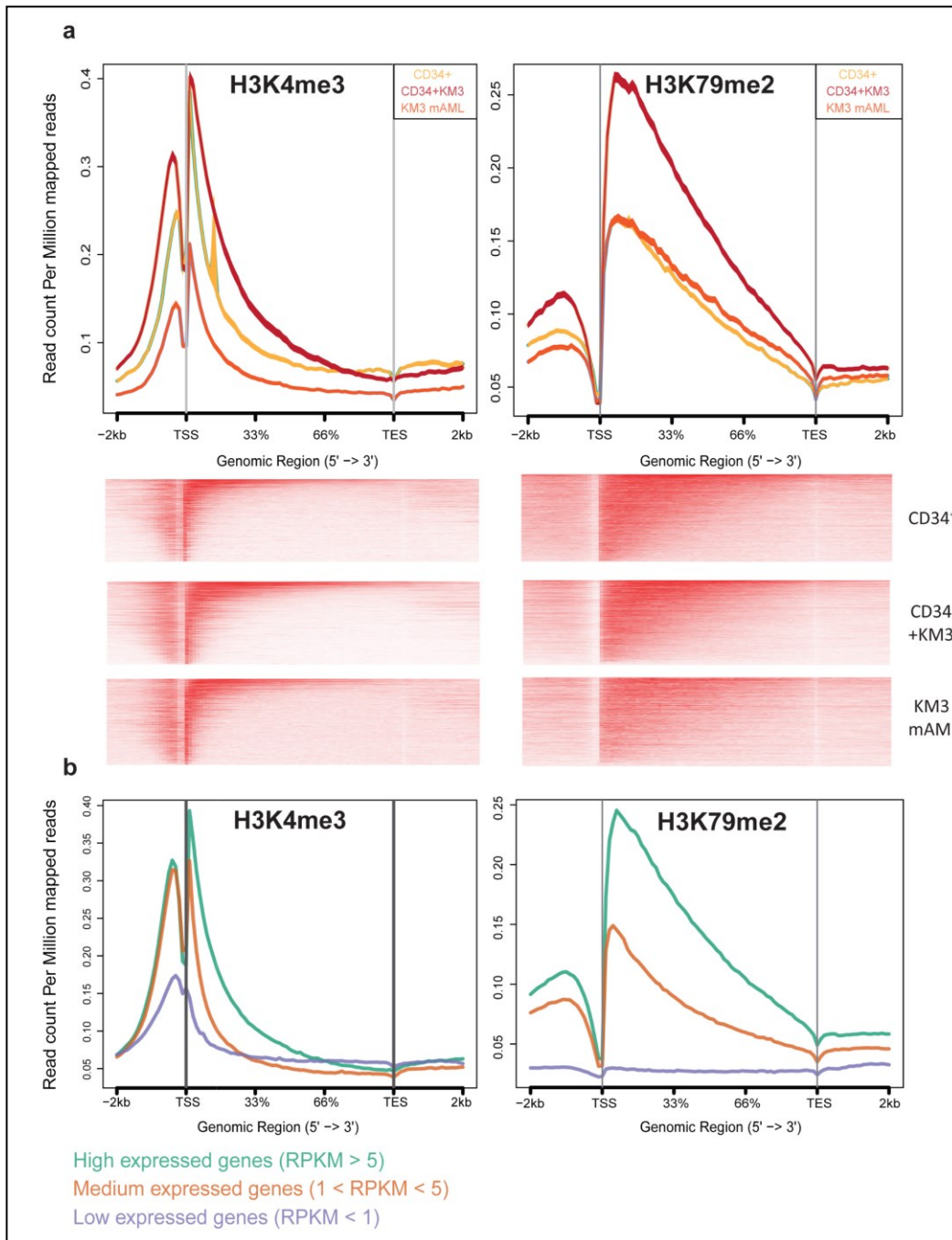
**(C)** Boxplots show the expression of *CCR1* and its ligands in different stages of the model leukemias and in pediatric *KMT2A*-AML patients. The dotted green line represents an arbitrary threshold for an expressed gene (FPKM >1).

**(D)** Annotation of ATAC-sequencing peaks unique to model leukemias (mAML) show that the majority are either in intronic or intergenic regions.

**(E)** Hierarchical clustering of ATAC-sequencing peaks in published data from different normal blood cell types compared to leukemic cells. NK = Natural Killer cells. CD4T = CD4+ T lymphocytes. CD8T = CD8+ T lymphocytes. ERY = Erythrocytes. AML = Acute Myeloid Leukemia (model). Mono = Monocytes. CD34 HSPC = CD34+ Hematopoietic Stem and Progenitor Cells. GMP = Granulocyte-Monocyte Progenitor cells. B = B lymphocytes. CLP = Common Lymphoid Progenitor. MEP = Megakaryocyte-Erythroid Progenitor cells. HSC = Hematopoietic Stem Cells. MPP = Multipotential Progenitor cells. LMPP = Lymphoid-primed MultiPotent Progenitors cells. CMP = Common Myeloid Progenitor cells.

**(F)** A list of transcription factors is shown ranked by their abundance in the mAML and with their corresponding rank in the entire dataset (ENCODE consortium) and their expression level in the model system stages in the columns to the right.

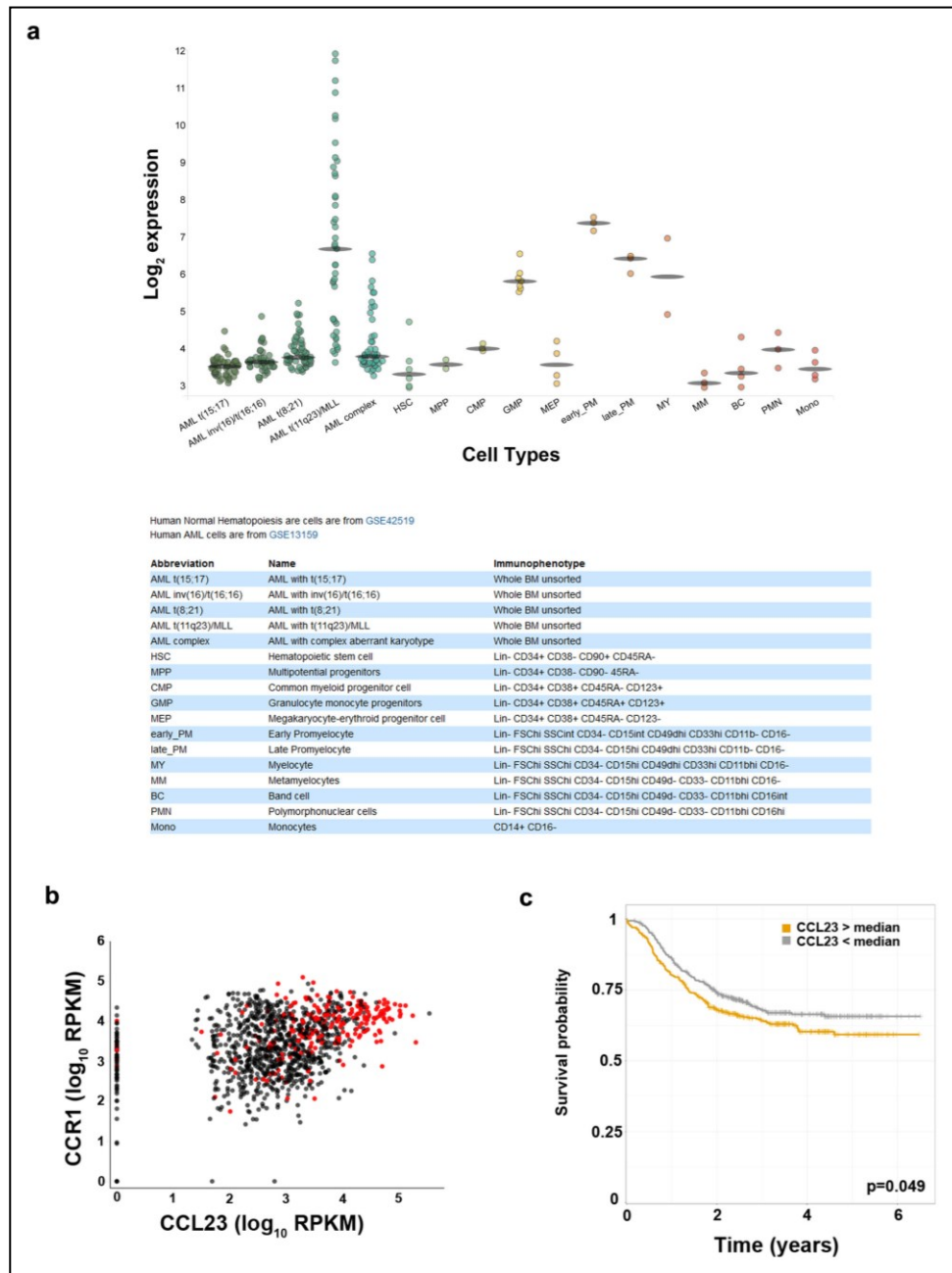
**(G)** UCSC snapshot showing an example of a gene with ATAC-sequencing peaks in CD34+KM3, mAML, CD34+ HSPC, hematopoietic stem cells and monocytes. The blue vertical line shows a peak that is unique to the mAML samples, whereas red vertical lines highlight common peaks that also overlap ENCODE transcription factor clusters.



**Figure 2.11 – Genome-wide distribution of histone modifications in model leukemias.**

(A) The spatial distribution of H3K4me3 and H3K79me2 was plotted across average genes in each of the three model leukemia stages, as well as by ranked enrichment in individual genes, using EaSeq package.

(B) Histone modifications across average genes are shown with genes binned into high, medium and low gene expression levels.



**Figure 2.12 – Expression of the *CCL23* gene in leukemic and normal blood cells.**

(A) Normalized gene expression data for *CCL23* from previously published studies of normal and leukemic samples is shown. The table in the lower panel details the sample abbreviations and the immunophenotype for each of the samples represented.

(B) Expression of *CCR1* and *CCL23* in the TARGET pediatric AML cohort, with *KMT2A*-translocated AML patients coloured in red.

(C) Kaplan-Meier survival curve based on expression of *CCL23* above/below the median from the TARGET AML cohort.

Chromosome	Start	End	Strand	Gene
chr8	9757574	9760839	+	LOC157627
chr2	219646472	219680016	+	CYP27A1
chr17	56270089	56282535	+	EPX
chr19	7733972	7735340	+	RETN
chr1	153362508	153363664	+	S100A8
chr19	7741943	7744719	+	C19orf59
chr1	182761394	182798588	+	NPL
chr3	46243200	46249832	+	CCR1
chr5	140011313	140013286	+	CD14
chr11	124933013	124960412	+	SLC37A2
chr19	55385549	55401839	+	FCAR
chr17	78194200	78227308	+	SLC26A11
chr9	120466460	120479766	+	TLR4
chr1	153330330	153333503	+	S100A9
chr2	218990013	219001976	+	CXCR2
chr13	111175413	111214071	+	RAB20
chr5	179220986	179223513	+	LTC4S
chr16	31271288	31344213	+	ITGAM
chr6	151815175	151942328	+	C6orf97
chr19	4537227	4540036	+	LRG1
chr2	138721808	138773934	+	HNMT
chr13	31309669	31338556	+	ALOX5AP
chr19	18284579	18288927	+	IFI30
chr19	51645558	51656783	+	SIGLEC7
chr19	52145806	52150132	+	SIGLEC14
chr5	122181160	122344902	+	SNX24
chr21	34442450	34444728	+	OLIG1
chr19	51628137	51639520	+	SIGLEC9
chr5	135364584	135399507	+	TGFBI
chr18	61582745	61602476	+	SERPINB10
chr14	88471468	88481155	+	GPR65
chr6	160390131	160527583	+	IGF2R
chr14	55595935	55612148	+	LGALS3
chr12	6898638	6929976	+	CD4
chr1	158801168	158819270	+	MNDA
chr3	122044011	122060815	+	CSTA
chr17	42422491	42430470	+	GRN
chr7	150497854	150502208	+	TMEM176A
chr16	31366509	31394318	+	ITGAX
chr17	80186293	80197369	+	SLC16A3
chr19	55174124	55179846	+	LILRB4
chr8	74903564	74941307	+	LY96
chr12	69742134	69748013	+	LYZ

Chromosome	Start	End	Strand	Gene
chr11	59824101	59838588	+	MS4A3
chr5	149432854	149492935	+	CSF1R
chr12	100967489	101018685	+	GAS2L3
chr9	80912059	80945009	+	PSAT1
chr3	169755735	169803183	+	GPR160
chr3	132036211	132087146	+	ACPP
chr19	36395303	36399211	+	TYROBP
chr1	161185087	161189038	+	FCER1G
chr14	65171193	65211060	+	PLEKHG3
chr5	150827163	150871940	+	SLC36A1
chr22	17597189	17602257	+	CECR6
chr11	72929344	72947395	+	P2RY2
chr14	25042724	25045466	+	CTSG
chr3	46395235	46402413	+	CCR2
chr17	78183079	78194199	+	SGSH
chr7	27220776	27224835	+	HOXA11
chrX	12924739	12941288	+	TLR8
chr19	852291	856246	+	ELANE
chr8	48649476	48650726	+	CEBPD
chr14	21249210	21250626	+	RNASE6
chr16	28505993	28510282	+	APOB48R
chr3	69788586	70017488	+	MITF
chr11	58976061	58980494	+	MPEG1
chr12	129338081	129469509	+	GLT1D1
chr16	50731050	50766987	+	NOD2
chr11	48002110	48192394	+	PTPRJ
chr5	82767493	82878122	+	VCAN
chr18	9708228	9862553	+	RAB31
chr16	85061410	85127828	+	KIAA0513
chr20	1874813	1920540	+	SIRPA
chr12	32655041	32798984	+	FGD4
chr19	827831	832017	+	AZU1
chr1	12227060	12269277	+	TNFRSF1B
chr17	55055468	55084129	+	SCPEP1
chr1	28473677	28520437	+	PTAFR
chr20	48807376	48809212	+	CEBPB
chr10	45455219	45490172	+	RASSF4
chr14	76044940	76114512	+	FLVCR2
chr5	53813589	53842416	+	SNX18
chr7	27225027	27228912	+	HOXA11-AS1
chr17	64298926	64806862	+	PRKCA
chr16	1988234	1993294	+	SEPX1
chr10	65281123	65384883	+	REEP3
chr4	15704573	15733796	+	BST1
chr22	17565851	17596584	+	IL17RA

Chromosome	Start	End	Strand	Gene
chr20	23614294	23618574	+	CST3
chr3	42695176	42709072	+	ZBTB47
chr19	50015536	50029685	+	FCGRT
chr5	73980969	74017030	+	HEXB
chr10	60094739	60130509	+	UBE2D1
chr7	76822688	76829150	+	FGL2
chr17	28705942	28796675	+	CPD
chr11	65543378	65547822	+	DKFZp761E198
chr7	48128355	48148330	+	UPP1
chr1	40361096	40367687	+	MYCL1
chr17	15848231	15879210	+	ADORA2B
chr11	60145958	60163426	+	MS4A7
chr19	13947401	13947473	+	MIR23A
chr5	66478104	66492617	+	CD180
chr11	61447905	61514474	+	DAGLA
chr21	34638672	34669520	+	IL10RB
chr2	219246752	219261617	+	SLC11A1
chr6	35310335	35395968	+	PPARD
chr1	109822176	109825790	+	PSRC1
chr17	47287589	47300587	+	ABI3
chr20	30639991	30689657	+	HCK
chr12	113659260	113736389	+	TPCN1
chr12	40618813	40763086	+	LRRK2
chr8	20054704	20079207	+	ATP6V1B2
chr5	49961733	50142356	+	PARP8
chr15	66994674	67074337	+	SMAD6
chr5	14581891	14616287	+	FAM105A
chr3	53195223	53226733	+	PRKCD
chr17	4901243	4931694	+	KIF1C
chr17	72690452	72709108	+	CD300LF
chr17	42634812	42638630	+	FZD2
chr11	10326642	10328923	+	ADM
chr19	16771938	16799816	+	TMEM38A
chr6	151773422	151791232	+	C6orf211
chr11	62104774	62160887	+	ASRGL1
chr10	65224989	65226322	+	LOC84989
chr1	150521898	150533412	+	ADAMTSL4
chr9	108006929	108153682	+	SLC44A1
chr3	71820806	71834357	+	PROK2
chr8	21766384	21771205	+	DOK2
chr11	117857106	117872199	+	IL10RA
chr13	49063099	49107316	+	RCBTB2
chr5	79783800	79838206	+	FAM151B
chr1	28199055	28213193	+	C1orf38
chr12	109015680	109027670	+	SELPLG

Chromosome	Start	End	Strand	Gene
chr16	84599204	84651669	+	COTL1
chr19	36393382	36395173	+	HCST
chr9	36136742	36163903	+	GLIPR2
chr21	45138978	45182188	+	PDXK
chr1	178063129	178448648	+	RASAL2
chr4	154605441	154627242	+	TLR2
chr5	148206156	148208197	+	ADRB2
chr1	154975112	154990999	+	ZBTB7B
chr4	7045156	7059677	+	TADA2B
chr6	3020390	3025005	+	LOC401233
chr1	153507076	153508717	+	S100A6
chr9	128509617	128729655	+	PBX3
chr12	10124008	10138194	+	CLEC12A
chr19	38810484	38819649	+	KCNK6
chr3	39304985	39323226	+	CX3CR1
chr12	56078395	56106089	+	ITGA7
chr1	11796142	11810828	+	AGTRAP
chr1	51434367	51440306	+	CDKN2C
chr11	76571917	76734850	+	ACER3
chr19	33685599	33699569	+	LRP3
chr11	94300474	94354587	+	PIWIL4
chr6	11538511	11583757	+	TMEM170B
chr22	38071613	38075809	+	LGALS1
chr17	41166622	41174459	+	VAT1
chr17	7461609	7463210	+	TNFSF13
chr5	173315331	173387313	+	CPEB4
chr3	171757418	172118492	+	FNDC3B
chr12	113860216	113876081	+	SDSL
chr10	82214038	82282391	+	TSPAN14
chr18	2847028	2914090	+	EMILIN2
chr18	46446223	46477081	+	SMAD7
chr2	111878491	111926022	+	BCL2L11
chr21	45773484	45862964	+	TRPM2
chr2	120124504	120130122	+	DBI
chr1	10459085	10480201	+	PGD
chr12	132312941	132336316	+	MMP17
chr14	75745481	75748937	+	FOS
chr1	154377669	154440188	+	IL6R
chr22	35695268	35743987	+	TOM1
chr6	6588934	6655216	+	LY86
chr11	67806462	67818366	+	TCIRG1
chr5	142150292	142608572	+	ARHGAP26
chr20	62526455	62567384	+	DNAJC5
chr15	77287465	77329671	+	PSTPIP1
chr10	64893007	64914786	+	NRBF2

Chromosome	Start	End	Strand	Gene
chr11	2466221	2870340	+	KCNQ1
chr15	63413999	63434260	+	LACTB
chr17	33570086	33594761	+	SLFN5
chr1	151584662	151671559	+	SNX27
chr1	53392901	53517289	+	SCP2
chr15	92396938	92715665	+	SLCO3A1
chr11	94277017	94283064	+	FUT4
chr19	33182867	33204702	+	NUDT19
chr18	47013413	47013644	+	C18orf32
chr21	37536839	37666572	+	DOPEY2
chr1	92940318	92952433	+	GFI1
chr9	115513134	115637267	+	SNX30
chr1	179851420	179889211	+	TOR1AIP1
chr10	75672043	75677258	+	PLAU
chr13	113344643	113541482	+	ATP11A
chr11	32851489	32876105	+	PRRG4
chr13	32605437	32870776	+	FRY
chr10	101419263	101468504	+	ENTPD7
chr20	58630980	58648008	+	C20orf197
chr2	160654643	160654766	+	CD302
chr11	14665269	14893604	+	PDE3B
chr9	94171327	94186144	+	NFIL3
chr1	151129105	151132223	+	TNFAIP8L2
chr16	28996147	29002104	+	LAT
chr19	840985	848175	+	PRTN3
chr12	125478194	125510349	+	BRI3BP
chr19	51728335	51743274	+	CD33
chr3	58318617	58410878	+	PXK
chr17	57697050	57774317	+	CLTC
chr1	201951766	201975275	+	RNPEP
chr22	38035684	38052050	+	SH3BP1
chr5	175085040	175113245	+	HRH2
chr18	61637263	61656608	+	SERPINB8
chr6	150070831	150132557	+	PCMT1
chr19	51874874	51875960	+	NKG7
chr2	48541795	48606434	+	FOXN2
chr19	33790840	33793430	+	CEBPA
chr9	140172280	140177093	+	C9orf167
chr1	36273828	36321188	+	EIF2C4
chr1	40420784	40435628	+	MFSD2A
chr8	8860314	8890849	+	ERI1
chr6	127588020	127609504	+	RNF146
chr5	56110900	56191978	+	MAP3K1
chr9	110247133	110252047	+	KLF4
chr19	10381517	10397291	+	ICAM1



Chromosome	Start	End	Strand	Gene
chr6	36973423	36996845	+	FGD2
chr12	4430359	4469190	+	C12orf5
chr11	64107690	64125006	+	CCDC88B
chr17	7076751	7082883	+	ASGR1
chr4	2794750	2842823	+	SH3BP2
chr9	139971953	139978990	+	UAP1L1
chr11	6232564	6255941	+	FAM160A2
chr14	103592664	103603776	+	TNFAIP2
chr16	30483983	30534506	+	ITGAL
chr16	2479395	2508859	+	CCNF
chr12	45609790	45834187	+	ANO6
chr6	31626075	31628549	+	C6orf47
chr7	38299244	38313248	+	TARP
chr7	143078360	143088204	+	ZYX
chr1	150524405	150524490	+	MIR4257
chr17	46184920	46200105	+	SNX11
chr4	148653453	148993927	+	ARHGAP10
chr9	132575221	132586441	+	TOR1A
chr3	196466728	196559518	+	PAK2
chr16	30960405	30966259	+	ORAI3
chr9	82186878	82341656	+	TLE4
chr3	190231840	190374986	+	IL1RAP
chr1	2160134	2241652	+	SKI
chr10	17270258	17279592	+	VIM
chr1	247579458	247612406	+	NLRP3
chr12	21590538	21624182	+	PYROXD1
chr11	73087405	73108519	+	RELT
chr16	68119269	68262449	+	NFATC3
chr19	6531010	6535939	+	TNFSF9
chr2	131113580	131132982	+	PTPN18
chr17	57918627	57918698	+	MIR21
chr1	111729801	111743281	+	DENND2D
chr4	177241090	177253396	+	SPCS3
chr4	106067943	106200958	+	TET2
chr12	121124949	121139667	+	MLEC
chr7	30323923	30407308	+	ZNRF2
chr3	52740049	52742197	+	SPCS1
chr1	15736391	15756839	+	EFHD2
chr9	100174302	100258405	+	TDRD7
chr9	140500096	140509783	+	ARRDC1
chr16	23847300	24231932	+	PRKCB
chr10	102790996	102800998	+	SFXN3
chr2	74425690	74442424	+	MTHFD2
chr5	148724977	148734146	+	GRPEL2
chr14	74960450	74962271	+	ISCA2

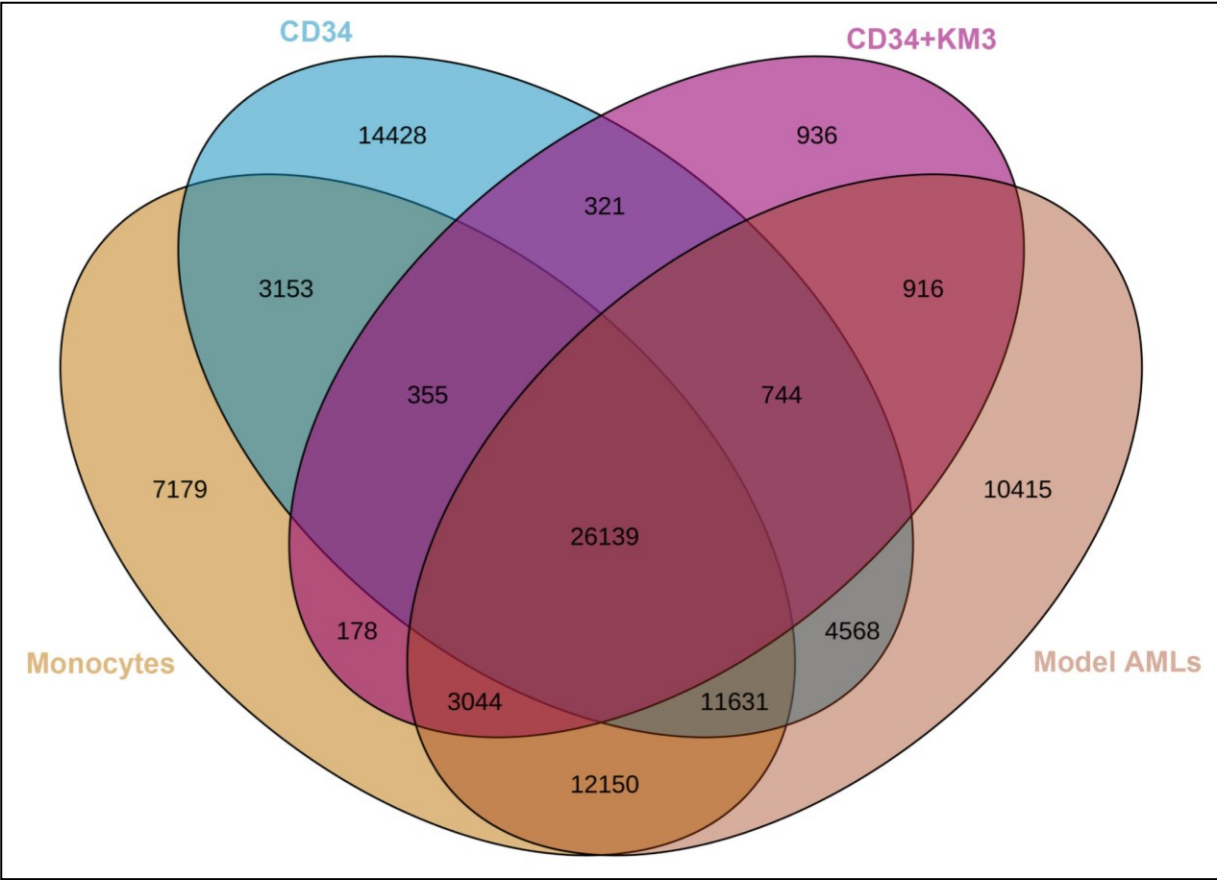
Chromosome	Start	End	Strand	Gene
chr19	16244838	16269377	+	HSH2D
chr16	21610856	21668792	+	METTL9
chr1	153516095	153518282	+	S100A4
chr15	45927256	45983479	+	SQRDL
chr22	50968838	50971008	+	ODF3B
chr19	12902310	12904125	+	JUNB
chr19	571325	583493	+	BSG
chr10	129705325	129884164	+	PTPRE
chr11	33279168	33375939	+	HIPK3
chr6	31465855	31478901	+	MICB
chr9	128024111	128127290	+	GAPVD1
chr11	63742079	63744015	+	COX8A
chrX	100663121	100663295	+	HNRNPH2
chr17	1619817	1641893	+	WDR81
chr16	50099881	50139375	+	HEATR3
chr17	21030258	21094836	+	DHRS7B
chr17	2496923	2588909	+	PFAFH1B1
chr22	44568836	44603035	+	PARVG
chr13	25875666	25916561	+	NUPL1
chr15	76135622	76193388	+	UBE2Q2
chr11	65343509	65360116	+	EHBP1L1
chr9	130823512	130829599	+	NAIF1
chr5	133984479	134063601	+	SEC24A
chr15	57592563	57599967	+	LOC283663
chr15	31619083	31670102	+	KLF13
chr9	139924626	139926449	+	FUT7
chr10	75757872	75879914	+	VCL
chr14	20937538	20946165	+	PNP
chr16	2261603	2264822	+	PGP
chr7	2273926	2281833	+	FTSJ2
chr2	234263153	234380743	+	DGKD
chr2	218923878	218926013	+	CXCR2P1
chr14	97263684	97347951	+	VRK1
chr16	11348274	11350039	+	SOCS1
chr12	66696356	66731958	+	HELB
chr15	32812049	32825942	+	LOC100288615
chr18	9913955	9960018	+	VAPA
chr2	28113557	28561767	+	BRE
chr12	54756229	54758270	+	GPR84
chr4	174292093	174298683	+	SAP30
chr16	31085743	31094744	+	ZNF646
chr19	33793763	33795963	+	LOC80054
chr5	134784558	134788089	+	TIFAB
chr20	30193092	30194313	+	ID1
chr17	34900737	34946276	+	GGNBP2

Chromosome	Start	End	Strand	Gene
chr4	2932288	2936586	+	MFSD10
chr13	73302042	73329539	+	C13orf34
chr16	4853204	4897303	+	GLYR1
chr1	179051112	179065129	+	TOR3A
chr1	38326369	38412729	+	INPP5B
chr16	3072626	3074287	+	HCFC1R1
chr1	28218049	28241236	+	RPA2
chr5	131630145	131679899	+	SLC22A4
chr7	156931655	157062066	+	UBE3C
chr6	26538572	26547164	+	HMGN4
chr15	85291818	85349663	+	ZNF592
chr5	141018869	141030986	+	FCHSD1
chr11	64073044	64084164	+	ESRRA
chr12	12870302	12875305	+	CDKN1B
chr22	21996542	21998588	+	SDF2L1
chr16	85932774	85956211	+	IRF8
chr12	46123620	46301819	+	ARID2
chr1	109756515	109780804	+	SARS

**Table 2.2 – Genes with correlated gene expression and active histone modifications.**

#### 2.4.3.4 KM3-mediated epigenetic changes are minimal relative to normal blood cells.

Location analysis of the annotated ATAC-sequencing peaks unique to the model AML cells revealed the expected peaks at promoters of active genes, although the majority of open chromatin regions were either intergenic or intronic (**Figure 2.10d**). As a point of comparison for our ATAC-sequencing data, we downloaded previously published and normalized ATAC-sequencing data generated from normal blood cells (Corces, Buenrostro et al. 2016) along with raw ATAC-sequencing data from normal CD34<sup>+</sup> cells and monocytes from the BLUEPRINT project (Stunnenberg, International Human Epigenome et al. 2016). Using peaks identified using MACS2 in our ATAC-sequencing data or remapped published data, or through supervised thresholding of normalized UCSC tracks, we compared regions of open chromatin in normal and leukemic blood cells. As with the histone modifications, the majority of peaks identified in the model leukemia samples were shared with one or more normal types of blood cell. Considering the 10,415 peaks identified by MACS2 that were specific to the model AML (**Figure 2.13**), analysis of potential transcription factor binding motifs once again identified factors, such as PU.1, that are critical for normal myeloid cell differentiation (**Figure 2.14**). Hierarchical clustering of all ATAC-sequencing peaks from the CD34+KM3 and model AML revealed that the chromatin landscape in the model AML is close to that of normal monocytes and CD34<sup>+</sup> HSPC (**Figure 2.10e**). The clear separation of CD34+KM3 cells, despite the small absolute number of unique peaks (and more peaks common to CD34<sup>+</sup> HPSC than to monocytes), suggests that the chromatin of these cells is subtly impacted by their culture *in vitro*.



**Figure 2.13 – Overlapping peaks in ATAC-seq data from normal and leukemic cells.**

Venn diagram shows the number of peaks detected by MACS2 in normal blood cells (monocytes, CD34<sup>+</sup> cells) or model leukemia stages (CD34+KM3, mAML).

Rank	Motif	Name	P-value	log P-p-value	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif
1		SpiB(ETS) OCILY3-SPIB-ChIP-Seq(GSE56837)/Homer	1e-1014	-2.336e+03	0.0000	1917.0	18.41%	965.9	2.48%
2		PU.1(ETS) ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-986	-2.272e+03	0.0000	2464.0	23.66%	1791.8	4.59%
3		ELF3(ETS) T47D-ELF3-ChIP-Seq(GSE30407)/Homer	1e-815	-1.879e+03	0.0000	2771.0	26.61%	2751.2	7.05%
4		PU.1-IRF8(ETS-IRF) pDC-IrF8-ChIP-Seq(GSE66899)/Homer	1e-676	-1.559e+03	0.0000	1443.0	13.86%	833.7	2.14%
5		ETS1(ETS) Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-604	-1.393e+03	0.0000	2792.0	26.81%	3510.2	9.00%
6		IRF8(IRF) BMDM-IRF8-ChIP-Seq(GSE77884)/Homer	1e-572	-1.310e+03	0.0000	1726.0	16.57%	1473.7	3.78%
7		ELF3(ETS) PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-571	-1.316e+03	0.0000	2416.0	23.20%	2812.1	7.21%
8		EHF(ETS) LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-549	-1.265e+03	0.0000	3317.0	31.85%	5024.3	12.88%
9		JunB(ZIP) DendriticCells-JunB-ChIP-Seq(GSE36099)/Homer	1e-533	-1.228e+03	0.0000	1895.0	18.19%	1885.8	4.83%
10		ETV1(ETS) GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	1e-528	-1.218e+03	0.0000	3147.0	30.22%	4698.4	12.04%
11		Fra1(ZIP) BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-527	-1.215e+03	0.0000	1888.0	18.13%	1888.7	4.84%
12		BA1F(ZIP) Th17-BA1F-ChIP-Seq(GSE39756)/Homer	1e-527	-1.214e+03	0.0000	2099.0	20.15%	2306.3	5.91%
13		Atf3(ZIP) GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-527	-1.214e+03	0.0000	2120.0	20.36%	2349.8	6.02%
14		AP-1(ZIP) ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-514	-1.184e+03	0.0000	2227.0	21.38%	2618.8	6.71%
15		Fra2(ZIP) Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-494	-1.139e+03	0.0000	1663.0	15.97%	1564.0	4.01%
16		Etv2(ETS) ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	1e-463	-1.067e+03	0.0000	2448.0	23.50%	3318.9	8.51%
17		PU.1-IRF(ETS-IRF) Beutl-PU.1-ChIP-Seq(GSE21512)/Homer	1e-453	-1.045e+03	0.0000	3438.0	33.01%	5878.9	15.07%
18		Flt1(ETS) CD8-FL1-ChIP-Seq(GSE20898)/Homer	1e-452	-1.041e+03	0.0000	2558.0	24.56%	3628.8	9.30%
19		GABPA(ETS) Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	1e-451	-1.039e+03	0.0000	2245.0	21.56%	2905.8	7.45%
20		ERG(ETS) VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-449	-1.034e+03	0.0000	3421.0	32.85%	5862.2	15.02%
21		Foxl2(ZIP) 3T3L1-Foxl2-ChIP-Seq(GSE56872)/Homer	1e-431	-9.941e+02	0.0000	1256.0	12.06%	1016.8	2.61%
22		Jun-AP1(ZIP) K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-370	-8.536e+02	0.0000	998.0	9.58%	739.7	1.90%
23		CTCF(Zf) CD4+ CTCF-ChIP-Seq(Barski_et_al)/Homer	1e-354	-8.173e+02	0.0000	579.0	5.56%	222.5	0.57%
24		IRF3(IRF) BMDM-IrF3-ChIP-Seq(GSE67343)/Homer	1e-305	-7.023e+02	0.0000	1333.0	12.80%	1527.2	3.91%
25		EWS-ERG-fusion(ETS) CAD0_ES1-EWS-ERG-ChIP-Seq(SRA014231)/Homer	1e-296	-6.817e+02	0.0000	1892.0	18.17%	2813.9	7.21%

**Figure 2.14 – Predicted transcription factor binding sites in ATAC-seq peaks.**

The table presents the predictive output of Homer for the top 25 transcription factors whose binding motifs are over-represented in the ATAC-seq peaks that are uniquely found in the model leukemia samples.

Using data from the ENCODE consortium, we further examined the transcription factors that have been experimentally observed to bind to leukemia-specific open chromatin regions. For this, we ranked the abundance of individual factors in the entire dataset, and in the leukemia-specific peaks, to account for the difference in frequencies of binding sites for individual factors (**Figure 2.10f**). This analysis revealed an enrichment of factors in the leukemia-specific peaks that did not necessarily correlate with an increase in expression of the factor. Given the ChIP-sequencing data, this suggests that the novel, leukemia-specific peaks may create competition for the binding of a range of expressed transcription factors, potentially co-opting their activity. In other cases, while there was no difference in the ranking, the factor itself was more highly expressed; in cases in which such factors have been demonstrated to be essential for leukemia development (e.g. FOS (Assi, Imperato et al. 2019)), the higher expression level may be important to ensure a high/continuous occupancy at the open regions in order to drive the leukemic transcriptional program.

Across the genome, we observed many examples of genomic regions in which ATAC-peaks common to both leukemic and normal cells were seen in the vicinity of other leukemia-specific peaks (**Figure 2.10g**). Given our observations, we questioned whether all of the leukemia-specific peaks we observed were potentially functional or whether they might represent stochastic changes in accessibility. To address this question, we compared data for chromatin accessibility (DNaseI hypersensitive sites) and ChIP-sequencing data from the ENCODE consortium to the ATAC-sequencing peaks that were unique to the model leukemias. Interestingly, the peaks determined by either MACS2 or supervised thresholding showed high levels of overlap (65% and 75%, respectively) with both previously defined transcription factor clusters, DNaseI hypersensitive sites and overlapped chromatin marks associated with enhancers seen in other cell types. These data argue that despite their unique presence in leukemic cells, the large majority of leukemia-specific open chromatin regions likely represent enhancers.

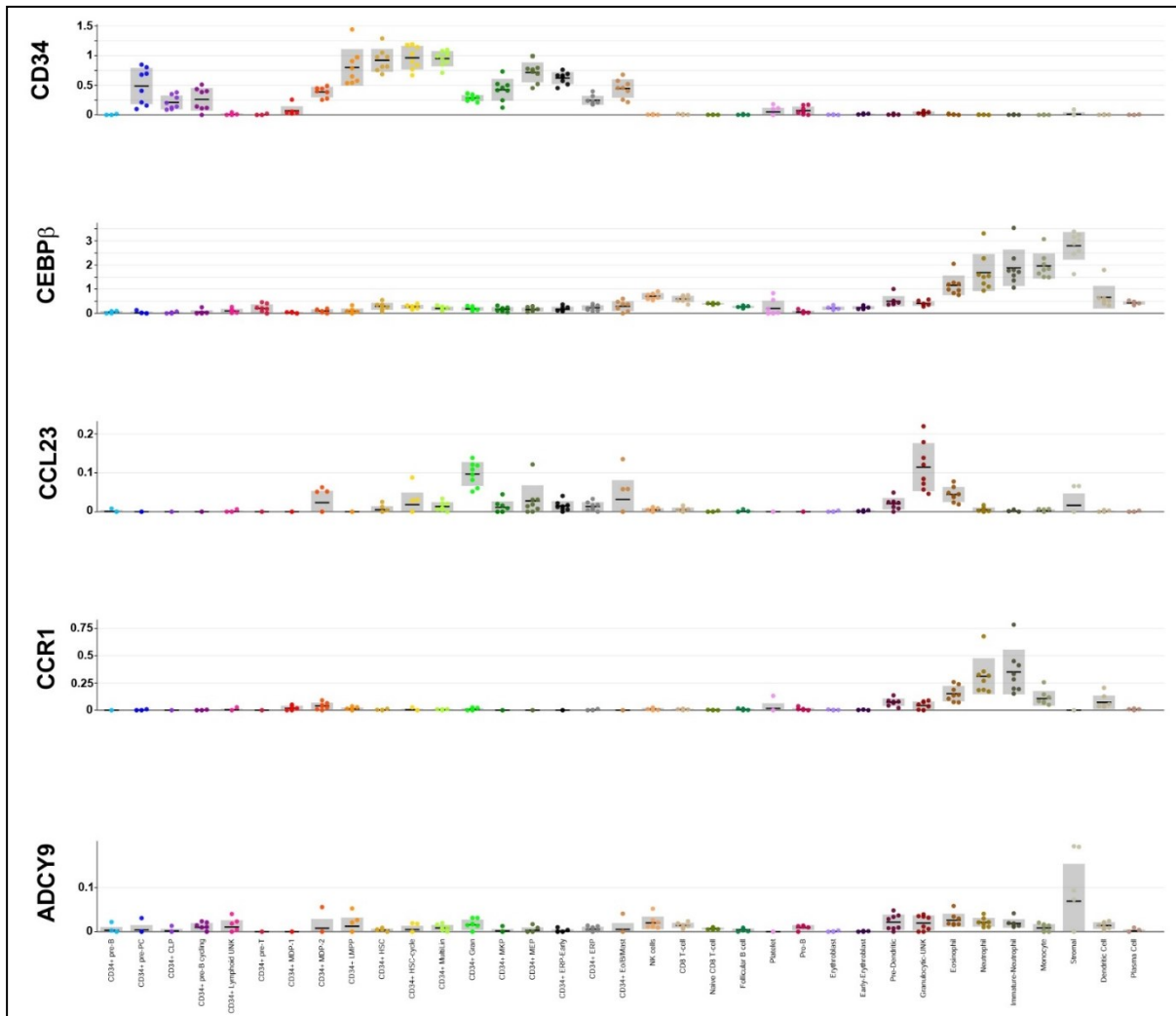
## 2.4.4 Discussion

The *KMT2A-MLL3* gene fusion is one of the most common genetic anomalies found in pediatric AML and is an epigenetic regulator that is essential, directly or indirectly, for the alterations of chromatin structure (Guenther, Lawton et al. 2008, Bernt, Zhu et al. 2011, Garcia-Cuellar, Buttner et al. 2016) and DNA methylation seen in the disease (Erfurth, Popovic et al. 2008). Despite its extensive characterization, less is known regarding its role in the initial transformation of HSPC into leukemic cells. Our current study leverages our previously described human model leukemia system to integrate epigenetic and expression changes to better characterize these changes. With respect to DNA methylation changes, our data confirm previously published results (Alvarez, Suela et al. 2010) showing that an aberrant hypomethylation profile is found in KM3 leukemias. Interestingly, relative to the initial CD34<sup>+</sup> cells, CD34<sup>+</sup>KM3 and B-cell ALL have a hypermethylation phenotype, whereas the model and patients' AML show the reverse. While the significance of all of these global differences is not yet clear, previous single cell studies showed that differentially methylated promoter regions in normal blood cells were enriched in binding sites for transcription factors promoting myeloid differentiation (Farlik, Halbritter et al. 2016). Our data (e.g **Figure 2.9**) agree with this finding, which may reflect a default lymphoid differentiation pathway in the absence of demethylation of these regions. Recent studies have highlighted the importance of DNA methyltransferase in KM3-mediated leukemogenesis in which *DNMT3A* and *DNMT3B* loss-of-function mutations accelerate leukemia progression by enhancing cell cycle progression (Zheng, Zhang et al. 2016). Forced expression of *DNMT3B* delayed KM3 leukemia development (Schulze, Rohde et al. 2016) whereas the deletion of *DNMT1* prevents KM3 AML progression (Trowbridge, Sinha et al. 2012). These findings and our own data demonstrate that the loss of *de novo* methylation, but not necessarily maintenance methylation, is required for the induction of AML.

Analysis of the correlated epigenetic and gene expression changes highlighted a number of genes of interest, including *ADCY9*, an atypical member of the adenylate cyclase family. *ADCY9* shows low but relatively broad expression in most tissues except for blood (of all lineages (Hay, Ferchen et al. 2018); **Figure 2.15**), and is capable of inducing cAMP accumulation (Cumbay and Watts 2004), although its activation with G protein subunits appears to be cell-type specific. Our functional studies demonstrate that *ADCY9* function is critical for the growth of KM3-



AML cells, and levels of cAMP in leukemias (Gausdal, Wergeland et al. 2013, Perez, Smagley et al. 2016) have been shown to be important. Nevertheless, whether the growth defects seen in response to the loss of *ADCY9* expression are mediated through its ability to produce cAMP remains unclear and will require further study. The integration of our data also highlighted the potential importance of the chemokine receptor gene *CCR1* and one of its ligands, *CCL23*. This receptor-ligand pair was noteworthy in part because of their strong level of upregulation, but also because the expression of both increased immediately after the addition of the gene fusion, prior to the complete transformation of the cells. The *CCR1* receptor can recognize at least ten different ligands including *CCL3* and *CCL5*, which are both ubiquitously expressed in normal blood cells. In contrast, *CCL23* is only expressed at very low levels in promyelocytes (Bagger, Kinalis et al. 2019) ( $\text{Lin}^- \text{FSC}^{\text{hi}} \text{SSC}^{\text{int}} \text{CD34}^- \text{CD15}^{\text{int}} \text{CD49d}^{\text{hi}} \text{CD33}^{\text{hi}} \text{CD11b}^- \text{CD16}^-$ ) and granulocyte-monocyte progenitors (Bagger, Kinalis et al. 2019) ( $\text{Lin}^- \text{CD34}^+ \text{CD38}^+ \text{CD45RA}^+ \text{CD123}^+$ ) but is highly expressed in *KMT2A*-AML according to a number of independent datasets. The impact of *CCR1* signaling in AML remains poorly defined, although one consequence appears to be activation of the  $\text{G}\alpha_i$  pathway resulting in the inhibition of cAMP production (Gilliland, Salanga et al. 2013). In this context, it is interesting to note that *ADCY9* is specifically upregulated in *KMT2A*-AML (**Figure 2.6f**), potentially as counterbalance to the activity of *CCR1* and  $\text{G}\alpha_i$  and to maintain cAMP levels.



**Figure 2.15 – Single-cell expression of different genes in normal blood cells.**

Published single-cell RNA-sequencing data from 35 different cell types shows the expression of 4 genes highlighted in this paper (*CEBPβ*, *CCR1*, *CCL23*, *ADCY9*) as well as *CD34* as a cell-type specific expression control. Graphs of expression were generated by the Human Cell Atlas interactive web portal (Hay, Ferchen et al. 2018) and are automatically scaled to the range of expression values for each gene.

Perhaps most surprisingly with respect to chromatin structure and accessibility, while we expected to see widespread changes, in fact there were relatively few modifications specific to the model leukemia cells compared to normal blood cells. This observation agrees with those of a recent study showing that *KMT2A-MLLT3* helps to preserve gene expression of the cellular states in which it is expressed (Chen, Burkhardt et al. 2019). As noted, most of the leukemia-specific regions overlap open chromatin regions and transcript factor clusters seen in large-scale ENCODE datasets. Although speculative, these observations suggest the interesting hypothesis there may be a finite number of regions in the genome that are competent to be open *in vivo* across all cell types. The addition of the KM3 fusion, while activating the expression of a subset of genes, may also cause re-localization of chromatin, leading to the opening of regions not normally seen in blood cells. Thus, while chromatin remodeling is required to give access to endogenously expressed transcription factors in some of the novel locations that we characterized in our study, others may simply be a response to this remodeling. The distribution of ATAC peaks that we observe, largely in introns and intergenic regions (**Figure 2.10d**), also raises the question of the potential biological role as novel enhancers. Similar distributions in ATAC-sequencing data have been noted (Yan, Powell et al. 2020) and intronic enhancers have been shown to regulate the expression of genes in a tissue-specific manner (Borsari, Villegas-Mirón et al. 2020), which may explain why many of the ATAC-sequencing peaks not present in blood cells nevertheless corresponded to open chromatin regions in other cell types. These results open a new avenue to look at some of these regions of potential interest in the context of AML, to better understand their role in enabling access to genes which may be essential for leukemic growth.

In conclusion, here we have leveraged the advantages of our model system to characterize the immediate impact of the KM3 fusion without the complicating factor of additional mutations and other genetics abnormalities typically found in patients' genomes. While our results highlight the subtlety of the changes that occur once the oncogene is introduced, they also support the role of previous genes implicated in leukemogenesis and highlight a novel receptor/ligand pair with potential functional relevance for *KMT2A*-AML. The novel insight into the nature and extent of epigenetic reprogramming that occurs in the disease will help to guide future studies to identify novel potential therapeutic targets.

## 2.4.5 Methods

### 2.4.5.1 Patients' samples and model leukemia generation

All pediatric and adult AML patients' samples and additional clinical information used in this study were collected by the Banque de Cellules Leucémiques du Québec (BCLQ) with informed consent and project approval from the Research Ethics Boards of CHU Ste-Justine, Maisonneuve-Rosemont Hospital and Université de Montréal. RNA was extracted from samples preserved in TRIzol (Invitrogen) according to standard protocols, while DNA for methylation-sequencing (methyl-sequencing) analysis was obtained from viably conserved samples in dimethylsulfoxide (DMSO). The characteristics of the patients' AML samples are shown in **Table 2.3**. The model leukemias used in this study were generated as previously described (Barabe, Gil et al. 2017).

ID	Age	FAB	Blast % (BM)	Karyotype
Ped_159	1 yr	M5A	70	46,XX,der(9)t(9;11)(p22;q23)del(9)(p21)[20]
Ped_111	5 yr	M5A	85	48,XX,+8,+8,t(9;11)(p22;q23)[20]
Ped_144	13 yr	M5A	90	46,XX,t(9;11)(p22;q23)[5]/47,XX,+8,t(9;11)(p22;q23)[16]

**Table 2.3 – Summary of pediatric patients in this study.**

#### 2.4.5.2 Human CD34<sup>+</sup> hematopoietic stem and progenitor cell isolation and cell culture

CD34<sup>+</sup> human HSPC were isolated from fresh umbilical cord blood, collected under informed consent at Sainte-Justine pediatric hospital (distributed by HémaQuébec's Public Cord Blood Bank (Montreal, Quebec, Canada)) or CHU de Québec and Hotel Dieu de Levis. CD34<sup>+</sup> cord blood cells were positively selected as previously described (Barabe, Gil et al. 2017) and processed immediately without *in vitro* culture. Briefly, cord blood was diluted with phosphate-buffered saline (PBS) + citrate 0.6% (1:2) and mixed with Ficoll. After spinning the mixture at 2,100 rpm, for 30 minutes at room temperature with the lowest deceleration speed, white blood cells found at the interphase were collected and diluted with PBS (1:2). Cells were then spun at 2,100 rpm for 10 minutes and resuspended in PBS + citrate 0.6% + bovine serum albumin (BSA) 0.5%. An isolation kit (Miltenyi Biotec; #130-100-453) composed of beads recognizing CD34<sup>+</sup> cells was used according to the manufacturer's protocol. The PosseldS method on an AutoMACS<sup>®</sup> machine was then used to isolate CD34<sup>+</sup> cells and the final yield was quantified using a hemocytometer. Leukemia cells lines (KG1a, THP-1, NOMO1 and MOLM-13) were grown in RPMI medium supplemented with 10% heat-inactivated fetal bovine serum, penicillin (100 U/ml) and streptomycin (100 µg/ml) at 37°C in 5% CO<sub>2</sub>. Cell lines were transfected using a lentiviral protocol as previously described (Barabe, Gil et al. 2017). The shRNA sequences used are shown in **Table 2.4**.

TARGET	SEQUENCE (5'->3')
SCRAMBLE	TGCTGTTGACAGTGAGCGCCCGCCTGAAGTCTCTGATTAATAGTGAAGCCACAGATGTATTAA TCAGAGACTTCAGGCGGTTGCCTACTGCCTCGGA
ADCY9 – SHRNA1	TGCTGTTGACAGTGAGCGCGACTGTCAAACCTTTGATAATAGTGAAGCCACAGATGTATTAT CAAAGGTTTTGACAGTCTTGCCCTACTGCCTCGGA
ADCY9 – SHRNA2	TGCTGTTGACAGTGAGCGCGGGTATTATTTGACTTTTAGATAGTGAAGCCACAGATGTATCTA AAAGTCAAATAATACCCATGCCTACTGCCTCGGA

**Table 2.4 – List of primers used for shRNA-mediated knockdown of *ADCY9* expression.**

### 2.4.5.3 Chromatin immunoprecipitation, RNA-sequencing and ATAC-sequencing

Cells were cross-linked using formaldehyde 1% (7 min, room temperature) before glycine quenching (0.125 nM, 5 min, room temperature), cell lysis and sonication (Covaris S2 sonicator). Protein G-coupled beads (Protein G Dynabeads™) were incubated overnight on a rotator at 4°C with 3 µg of antibodies against different histone marks (H3: #ab1791, H3K4me3: #ab8580-100, H3K79me2: #ab3594). Sonicated chromatin was incubated with 250 µL of beads for 4 hours on a rotator at 4°C before sequential washing with: ‘Low Salt’ Buffer (0.5% NP40, 15 mM KCl, 10 mM Tris pH 8.0, 1 mM EDTA), three different ‘High Salt’ buffers (0.5% Triton, 10 mM Tris pH 8.0, 100 mM (2) or 400 mM (3) or 500 mM (4) NaCl), twice in LiCl Buffer (0.5% NP40, 250 mM LiCl, 10 mM Tris pH 8.0, 1 mM EDTA). ChIPed material was then eluted (buffer: 10 mM Tris pH8.0, 1 mM EDTA) and incubated overnight at 65°C with shaking (1,200 rpm) to reverse-crosslink. An RNaseA (for 1h at 37°C) and Proteinase K digestion was performed (1 h at 37°C) before DNA was purified by phenol-chloroform. A TruSeq ChIP Library Preparation Kit (Illumina, #IP-202-9001) was used to construct DNA libraries as described in the manufacturer’s instructions with polymerase chain reaction (PCR) for a total of 12 cycles using the Illumina indexed library primers. An Illumina NextSeq 500 was used to sequence samples (75bp paired-end) according to standard Illumina protocols. For all RNA-sequencing experiments, RNA was purified from cells ( $1 \times 10^6$ ) in TRIzol reagent, purified on RNeasy columns (Qiagen) and then used as input for Illumina stranded Tru-seq protocols. For CD34<sup>+</sup> cells transduced with the KM3 fusion and cultured *in vitro*, RNA was collected after 30 days from GFP<sup>+</sup> cells while for all shRNA knockdown experiments RNA was collected 4 days after lentiviral transduction. Lastly, RNA from the xenografted model AML was collected from the bone marrow of mice sacrificed 24-30 weeks after injection. All sequencing in this study was performed by the genomics core facility at the Institute for Research in Immunology and Cancer. The ATAC-sequencing protocol used was based on a method published by Buenrostro et al (Buenrostro, Wu et al. 2015) and was performed using 50,000 cells. Samples were sequenced (75bp paired-end) according to Illumina protocols. The primers used are shown in **Table 2.5**.

PRIMER NAME	SEQUENCE (5'->3')
AD1_NOMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
AD2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTTCGTGGGCTCGGAGATGT
AD2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
AD2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
AD2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
AD2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT

**Table 2.5 – List of primers used to prepare assay for transposase-accessible chromatin (ATAC)-sequencing libraries.**

#### 2.4.5.4 DNA methylation capture sequencing

Genomic DNA from cells pellets from either KM3 patients' samples, pooled cord blood donors (freshly isolated), cultured cells (collected after 30 days) or model AML cells (collected 24 weeks after injection) was isolated using the PureLink Genomic DNA Mini Kit according to the manufacturer's protocol (Life Technologies, cat. #K1820-01). Three micrograms of DNA from each sample were fragmented to 140-180 bp DNA fragments using Covaris S2 (parameters: 10% duty cycle at intensity 5 for 6 cycles of 60 s with 200 cycles set at sweeping mode). Methyl-sequencing libraries for DNA methylation analysis were prepared using the SureSelectXT Human Methyl-Seq Target enrichment system (Agilent Technologies, cat. #5190-4661) according to the manufacturer's instructions. Briefly, DNA was fragmented, purified and used for end-repair and adenylated prior to ligation of methylation adaptors (synthesized with 5'-methylcytosine instead of cytosine). The adaptor-ligated DNA samples were purified using AMPure XP beads (Beckman Coulter Genomics, cat. #A63880). Nine hundred nanograms of each DNA sample was hybridized with Agilent SureSelect Methyl-Seq biotinylated RNA baits for 24 h at 65°C. The libraries were captured using Dynal MyOne Streptavidin T1 magnetic beads (Invitrogen, cat. #65601). The captured DNA was subjected to bisulfite conversion with the EZ DNA Methylation Gold Kit (Zymo Research, cat. #D5005) as described in the manufacturer's instructions followed by PCR for a total of 16 cycles using the Illumina indexed library primers. The final libraries were quantified by 2100 Bioanalyzer chips (Agilent technologies) with quality control on a MiSeq (Illumina) sequencer. Libraries were

sequenced on multiple lanes of a HiSeq2000 flow cell at the IRIC genomics platform and the FastQC R package was used to assess the quality of the reads.

#### **2.4.5.5 Bioinformatics analysis of ChIP-sequencing and ATAC-sequencing**

The quality of the raw sequence data was checked by FastQC (version 0.11.4). Trimmomatic (version 0.32) was used to remove adapters and keep reads with quality scores >30. Paired-end sequencing reads were then mapped to the human reference genome (GRCH37/hg19) using BWA (version 0.7.10) with the default parameters. PCR duplicates were removed by Samtools (version 1.9). For ATAC-sequencing data, reads from mitochondrial DNA were removed (samtools idxstats input.bam | cut -f 1 | grep -v MT | xargs samtools view -b input.bam > output.bam). The EaSeq software package (version 1.04) (Lerdrup, Johansen et al. 2016) was used to analyze data and integrate gene expression and data from epigenetic marks from chromatin immunoprecipitation (ChIP)-sequencing experiments. To calculate the enrichment of each histone mark, the analysis focused on specific regions in the genome with a distance from the Transcription Start Site (TSS) of -1000 to +3000pb for H3K79me2, and -2000 to +2000pb for H3K4me3. MACS2 (v2.2.4) (Thomas, Thomas et al. 2017) was the peak calling algorithm used to determine the list of significant enrichment peaks, following the default parameters (callpeak -f BAMPE -g hs -q 0.05). Intervene (Khan and Mathelier 2017) was used to find peaks in common in our datasets (bedtools intersect -a -b -f 0,50 -r) and generate Venn diagrams (Pybedtools). A ngs.plot (Shen, Shao et al. 2014) was used to visualize global repartition of mapped reads across the genome. Previously published data for lineage-specific ATAC-sequencing in a range of normal blood cells (Corces, Buenrostro et al. 2016) was accessed through the UCSC track hub ([https://s3-us-west-1.amazonaws.com/chang-public-data/2016\\_NatGen\\_ATAC-AML/hub.txt](https://s3-us-west-1.amazonaws.com/chang-public-data/2016_NatGen_ATAC-AML/hub.txt)). DNase HSS and TF clusters generated from the ENCODE consortium V3 were analyzed by custom Perl Scripts. Public ATAC-sequencing peaks for all cell types were identified using a threshold of a normalized score >60 for all tracks for starts/ends of peaks. Fastq files for BLUEPRINT ATAC-sequencing data (accessions EGAD00001002710 and EGAD00001002709) were downloaded, remapped and analyzed with MACS2 as described above. Gene Set Enrichment Analysis (GSEA) was performed using software (version 3.0) from the Broad Institute (Subramanian, Tamayo et al. 2005) and a pre-



ranked list based on significantly differentially expressed genes between CD34<sup>+</sup> and CD34+KM3 cells using the R package DESeq2 (Love, Huber et al. 2014).

#### **2.4.5.6 Mapping methyl-sequencing reads and methylation analysis**

Paired end sequencing reads (2x100 bp) were mapped to the human reference genome (GRCh37/hg19) using Bismark with a directional alignment mode and Bowtie2 as the alignment process. All putative duplicate reads were removed by previously described Perl scripts (Krueger and Andrews 2011). The methylation calls were extracted with Bismark-methylation-extractor and the analysis was performed using custom python scripts, R package MethylKit 0.9.2 and SeqMonk 1.45.2 as the visualization tool ([www.bioinformatics.babraham.ac.uk/projects/seqmonk/](http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/)). Hierarchical clustering of the samples based on the differentially methylated cytosines was performed using the computed Pearson correlation distance between samples. For each cell type (CD34<sup>+</sup>, CD34+KM3, KM3-AML model) the DNA methylation data corresponded to the combination of methyl-sequencing libraries from two biological replicates. Analysis of methylated cytosines was restricted to positions covered by a minimum of eight reads coverage for each biological replicate. To identify differential methylated cytosines (DMC) a logistic regression model was applied to compare the fraction of methylated cytosines across the three types of cells and a sliding linear model (SLIM) method was used to correct *P*-values for multiple testing and derive *q*-values (Wang, Tuominen et al. 2011). To define the DMC, we specified two parameters: *q*-values below 0.01 and a threshold of percentage methylation difference  $\leq 40\%$ . All functional annotations for genomic regions were defined based on GRCh37/hg19 assembly download from UCSC website. Gene bodies were defined as the transcribed regions from the start to the end of transcription sites for each RefSeq entry and promoter regions were defined as 1 kb of sequence upstream of the TSS for each RefSeq transcript. Annotations for regulatory elements were defined as CpG islands, TFBS (TFBS conserved sites), enhancers (Vista Enhancers (Pennacchio, Ahituv et al. 2006)), and DNase I hyper sensitive sites.

#### **2.4.5.7 Data-sharing statement**

All sequencing data for this manuscript has been deposited in the Gene Expression Omnibus and is accessible through the following accession numbers: GSE142579 and GSE70755.

## **2.4.6 Acknowledgments**

This work was supported by grants from the Cole Foundation (TM, fellowship), the Impact Grant of the Canadian Cancer Society in collaboration with The Cole Foundation, Molson Foundation, R. Howard Webster Foundation, Mirella and Lino Saputo Foundation, Fonds de recherche du Quebec - Santé, Faculté de Médecine, Université de Montréal, Letko Brosseau, Birks Family Foundation, Maryse and William Brock, CHU Sainte-Justine Foundation, Montreal Children's Hospital Foundation, Morris and Rosalind Goodman Family Foundation, Zeller Family Foundation, David H. Laidley Foundation, Drummond Foundation, and the Henry and Berenice Kaufmann Foundation (grant 705047-IMP-17; BW, FB, SC), and the Leukemia & Lymphoma Society of Canada (to FB). Human leukemia specimens were collected and analyzed by the Banque de Cellules Leucémiques du Québec (BCLQ), supported by the Cancer Research Network of the Fonds de Recherche du Québec en Santé. We would like to thank the clinicians and nurses at CHU Sainte-Justine, CHU de Québec, Hotel Dieu de Lévis for cord blood collection and we also wish to acknowledge the contribution of all of the courageous patients who provided samples used in this study. Finally, we would like to acknowledge Jennifer Huber, Sarah Boissel, Raphaëlle Lambert, Danièle Gagné and Annie Gosselin for expert technical assistance with experimental work.

All authors discussed the results and participated in the revision of the manuscript and declare that they have no conflict of interest.

## **2.4.7 Author contributions**

BTW and TM designed the study, analyzed data and drafted the manuscript. TM and KL performed ChIP-seq and ATAC-seq, analysed the data and generated figures. MC performed Methyl-seq experiments, analysed data, generated figures and drafted the manuscript. SS-T-H assisted with the ATAC-seq analysis and generation of figures. KL, ER, EB, and AB helped with molecular biology work, cloning and cell culture. EB, AB and FB generated the model on mice from CD34+ cells. SM provided expression and clinical data for patient samples and JH and SC collected, characterized, and banked the local patient samples used in this study.

## 2.4.8 References

- Agger, K., S. Miyagi, M. T. Pedersen, S. M. Kooistra, J. V. Johansen and K. Helin (2016). "Jmjd2/Kdm4 demethylases are required for expression of *Il3ra* and survival of acute myeloid leukemia cells." *Genes Dev* **30**(11): 1278-1288.
- Akalin, A., F. E. Garrett-Bakelman, M. Kormaksson, J. Busuttil, L. Zhang, I. Khrebtukova, T. A. Milne, Y. Huang, D. Biswas, J. L. Hess, C. D. Allis, R. G. Roeder, P. J. Valk, B. Lowenberg, R. Delwel, H. F. Fernandez, E. Paietta, M. S. Tallman, G. P. Schroth, C. E. Mason, A. Melnick and M. E. Figueroa (2012). "Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia." *PLoS Genet* **8**(6): e1002781.
- Alvarez, S., J. Suela, A. Valencia, A. Fernandez, M. Wunderlich, X. Agirre, F. Prosper, J. I. Martin-Subero, A. Maiques, F. Acquadro, S. Rodriguez Perales, M. J. Calasanz, J. Roman-Gomez, R. Siebert, J. C. Mulloy, J. Cervera, M. A. Sanz, M. Esteller and J. C. Cigudosa (2010). "DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia." *PLoS One* **5**(8): e12197.
- Andersson, A. K., J. Ma, J. Wang, X. Chen, A. L. Gedman, J. Dang, J. Nakitandwe, L. Holmfeldt, M. Parker, J. Easton, R. Huether, R. Kriwacki, M. Rusch, G. Wu, Y. Li, H. Mulder, S. Raimondi, S. Pounds, G. Kang, L. Shi, J. Becksfort, P. Gupta, D. Payne-Turner, B. Vadodaria, K. Boggs, D. Yergeau, J. Manne, G. Song, M. Edmonson, P. Nagahawatte, L. Wei, C. Cheng, D. Pei, R. Sutton, N. C. Venn, A. Chetcuti, A. Rush, D. Catchpoole, J. Heldrup, T. Fioretos, C. Lu, L. Ding, C. H. Pui, S. Shurtleff, C. G. Mullighan, E. R. Mardis, R. K. Wilson, T. A. Gruber, J. Zhang, J. R. Downing and P. St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome (2015). "The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias." *Nat Genet* **47**(4): 330-337.
- Assi, S. A., M. R. Imperato, D. J. L. Coleman, A. Pickin, S. Potluri, A. Ptasinska, P. S. Chin, H. Blair, P. Cauchy, S. R. James, J. Zacarias-Cabeza, L. N. Gilding, A. Beggs, S. Clokie, J. C. Loke, P. Jenkin, A. Uddin, R. Delwel, S. J. Richards, M. Raghavan, M. J. Griffiths, O. Heidenreich, P. N. Cockerill and C. Bonifer (2019). "Subtype-specific regulatory network rewiring in acute myeloid leukemia." *Nat Genet* **51**(1): 151-162.
- Bagger, F. O., S. Kinalis and N. Rapin (2019). "BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles." *Nucleic Acids Res* **47**(D1): D881-D885.
- Barabe, F., L. Gil, M. Celton, A. Bergeron, V. Lamontagne, E. Roques, K. Lagace, A. Forest, R. Johnson, L. Pecheux, J. Simard, J. Pelloux, A. Bellemare-Pelletier, E. Gagnon, J. Hebert, S. Cellot and B. T. Wilhelm (2017). "Modeling human MLL-AF9 translocated acute myeloid leukemia from single donors reveals RET as a potential therapeutic target." *Leukemia* **31**(5): 1166-1176.
- Barabe, F., J. A. Kennedy, K. J. Hope and J. E. Dick (2007). "Modeling the initiation and progression of human acute leukemia in mice." *Science* **316**(5824): 600-604.
- Bernt, K. M., N. Zhu, A. U. Sinha, S. Vempati, J. Faber, A. V. Krivtsov, Z. Feng, N. Punt, A. Daigle, L. Bullinger, R. M. Pollock, V. M. Richon, A. L. Kung and S. A. Armstrong (2011).

"MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L." Cancer Cell **20**(1): 66-78.

Bolouri, H., J. E. Farrar, T. Triche, Jr., R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvil, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard and S. Meshinchi (2018). "The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions." Nat Med **24**(1): 103-112.

Borsari, B., P. Villegas-Mirón, H. Laayouni, A. Segarra-Casas, J. Bertranpetit, R. Guigó and S. Acosta (2020). "Intronic enhancers regulate the expression of genes involved in tissue-specific functions and homeostasis." bioRxiv: 2020.2008.2021.260836.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21 29 21-29.

Cancer Genome Atlas Research, N., T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing and G. Eley (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." N Engl J Med **368**(22): 2059-2074.

Celton, M., A. Forest, G. Gosse, S. Lemieux, J. Hebert, G. Sauvageau and B. T. Wilhelm (2014). "Epigenetic regulation of GATA2 and its impact on normal karyotype acute myeloid leukemia." Leukemia **28**(8): 1617-1626.

Chen, X., D. B. Burkhardt, A. A. Hartman, X. Hu, A. E. Eastman, C. Sun, X. Wang, M. Zhong, S. Krishnaswamy and S. Guo (2019). "MLL-AF9 initiates transformation from fast-proliferating myeloid progenitors." Nat Commun **10**(1): 5767.

Corces, M. R., J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti and H. Y. Chang (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." Nat Genet **48**(10): 1193-1203.

Cumbay, M. G. and V. J. Watts (2004). "Novel regulatory properties of human type 9 adenylate cyclase." J Pharmacol Exp Ther **310**(1): 108-115.

De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." Blood Cancer J **6**(7): e441.

Endoh, Y., Y. M. Chung, I. A. Clark, C. L. Geczy and K. Hsu (2009). "IL-10-dependent S100A8 gene induction in monocytes/macrophages by double-stranded RNA." *J Immunol* **182**(4): 2258-2268.

Erfurth, F. E., R. Popovic, J. Grembecka, T. Cierpicki, C. Theisler, Z. B. Xia, T. Stuart, M. O. Diaz, J. H. Bushweller and N. J. Zeleznik-Le (2008). "MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression." *Proc Natl Acad Sci U S A* **105**(21): 7517-7522.

Farlik, M., F. Halbritter, F. Muller, F. A. Choudry, P. Ebert, J. Klughammer, S. Farrow, A. Santoro, V. Ciaurro, A. Mathur, R. Uppal, H. G. Stunnenberg, W. H. Ouwehand, E. Laurenti, T. Lengauer, M. Frontini and C. Bock (2016). "DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation." *Cell Stem Cell* **19**(6): 808-822.

Garcia-Cuellar, M. P., C. Buttner, C. Bartenhagen, M. Dugas and R. K. Slany (2016). "Leukemogenic MLL-ENL Fusions Induce Alternative Chromatin States to Drive a Functionally Dichotomous Group of Target Genes." *Cell Rep* **15**(2): 310-322.

Gausdal, G., A. Wergeland, J. Skavland, E. Nguyen, F. Pendino, N. Rouhee, E. McCormack, L. Herfindal, R. Kleppe, U. Havemann, F. Schwede, O. Bruserud, B. T. Gjertsen, M. Lanotte, E. Segal-Bendirdjian and S. O. Doskeland (2013). "Cyclic AMP can promote APL progression and protect myeloid leukemia cells against anthracycline-induced apoptosis." *Cell Death Dis* **4**: e516.

Ge, S. X. and D. Jung (2018). "ShinyGO: a graphical enrichment tool for ani-mals and plants." *bioRxiv*: 315150.

Gilliland, C. T., C. L. Salanga, T. Kawamura, J. Trejo and T. M. Handel (2013). "The chemokine receptor CCR1 is constitutively active, which leads to G protein-independent, beta-arrestin-mediated internalization." *J Biol Chem* **288**(45): 32194-32210.

Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C. M. Croce and E. Canaani (1992). "The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene." *Cell* **71**(4): 701-708.

Guenther, M. G., L. N. Lawton, T. Rozovskaia, G. M. Frampton, S. S. Levine, T. L. Volkert, C. M. Croce, T. Nakamura, E. Canaani and R. A. Young (2008). "Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia." *Genes Dev* **22**(24): 3403-3408.

Hay, S. B., K. Ferchen, K. Chetal, H. L. Grimes and N. Salomonis (2018). "The Human Cell Atlas bone marrow single-cell interactive web portal." *Exp Hematol* **68**: 51-61.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." *Mol Cell* **38**(4): 576-589.

Hodges, E., A. Molaro, C. O. Dos Santos, P. Thekkat, Q. Song, P. J. Uren, J. Park, J. Butler, S. Rafii, W. R. McCombie, A. D. Smith and G. J. Hannon (2011). "Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment." *Mol Cell* **44**(1): 17-28.

- Hsieh, J. J., P. Ernst, H. Erdjument-Bromage, P. Tempst and S. J. Korsmeyer (2003). "Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization." *Mol Cell Biol* **23**(1): 186-194.
- Jung, N., B. Dai, A. J. Gentles, R. Majeti and A. P. Feinberg (2015). "An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis." *Nat Commun* **6**: 8489.
- Kato, T., Y. Fujita, K. Nakane, K. Mizutani, R. Terazawa, H. Ehara, Y. Kanimoto, T. Kojima, Y. Nozawa, T. Deguchi and M. Ito (2013). "CCR1/CCL5 interaction promotes invasion of taxane-resistant PC3 prostate cancer cells by increasing secretion of MMPs 2/9 and by activating ERK and Rac signaling." *Cytokine* **64**(1): 251-257.
- Khan, A. and A. Mathelier (2017). "Intervene: a tool for intersection and visualization of multiple gene or genomic region sets." *BMC Bioinformatics* **18**(1): 287.
- Krueger, F. and S. R. Andrews (2011). "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." *Bioinformatics* **27**(11): 1571-1572.
- Lagace, K., F. Barabe, J. Hebert, S. Cellot and B. T. Wilhelm (2017). "Identification of novel biomarkers for MLL-translocated acute myeloid leukemia." *Exp Hematol* **56**: 58-63.
- Laouedj, M., M. R. Tardif, L. Gil, M. A. Raquil, A. Lachhab, M. Pelletier, P. A. Tessier and F. Barabe (2017). "S100A9 induces differentiation of acute myeloid leukemia cells through TLR4." *Blood* **129**(14): 1980-1990.
- Lee, E. J., J. Luo, J. M. Wilson and H. Shi (2013). "Analyzing the cancer methylome through targeted bisulfite sequencing." *Cancer Lett* **340**(2): 171-178.
- Lerdrup, M., J. V. Johansen, S. Agrawal-Singh and K. Hansen (2016). "An interactive environment for agile analysis and visualization of ChIP-sequencing data." *Nat Struct Mol Biol* **23**(4): 349-357.
- Ley, T. J., L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch, C. C. Harris, C. F. Lichti, R. R. Townsend, R. S. Fulton, D. J. Dooling, D. C. Koboldt, H. Schmidt, Q. Zhang, J. R. Osborne, L. Lin, M. O'Laughlin, J. F. McMichael, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, V. J. Magrini, T. L. Vickery, J. Hundal, L. L. Cook, J. J. Conyers, G. W. Swift, J. P. Reed, P. A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M. A. Watson, S. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M. H. Tomasson, D. C. Link, T. A. Graubert, J. F. DiPersio, E. R. Mardis and R. K. Wilson (2010). "DNMT3A mutations in acute myeloid leukemia." *N Engl J Med* **363**(25): 2424-2433.
- Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biol* **15**(12): 550.
- Mankuzhy, N. P., E. Walling, B. Anderson and R. Mody (2019). "Cryptic ETV6-ABL1 Fusion and MLL2 Truncation Revealed by Integrative Clinical Sequencing in Multiply Relapsed Acute Lymphoblastic Leukemia." *J Pediatr Hematol Oncol* **41**(8): 653-656.
- Meyer, C., T. Burmeister, D. Groger, G. Tsaour, L. Fechina, A. Renneville, R. Sutton, N. C. Venn, M. Emerenciano, M. S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal,

V. H. J. van der Velden, E. Delabesse, T. S. Park, S. H. Oh, M. L. M. Silva, T. Lund-Aho, V. Juvonen, A. S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenskaya, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. Zur Stadt, J. Lentes, G. Gohring, A. Kustanovich, O. Aleinikova, B. W. Schafer, S. Kubetzko, H. O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J. M. Cayuela, E. Clappier, C. N. Alonso, C. M. Zwaan, M. M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Moricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O. A. Haas, R. Panzer-Grumayer, L. Sedek, T. Szczepanski, A. Caye, L. Suarez, H. Cave and R. Marschalek (2018). "The MLL recombinome of acute leukemias in 2017." Leukemia **32**(2): 273-284.

Milne, T. A. (2017). "Mouse models of MLL leukemia: recapitulating the human disease." Blood **129**(16): 2217-2223.

Naya, A., Y. Sagara, K. Ohwaki, T. Saeki, D. Ichikawa, Y. Iwasawa, K. Noguchi and N. Ohtake (2001). "Design, synthesis, and discovery of a novel CCR1 antagonist." J Med Chem **44**(9): 1429-1435.

Pabst, C., A. Bergeron, V. P. Lavallee, J. Yeh, P. Gendron, G. L. Norddahl, J. Krosil, I. Boivin, E. Deneault, J. Simard, S. Imren, G. Boucher, K. Eppert, T. Herold, S. K. Bohlander, K. Humphries, S. Lemieux, J. Hebert, G. Sauvageau and F. Barabe (2016). "GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo." Blood **127**(16): 2018-2027.

Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel and E. M. Rubin (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Perez, D. R., Y. Smagley, M. Garcia, M. B. Carter, A. Evangelisti, K. Matlawska-Wasowska, S. S. Winter, L. A. Sklar and A. Chigaev (2016). "Cyclic AMP efflux inhibitors as potential therapeutic agents for leukemia." Oncotarget **7**(23): 33960-33982.

Pomaznoy, M., B. Ha and B. Peters (2018). "GOnet: a tool for interactive Gene Ontology analysis." BMC Bioinformatics **19**(1): 470.

Roopra, A. (2019). "MAGICTRICKS: A tool for predicting transcription factors and cofactors that drive gene lists." bioRxiv: 492744.

Sam, T. N., J. H. Kersey, A. M. Linabery, K. J. Johnson, N. A. Heerema, J. M. Hilden, S. M. Davies, G. H. Reaman and J. A. Ross (2012). "MLL gene rearrangements in infant leukemia vary with age at diagnosis and selected demographic factors: a Children's Oncology Group (COG) study." Pediatr Blood Cancer **58**(6): 836-839.

Sandoval, J. E., Y. H. Huang, A. Muise, M. A. Goodell and N. O. Reich (2019). "Mutations in the DNMT3A DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners." J Biol Chem **294**(13): 4898-4910.

Saultz, J. N. and R. Garzon (2016). "Acute Myeloid Leukemia: A Concise Review." J Clin Med **5**(3).

Schaller, M. A., L. E. Kallal and N. W. Lukacs (2008). "A key role for CC chemokine receptor 1 in T-cell-mediated respiratory inflammation." *Am J Pathol* **172**(2): 386-394.

Schulze, I., C. Rohde, M. Scheller-Wendorff, N. Baumer, A. Krause, F. Herbst, P. Riemke, K. Hebestreit, P. Tschanter, Q. Lin, H. Linhart, L. A. Godley, H. Glimm, M. Dugas, W. Wagner, W. E. Berdel, F. Rosenbauer and C. Muller-Tidow (2016). "Increased DNA methylation of Dnmt3b targets impairs leukemogenesis." *Blood* **127**(12): 1575-1586.

Sharma, A., N. Jyotsana, R. Gabdoulline, D. Heckl, F. Kuchenbauer, R. K. Slany, A. Ganser and M. Heuser (2019). "MN1 is indispensable for MLL-rearranged acute myeloid leukemia." *Haematologica*.

Shen, L., N. Shao, X. Liu and E. Nestler (2014). "ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases." *BMC Genomics* **15**: 284.

Shin, S. Y., D. H. Lee, J. Lee, C. Choi, J. Y. Kim, J. S. Nam, Y. Lim and Y. H. Lee (2017). "C-C motif chemokine receptor 1 (CCR1) is a target of the EGF-AKT-mTOR-STAT3 signaling axis in breast cancer cells." *Oncotarget* **8**(55): 94591-94605.

Slany, R. K. (2009). "The molecular biology of mixed lineage leukemia." *Haematologica* **94**(7): 984-993.

Stunnenberg, H. G., C. International Human Epigenome and M. Hirst (2016). "The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery." *Cell* **167**(7): 1897.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proc Natl Acad Sci U S A* **102**(43): 15545-15550.

Thomas, R., S. Thomas, A. K. Holloway and K. S. Pollard (2017). "Features that define the best ChIP-seq peak calling algorithms." *Brief Bioinform* **18**(3): 441-450.

Tkachuk, D. C., S. Kohler and M. L. Cleary (1992). "Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias." *Cell* **71**(4): 691-700.

Trowbridge, J. J., A. U. Sinha, N. Zhu, M. Li, S. A. Armstrong and S. H. Orkin (2012). "Haploinsufficiency of Dnmt1 impairs leukemia stem cell function through derepression of bivalent chromatin domains." *Genes Dev* **26**(4): 344-349.

Tsukada, J., Y. Yoshida, Y. Kominato and P. E. Auron (2011). "The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation." *Cytokine* **54**(1): 6-19.

Wang, H. Q., L. K. Tuominen and C. J. Tsai (2011). "SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures." *Bioinformatics* **27**(2): 225-231.

Winters, A. C. and K. M. Bernt (2017). "MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches." *Front Pediatr* **5**: 4.

Yan, F., D. R. Powell, D. J. Curtis and N. C. Wong (2020). "From reads to insight: a hitchhiker's guide to ATAC-seq data analysis." *Genome Biol* **21**(1): 22.



Zheng, Y., H. Zhang, Y. Wang, X. Li, P. Lu, F. Dong, Y. Pang, S. Ma, H. Cheng, S. Hao, F. Tang, W. Yuan, X. Zhang and T. Cheng (2016). "Loss of Dnmt3b accelerates MLL-AF9 leukemia progression." Leukemia **30**(12): 2373-2384.

# CHAPITRE 3

## CARACTÉRISATION DES DIFFÉRENCES GÉNÉTIQUES ET ÉPIGÉNÉTIQUES ENTRE LES MODÈLES DE LEUCÉMIES HUMAINES DE TYPE *KMT2A-MLLT3* ET *KMT2A-MLLT4*

### **3 Caractérisation des différences génétiques et épigénétiques entre les modèles de leucémies humaines de type *KMT2A-MLLT3* et *KMT2A-MLLT4***

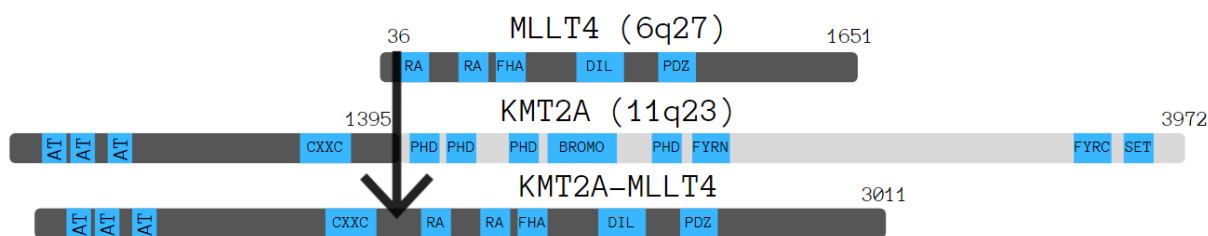
#### **3.1 Introduction**

Le premier chapitre a mis en évidence l'existence de multiples fusions oncogéniques responsables de la LMA chez les patients. Alors que la fréquence de chacune d'entre elles varie selon l'âge des malades et le type de leucémies, la probabilité de survie est également plus ou moins élevée selon la translocation chromosomique. L'objectif de ce chapitre est de comprendre les particularités génétiques et épigénétiques induites par une fusion de gènes peu étudiée : *KMT2A-MLLT4* (*KM4*).

Parmi les multiples fusions de gènes qui impliquent *KMT2A*, les mécanismes biologiques faisant intervenir la protéine partenaire *MLLT4* (*AF6* ou *AFDN*) restent peu compris. La translocation chromosomique *t(6;11)(q27;q23)* n'est pas considérée comme rare chez les patients pédiatriques atteints d'une LMA (Balgobind, Raimondi et al. 2009, Pigazzi, Masetti et al. 2011). Cependant, elle est associée à un diagnostic vital fortement défavorable comparativement aux autres fusions de type *KMT2A* (Martineau, Berger et al. 1998). En effet, il a été montré que la probabilité de survie varie entre 11% et 25% après 5 années (*5-year event free survival*) ce qui en fait l'une des LMA pédiatriques les plus agressives parmi celles causées par une fusion de gènes de type *KMT2A* (Balgobind, Raimondi et al. 2009). Ces données laissent à penser que les protéines de fusion utilisent des mécanismes différents de transformation pour activer des programmes génétiques particuliers qui mènent à des LMA plus agressives (Chen, Sun et al. 2016).

Bien que les fonctions des protéines de fusion ne soient pas complètement caractérisées, il est possible de mettre en évidence des points communs entre celles-ci, notamment la surexpression de certains gènes comme *Hoxa9* ou *Meis1*, qui sont connus pour induire des leucémies agressives chez la souris (Thorsteinsdottir, Kroon et al. 2001, Dasse, Volpe et al. 2012). Comme expliqué dans le chapitre 1, les partenaires de fusion peuvent être classifiés selon leur localisation dans la cellule (noyau ou cytoplasme). À l'état physiologique, la protéine

MLLT4 est localisée dans le cytoplasme, contrairement à la plupart des partenaires de fusion de KMT2A généralement retrouvés dans le noyau. Elle possède deux caractéristiques principales. La première est la présence d'un domaine de type PDZ ayant pour rôle de mener la protéine MLLT4 à la membrane pour lui permettre d'interagir avec des composants cellulaires, comme le récepteur tyrosine kinase EPH (Buchert, Schneider et al. 1999), la protéine kinase BCR (Radziwill, Erdmann et al. 2003), la protéine PRR au niveau des jonctions adhérentes (Takahashi, Nakanishi et al. 1999), les protéines Ras, Notch et Dishevelled (Carmena, Speicher et al. 2006) ou encore le membre c-Src de la famille des protéines kinases Src (Radziwill, Weiss et al. 2007). La seconde caractéristique est la présence de deux domaines homologues aux domaines RA (*RAS-Association*) des effecteurs RAS (Kuriyama, Harada et al. 1996). Cela permet à la protéine MLLT4, seule ou fusionnée à KMT2A, de moduler la transduction de voies de signalisation *in vivo*, notamment celles impliquant RAS, NOTCH ou WNT (Kuriyama, Harada et al. 1996). La protéine MLLT4, dans le cytoplasme, agit comme régulateur du niveau de RAS-GTP dans les cellules (Manara, Baron et al. 2014).



**Figure 3.1 – Structure de la protéine KMT2A-MLLT4.**

Adaptée de (Smith, Ottoni et al. 2017). La protéine se compose des 1395 premiers résidus de KMT2A ainsi que des résidus 36 à 1651 de MLLT4. FHA = Forkhead associated domain. DIL = Dilute domain. Les autres domaines ont déjà été décrits dans les chapitres précédents.

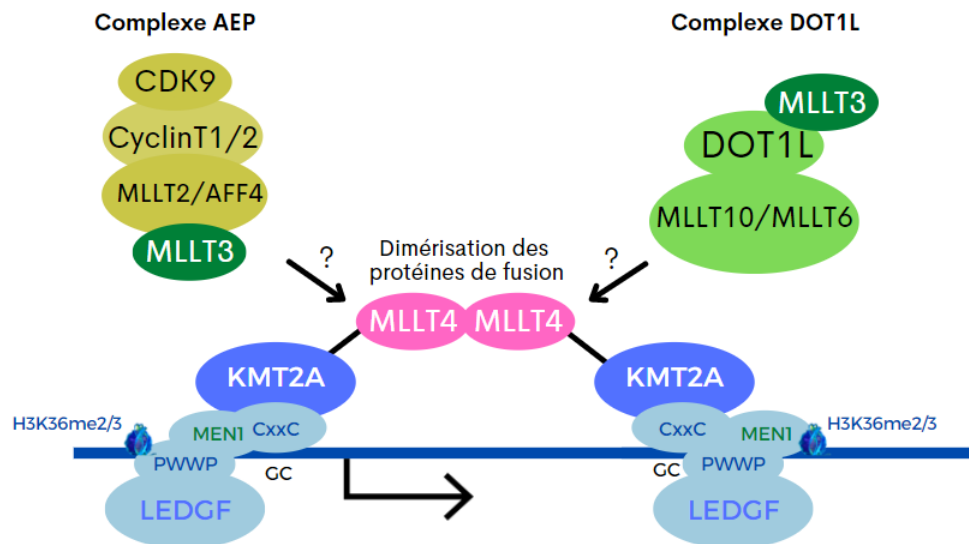
La translocation chromosomique *KM4* est à l'origine de la création d'une protéine nucléaire d'environ 334 kDa composée de la partie N-terminale de KMT2A et de la quasi-totalité de la protéine MLLT4 (**Figure 3.1**). Cette dernière conserve ses domaines fonctionnels, tels que PDZ et RA, et ne démontre aucune homologie (ni dans les séquences, ni dans les fonctions) avec le produit d'autres partenaires de gènes de *KMT2A* (Prasad, Gu et al. 1993). Il a d'ailleurs été

démontré que les motifs RA de la protéine MLLT4 sont nécessaires et suffisants pour la transformation leucémique (Liedtke, Ayton et al. 2010). La propriété oncogénique de *KM4* est décrite comme étant dépendante de la dimérisation de la protéine *KM4*, causée par l'exposition de résidus hydrophobes suite à la perte de l'hélice  $\alpha$  à l'extrémité N-terminale de la protéine MLLT4 (Smith, Ottoni et al. 2017). Ce processus est un mécanisme clé dans la leucémogénèse médiée par *KM4* : l'inhibition de la dimérisation perturbe la formation du complexe et bloque le développement de la leucémie chez la souris (Smith, Ottoni et al. 2017). Par ailleurs, cette oligomérisation pourrait être médiée par MLLT4, qui agirait comme une protéine d'échafaudage pour permettre l'interaction avec les gènes cibles (Liedtke, Ayton et al. 2010). Il a été montré que MLLT4 est recruté dans le noyau des cellules porteuses de la fusion *KM4*, ce qui entraîne une activation anormale de RAS et de l'ensemble de ses cibles en aval. L'inactivation de *KM4* restaure la localisation cytoplasmique de MLLT4 et permet une réduction significative des niveaux de RAS-GTP. Cela démontre que *KM4* agit sur la voie de signalisation médiée par RAS en agissant sur la rétention de MLLT4 dans le noyau (Manara, Baron et al. 2014). Par ailleurs, il a été récemment démontré que *KM4* a besoin de l'activité d'histone méthyltransférase de DOT1L pour maintenir son programme d'expression génique : c'est la raison pour laquelle DOT1L est considérée comme ayant un rôle oncogénique important (Deshpande, Chen et al. 2013).

D'ailleurs, certains groupes ont tenté d'identifier des molécules thérapeutiques efficaces spécifiquement chez les patients porteurs de la translocation chromosomique *KM4*. En effet, la plupart d'entre eux succombe dans l'année qui suit le diagnostic en raison de la résistance aux thérapies cytotoxiques conventionnelles. Certaines études amènent des perspectives thérapeutiques prometteuses. Alors que l'inhibition pharmacologique ou génétique de DOT1L semble être efficace pour endiguer la leucémogénèse dans un modèle murin (Deshpande, Chen et al. 2013), l'utilisation d'inhibiteur de RAS sur des lignées cellulaires leucémiques et des cellules dérivées de patients démontre un potentiel thérapeutique en clinique (Manara, Baron et al. 2014). Une étude récente ayant pour objectif de réaliser un criblage de molécules thérapeutiques pour trouver de nouveaux composés et accélérer l'accès à de nouveaux traitements en clinique a permis d'identifier la molécule thioridazine (TDZ) (Tregnago, Da Ros et al. 2020). Ce composé entraîne un remodelage du cytosquelette et agit sur le niveau de

calcium intracellulaire pour déclencher une réponse apoptotique dans les cellules LMA porteuses de la fusion de type *KM4*.

Les mécanismes biochimiques qui sous-tendent les propriétés oncogéniques de la fusion restent encore à élucider. Malgré tout, il est connu que MLLT4 n'a pas été identifié comme faisant partie du complexe d'élongation transcriptionnelle (Liedtke, Ayton et al. 2010, Yokoyama, Lin et al. 2010). Néanmoins, *KM4* recrute les complexes DOT1L et AEP sur la chromatine cible via un mécanisme encore méconnu ce qui permet l'activation de l'élongation transcriptionnelle de gènes cibles (**Figure 3.2**) (Yokoyama, Lin et al. 2010, Deshpande, Chen et al. 2013). Une étude récente a permis d'identifier les gènes différemment exprimés dans des lignées cellulaires porteuses (*ML-2* ou *SHI-1*) ou non de la fusion *KM4* (Numata, Kwok et al. 2018). *SHARPI* s'avère jouer un rôle essentiel dans la leucémogénèse médiée par *KM4* : la surexpression de ce gène, spécifiquement chez les patients porteurs de cette fusion, permettrait l'activation de gènes cibles impliqués dans la survie cellulaire (Coenen, Zwaan et al. 2014).



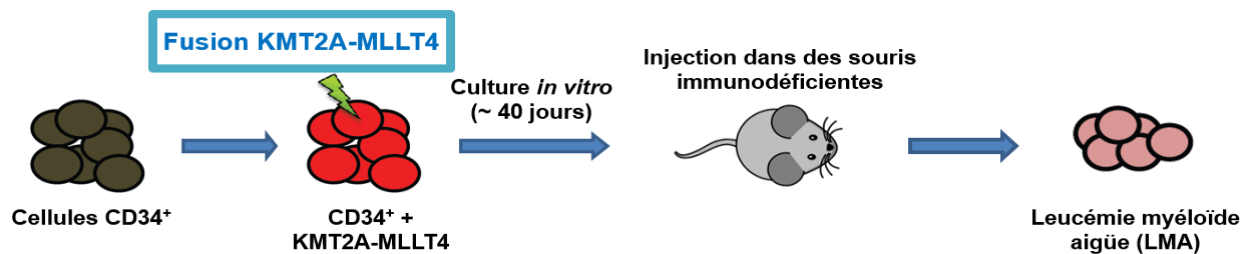
**Figure 3.2 – Mécanismes moléculaires menant à l'activité oncogénique des protéines de fusion KMT2A-MLLT4.**

La protéine recrute les complexes AEP et DOT1L par un mécanisme encore inconnu. Figure tirée et adaptée de la revue (Takahashi and Yokoyama 2020).

Cette revue de littérature résume les connaissances actuelles de la leucémogénèse causée par la translocation chromosomique *KM4*. La majorité des études ont été réalisées sur des lignées cellulaires. Décrire les conséquences génétiques directes de la fusion reste à déterminer.

### 3.2 Objectifs du projet

L'objectif de ce chapitre est de caractériser les différences génétiques et épigénétiques entre les modèles leucémiques médiés par les fusions *KM3* et *KM4* (pour simplifier l'écriture, nous parlerons des modèles ou de leucémies *KM3* et *KM4* dans le reste du chapitre). En exploitant les avantages du modèle décrit dans le chapitre précédent, il est possible d'adapter le modèle et d'induire l'expression de la fusion oncogénique *KM4* pour développer une leucémie humaine dans les souris immunodéficientes (**Figure 3.3**). La présence de la fusion est suffisante pour déclencher la leucémogénèse.



**Figure 3.3 – Représentation du modèle leucémique humain de type *KMT2A-MLLT4*.**

Des cellules sanguines  $CD34^+$  provenant d'un sang de cordon ombilical sont infectées par un rétrovirus portant la fusion oncogénique *KM4*. Ces cellules sont ensuite injectées dans une souris immunodéficiente pour produire une leucémie myéloïde aiguë humaine après quelques semaines.

Les approches expérimentales sont donc similaires à celles utilisées dans le chapitre précédent : séquençage de l'ARN (RNA-seq), immunoprécipitation de la chromatine (ChIP-seq) contre des marques d'histones de type H3K4me3 & H3K79me2 et analyse de la conformation de la chromatine (ATAC-seq).

Trois sous-objectifs ont été établis pour mener l'étude :

**1. Confirmer l'utilisation et l'adaptation du modèle leucémique**

C'est en comparant les données transcriptomiques d'échantillons du modèle avec celles de patients leucémiques que le modèle pourra être validé.

**2. Déterminer, à l'échelle du génome, les différences génétiques et épigénétiques entre les deux modèles**

L'idée est de cartographier et de comparer les profils transcriptomiques ainsi que les paysages épigénétiques des leucémies causées par les fusions *KM4* et *KM3*. L'analyse se base sur les données de RNA-seq auxquelles s'ajoutent les données épigénétiques qui pourraient expliquer comment les gènes sont régulés.

**3. Analyser et définir l'existence de potentiels biomarqueurs spécifiques à la LMA causée par la fusion oncogénique *KM4***

Ces résultats préliminaires permettront de paver la voie à de nouvelles analyses fonctionnelles pour déterminer le rôle de certains gènes dans la leucémogénèse.



### **3.3 Matériels et méthodes**

Les approches expérimentales sont très similaires à celles utilisées dans le chapitre précédent.

#### **3.3.1 Échantillons de patients leucémiques**

Tous les échantillons utilisés provenant de patients pédiatriques atteints d'une LMA et les informations cliniques associées ont été collectés par la Banque de Cellules Leucémiques du Québec (BCLQ) avec le consentement éclairé et l'approbation du projet des comités d'éthique de la recherche du CHU Ste-Justine, de l'hôpital Maisonneuve-Rosemont et de l'Université de Montréal. L'ARN a été extrait d'échantillons conservés au TRIzol (Invitrogen) selon des protocoles standards.

#### **3.3.2 Génération des modèles de leucémies *KMT2A-MLLT4***

Les modèles de leucémies ont été générés comme décrit dans de précédentes publications (Barabe, Gil et al. 2017) et adaptés à la fusion *KMT2A-MLLT4*. Le plasmide provient du laboratoire du Dr Michael Cleary (Lavau, Szilvassy et al. 1997, DiMartino, Miller et al. 2000, Liedtke, Ayton et al. 2010) et code pour les résidus 1-1395 du gène humain *KMT2A* fusionnés aux résidus 35-348 du gène humain *MLLT4*.

#### **3.3.3 RNA-seq, ChIP-seq et ATAC-seq – Protocoles et analyses bioinformatiques**

Tous les protocoles expérimentaux sont détaillés dans la partie 2.4.5 « methods » du chapitre précédent. Aucune modification n'a été apportée. Le protocole d'ATAC-seq a été réalisé en utilisant 50 000 cellules et est basé sur la méthode déjà publiée dans la littérature (Buenrostro, Wu et al. 2015). Le séquençage (*paired-end* 75 pb) a été réalisé par la plateforme de génomique de l'Institut de Recherche en Immunologie et Cancérologie.

Pour les données de ChIP-seq et d'ATAC-seq, l'analyse bioinformatique est détaillée dans la partie 2.4.5 « methods » du chapitre précédent.

Pour le RNA-seq, Trimmomatic (version 0.35) a été utilisé pour retirer les index et les *reads* de mauvaise qualité. STAR (version 2.5.1b) a permis d'aligner les *reads* au génome humain de

référence (GRCH38/hg38). La liste des gènes différentiellement exprimés entre les modèles *KM3* et *KM4* a été produite en utilisant le package sur R, DESeq2 (Love, Huber et al. 2014). L'analyse des sites de fixation des facteurs de transcription a été réalisée via l'outil ShinyGO (version 0.61) (Ge, Jung et al. 2020).

## 3.4 Résultats

### 3.4.1 Les signatures transcriptomiques du modèle *KM4* sont similaires à celles retrouvées chez les patients leucémiques porteurs de la fusion *KM4*

Les données sur le modèle *KM3* déjà publiées ont permis de démontrer que le modèle récapitule ce qui est observé chez des patients porteurs de la translocation chromosomique (Barabe, Gil et al. 2017). Comme le modèle leucémique *KM4* n'a jamais été étudié auparavant, la première étape a été de valider l'utilisation du modèle en exploitant les données de RNA-seq.

#### 3.4.1.1 Les données du modèle proches des observations cliniques

Nous avons comparé les données transcriptomiques de 15 échantillons du modèle *KM4* avec celles provenant d'autres modèles et de patients porteurs de multiples types de translocations chromosomiques (médiées ou non par *KMT2A*). La classification hiérarchique basée sur l'expression des 2000 gènes les plus variables permet d'établir plusieurs observations.

La première observation est le regroupement de la majorité des modèles *KM4* avec deux des trois patients porteurs de la même fusion (**Figure 3.4**). Cela démontre la proximité des données du modèle avec les données cliniques en notre possession. Seulement deux échantillons de patients pédiatriques sont situés sur la même branche de la classification, le troisième étant plus éloigné et porteur d'une signature transcriptomique plus distinct. Cela souligne l'enjeu de l'hétérogénéité génétique qu'il est possible d'observer chez les patients. Il faudrait intégrer davantage de patients *KM4* dans cette classification pour avoir une meilleure idée de la comparaison entre patients et modèles.

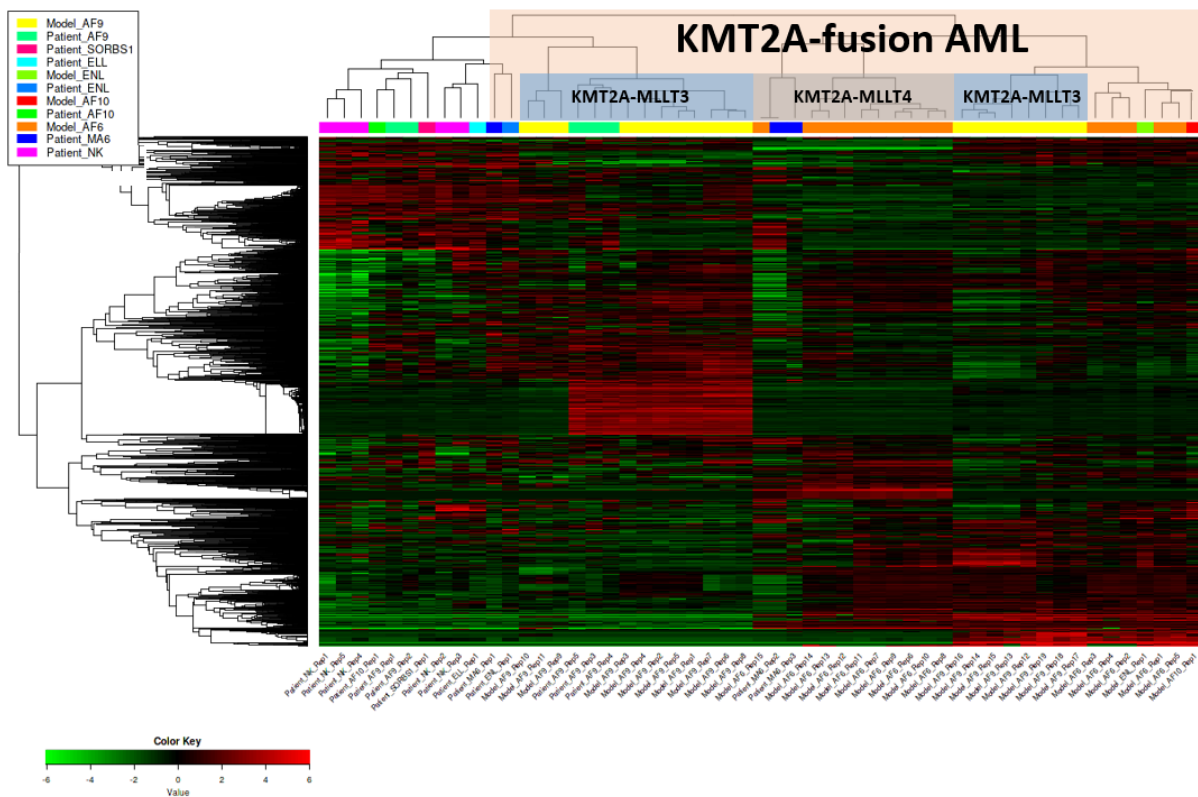
La deuxième observation réside dans le fait que les modèles *KM4* se situent sur une branche de la classification distincte – mais proche – de deux autres branches représentant les leucémies (modèles et patients) de type *KM3*. Cette répartition laisse à penser que les deux fusions causent chacune des changements transcriptomiques minimes mais suffisants pour les séparer dans la classification.

La troisième observation concerne la division en deux groupes des réplicats des leucémies *KM3* d'une part et des réplicats des leucémies *KM4* d'autre part. Dans les deux cas, une raison

expliquant cette séparation pourrait être la diversité biologique parmi les réplicats causée par l'intégration aléatoire de la fusion de gènes à la base du modèle.

Enfin, la quatrième observation est la distinction des LMA médiées par une fusion de type *KMT2A* des autres types de LMA (non-*KMT2A*). Cette représentation souligne les différences transcriptomiques entre les multiples leucémies causées par différentes translocations chromosomiques et démontre le rapprochement des données d'expression entre les leucémies impliquant le gène *KMT2A*.

Ensemble, les données démontrent la capacité des modèles *KM4* à récapituler ce qui est observé en clinique chez les patients leucémiques.



**Figure 3.4 – Classification hiérarchique d'échantillons de modèles LMA et de patients LMA basée sur leur signature transcriptomique respective.**

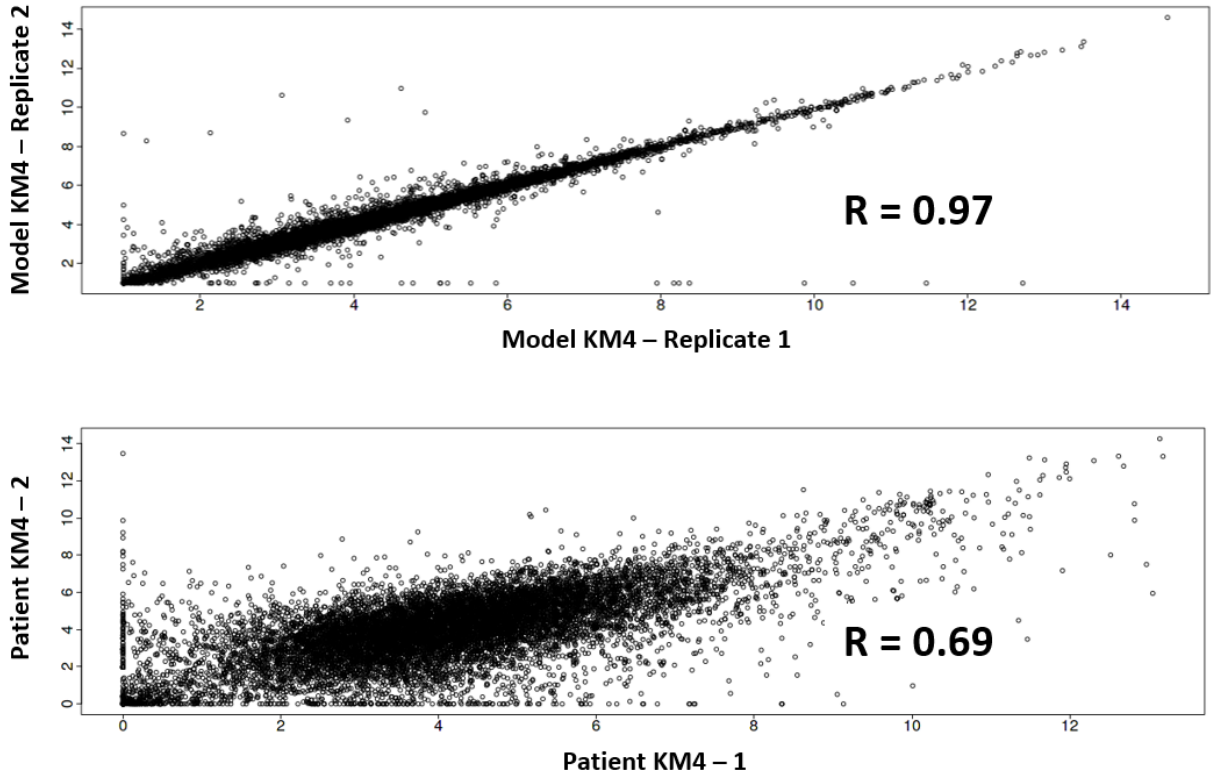
Le *clustering* tient compte de 2000 gènes les plus différenciellement exprimés dans les modèles et patients leucémiques. Chaque colonne représente un échantillon leucémique. Chaque ligne correspond à un gène. L'échelle de couleur permet de représenter les différences relatives d'expression des gènes. Le rectangle de couleur orange pâle représente la vaste majorité des échantillons LMA (modèle ou patient) possédant une fusion de type *KMT2A*. Les leucémies médiées par les fusions *KMT2A-MLLT3* et *KMT2A-MLLT4* ont été mises en évidence.

Model\_AF9 = modèle *KMT2A-MLLT3* ; patient\_AF9 = patient *KMT2A-MLLT3* ; patient\_ELL = patient *KMT2A-ELL* ; model\_ENL = modèle *KMT2A-MLLT1* ; patient\_ENL = patient *KMT2A-MLLT1* ; model\_AF10 = modèle *KMT2A-MLLT10* ; patient\_AF10 = patient *KMT2A-MLLT10* ; model\_AF6 = modèle *KMT2A-MLLT4* ; patient\_MA6 = patient *KMT2A-MLLT4* ; patient\_NK = patient Normal Karyotype.

### 3.4.1.2 Un modèle utile pour la production de réplicats biologiques

L'un des enjeux principaux dans l'étude de la leucémogenèse est l'accès aux échantillons de patients. Leur rareté limite les expérimentations. C'est donc l'une des forces de l'utilisation du modèle : produire rapidement des réplicats biologiques. S'assurer de leur reproductibilité est cependant une caractéristique à ne pas négliger.

Pour vérifier cet aspect, nous avons comparé des échantillons du modèle et de patients deux à deux (**Figure 3.5**). Comme observé dans le modèle *KM3* (Barabe, Gil et al. 2017), la corrélation entre deux échantillons du modèle *KM4* est très forte (coefficient de corrélation proche de 1) alors que deux échantillons provenant de deux différents patients *KM4* présentent un profil transcriptomique davantage éloigné l'un de l'autre (coefficient de corrélation = 0.69). Cette représentation apporte la preuve de l'utilité et de la force du modèle utilisé : il est plus facile de reproduire plusieurs réplicats biologiques pour faciliter l'analyse et trouver de potentiels biomarqueurs en commun. Le fond génétique des cellules souches CD34<sup>+</sup> qui sont à la base du modèle n'a donc pas de conséquence sur la formation de la leucémie humaine dans la souris immunodéficiente. Chez les patients, les translocations chromosomiques sont accompagnées de mutations récurrentes secondaires. À l'inverse, dans les modèles, seule la fusion de gènes *KM4* est présente et suffisante pour produire une LMA. L'ensemble des données transcriptomiques valident l'utilisation du modèle *KM4*.



**Figure 3.5 – Nuage de points (*scatterplot*) des valeurs d’expression génique entre deux réplicats biologiques du modèle *KM4* (en haut) et deux échantillons pédiatriques leucémiques *KM4* (en bas).**

Les coefficients de corrélation démontrent la très forte variabilité transcriptomique entre les patients comparativement aux modèles. 11,509 gènes sont représentés dans ce graphique. L’échelle est exprimée en  $\log(\text{FPKM} + 1)$ .

### 3.4.2 La comparaison des profils d'expression génique entre les deux modèles leucémiques révèle une baisse d'expression de gènes intervenant dans les mécanismes de traduction

Pour caractériser les différences au niveau du transcriptome, les profils d'expression génique des 15 réplicats biologiques des modèles *KM3* et *KM4* ont été comparés. L'algorithme utilisé (DESeq2) est une méthode quantitative d'analyse différentielle des données d'expression entre échantillons (Love, Huber et al. 2014).

#### 3.4.2.1 Des signatures distinctes

Afin de déterminer les processus biologiques reliés aux différences d'expression génique, une analyse des termes de *Gene Ontology* (GO-term) permet de faire ressortir des signatures. L'outil DESeq2 permet de distinguer deux groupes de gènes avec des profils d'expression similaires entre les 15 réplicats de chacun des modèles.

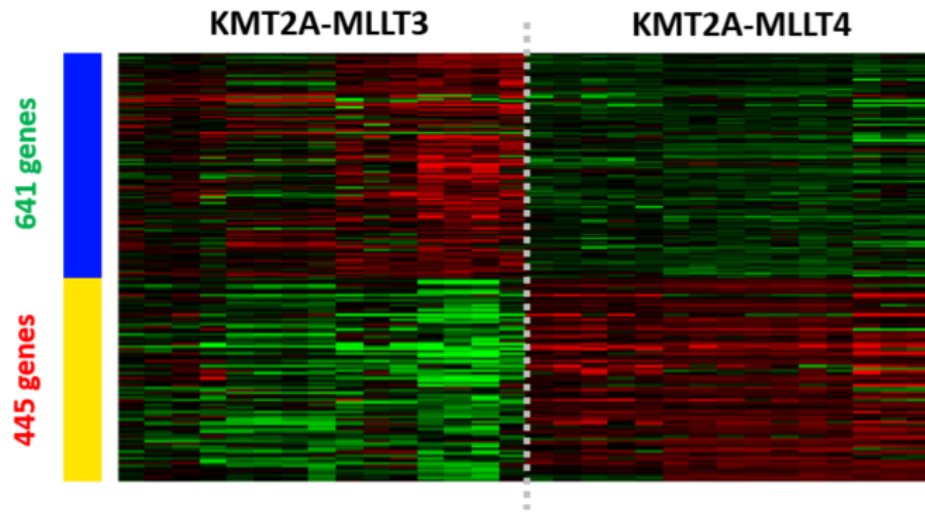
D'un côté, 445 gènes se retrouvent plus fortement exprimés dans le modèle *KM4* par rapport au modèle *KM3* (**Figure 3.6a**). L'analyse par *GO-term* révèle que les gènes concernés codent pour des composants cellulaires comme les éléments du cytosquelette, le centre d'organisation des microtubules, le centrosome, le centriole, l'assemblage des fuseaux mitotiques, les projections cellulaires, la condensation des chromosomes ou encore les vésicules sécrétoires. Par ailleurs, les processus biologiques impactés concernent la régulation du cycle cellulaire, la division cellulaire, la régulation de l'organisation des composants cellulaires, l'organisation des organelles ou encore le cycle du centrosome. En résumé, les gènes avec un niveau d'expression plus élevé dans le modèle *KM4* par rapport au modèle *KM3* sont principalement impliqués dans le cycle et le fonctionnement cellulaire (**Figure 3.6b**).

De l'autre côté, le niveau d'expression de 641 gènes est plus faible dans les 15 réplicats du modèle *KM4* comparativement aux 15 réplicats du modèle *KM3* (**Figure 3.6a**). Les ribosomes, les petites et grandes sous-unités ribosomales, les jonctions adhérentes ou encore les adhésions focales semblent être les composants cellulaires concernés par cette diminution d'expression génique. À cela s'ajoutent les processus cataboliques des ARN, l'initiation de la traduction ou encore l'établissement de la localisation des protéines à la membrane, comme exemples de processus biologiques impactés par la présence de la fusion *KM4* (**Figure 3.6c**). Cette analyse



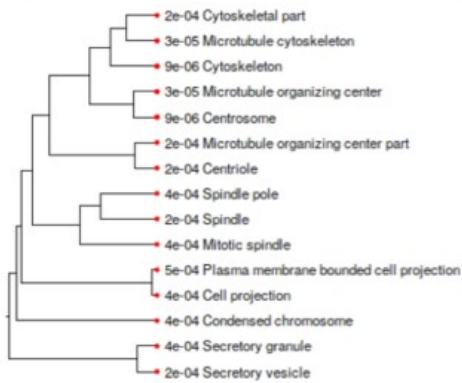
laisse à penser que les mécanismes de traduction des protéines sont impactés en présence de la fusion *KM4*. L'outil *ReactomePA* (Yu and He 2016) confirme que la traduction, la régulation transcriptionnelle par p53, les principales voies de maturation des ARN ribosomiaux dans le noyau ou dans le cytosol, l'épissage des ARN messagers sont les processus biologiques majoritairement affectés par la présence de la fusion oncogénique *KM4* en comparaison avec le modèle *KM3* (**Figure 3.6d**).

**A**

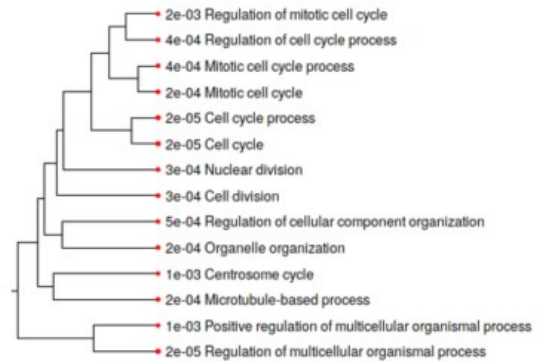


**B**

**GO Cellular Component**

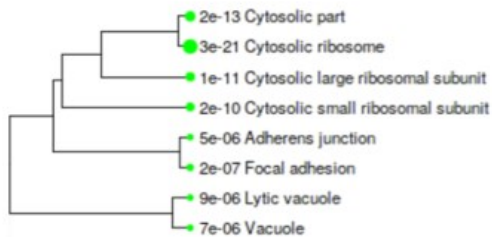


**GO Biological Process**

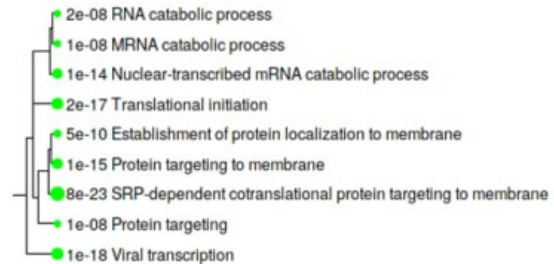


**C**

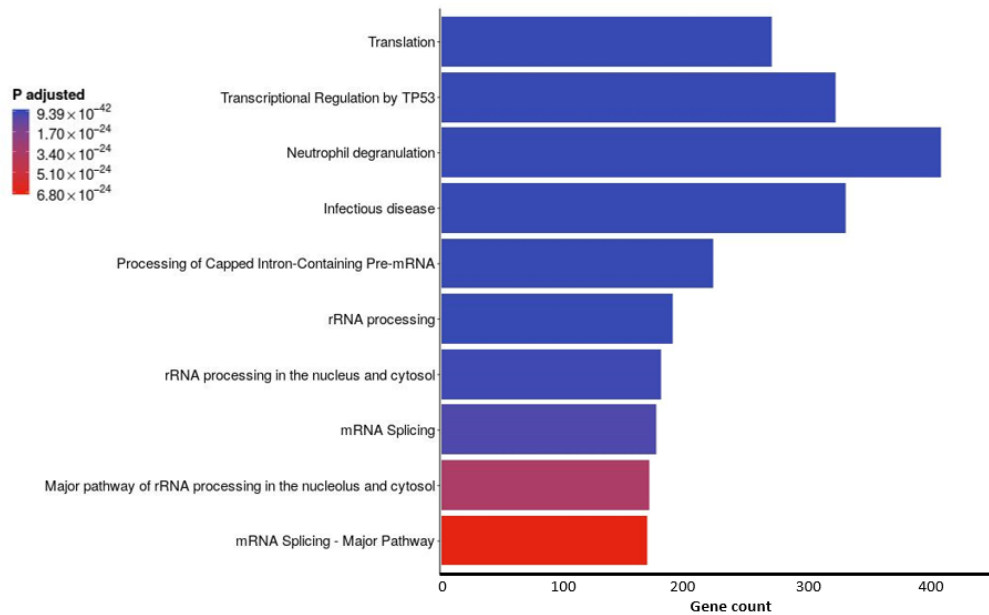
**GO Cellular Component**



**GO Biological Process**



D



**Figure 3.6 – Analyse transcriptomique des modèles *KM4* et *KM3* et caractérisation des processus biologiques impactés.**

(A) Représentation par *heatmap* de deux groupes de gènes différentiellement exprimés parmi les 15 réplicats du modèle *KM4* ou *KM3*. Une analyse de GO-term dans le groupe de gènes plus (B) ou moins (C) exprimés dans le modèle *KM4* relativement au modèle *KM3* révèle que les mécanismes traductionnels sont un élément clé qui différencie les deux leucémies. L’observation est confirmée par l’utilisation de l’algorithme ReactomePA (D) qui permet de représenter le nombre de gènes retrouvés dans certaines catégories de processus biologiques et d’associer le résultat à des valeurs statistiques.

GO Cellular Component = Gene Ontology des composants cellulaires ; GO Biological Process = Gene Ontology des processus biologiques

### 3.4.2.2 Une baisse d'expression de certains gènes codant pour des facteurs de transcription pourrait être à l'origine de ces signatures

Afin de mieux caractériser ce qui détermine ces changements transcriptionnels, nous avons utilisé des outils informatiques de type ShinyGO (Ge and Jung 2018) pour mesurer la fréquence des sites de fixation des facteurs de transcription situés en amont des gènes différentiellement exprimés. L'analyse des motifs enrichis permet de produire une liste de potentiels facteurs de transcription ayant un effet dans la régulation transcriptionnelle.

D'une part, les séquences en amont des 641 gènes plus faiblement exprimés dans le modèle *KM4* correspondent principalement à des sites de fixation pour des facteurs de transcription de type doigt de zinc (**Figure 3.7a**). La majorité des gènes codant pour ces derniers ne sont pas différentiellement exprimés (selon les tests statistiques), sauf quelques exceptions. En effet, le niveau d'expression de certains gènes codant pour des facteurs de transcription est plus faible dans le modèle *KM4* (**Figure 3.7b**) :

- PATZ1 est un facteur de transcription de type doigt de zinc et agit comme un modulateur positif ou négatif de l'expression de gènes dépendamment du contexte cellulaire (Valentino, Palmieri et al. 2013). La protéine PATZ1 agirait sur l'acétylation des histones en interagissant avec des protéines telles que NCOR ou SIRT1 et pourrait avoir un rôle dans le remodelage de la chromatine notamment via ses domaines « AT-hook » (Sakaguchi, Hombauer et al. 2010, Cho, Kim et al. 2012). Elle est impliquée dans de nombreux processus biologiques comme l'apoptose, la sénescence, la prolifération ou encore la différenciation des lymphocytes T (Fedele, Crescenzi et al. 2017) mais également certains cancers comme les lymphomes (Franco, Scognamiglio et al. 2016).
- E2F1 et E2F4 sont deux facteurs de transcription respectivement activateurs et répresseurs, tous deux impliqués dans la transition G1/S dans le cycle cellulaire et le développement myéloïde (Tripathi, Sharma et al. 2011). La famille E2F est connue pour jouer un rôle dans la cancérogenèse (Kent and Leone 2019).
- RREB1 est impliqué dans la signalisation de RAS. La surexpression du gène *RREB1* chez des patients LMA pourrait être à l'origine du blocage dans le processus de différenciation des granulocytes (Yao, Zhong et al. 2019).

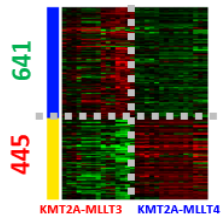
- MZF1 est un facteur de transcription impliqué dans la régulation transcriptionnelle du développement hématopoïétique. L'expression aberrante de *MZF1* joue un rôle dans la formation de cancers hématopoïétiques (Brix, Bundgaard Clemmensen et al. 2020).

- NFKB1/p50, une des 5 sous-unités de NF-KB, est un facteur de transcription impliqué dans la réaction immunitaire et qui peut jouer un rôle d'oncogène ou de suppresseur de tumeur dépendamment du type de cancer (Concetti and Wilson 2018).

D'autre part, les séquences des 445 gènes plus fortement exprimés dans le modèle *KM4* correspondent principalement à des sites de fixation pour des facteurs de transcription de type homéodomaine (**Figure 3.7c**). Seuls les gènes codant pour ARID3B, HOXC9 et HMG20B semblent être significativement moins exprimés (**Figure 3.7d**) mais cette tendance est peu marquée.

Pris ensemble, ces résultats permettent d'émettre l'hypothèse que les mécanismes de développement leucémique causé par les fusions *KM3* et *KM4* semblent se distinguer au niveau des mécanismes traductionnels.

A



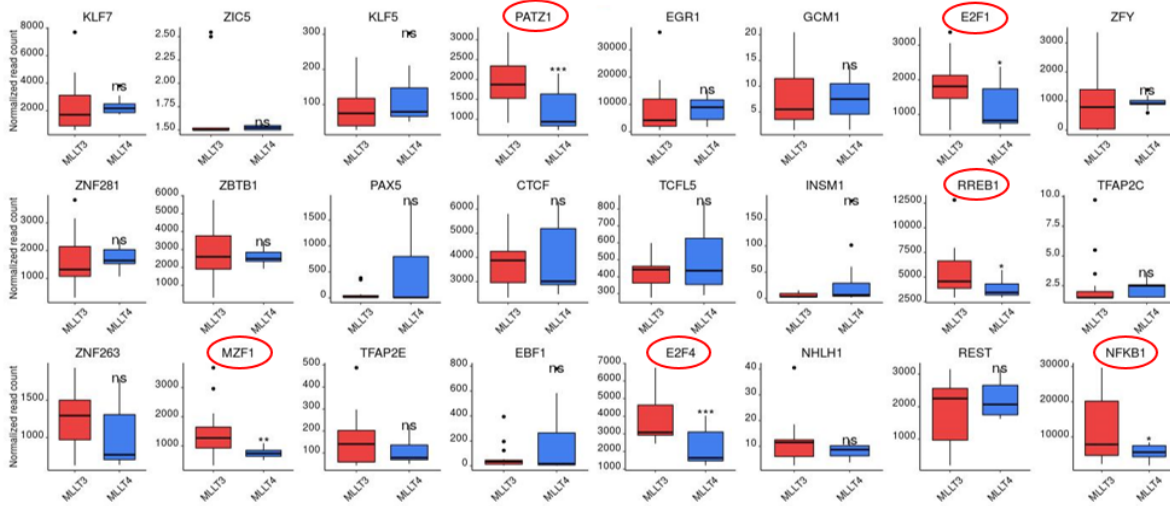
Upstream 300bp as promoter

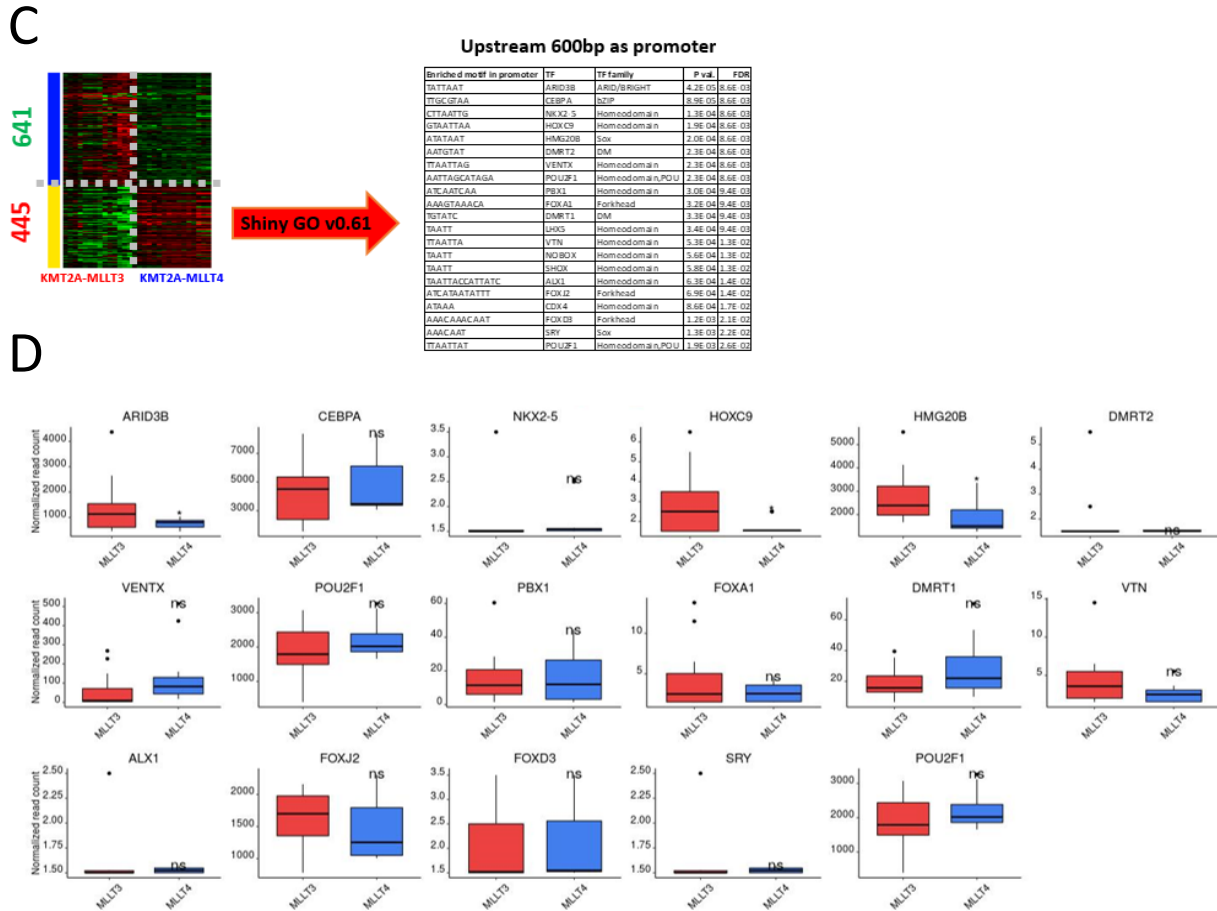
Enriched motif in promoter	TF	TF family	P val	FDR
AGGCGTGT	KLF7	CH2 ZF	4.4E-16	2.4E-14
GGGGGGT	TCF5	CH2 ZF	8.9E-16	4.8E-14
GGGGGGGGC	KLF5	CH2 ZF	1.2E-15	5.9E-14
GGGGGGGGCC	PATZ1	CH2 ZF	8.9E-15	3.3E-13
AGGGGGGGGG	RREB1	CH2 ZF	4.9E-14	1.3E-12
TCCGGG	GCM1	GCM	4.6E-12	1.4E-10
GGGGGGAG	E2F1	E2F	6.3E-12	1.7E-10
CCCGAGGCT	ZFY	CH2 ZF	6.6E-12	1.7E-10
AGGAGTAA	ZNF281	CH2 ZF	7.1E-12	1.7E-10
TCCAGG	ZBTB1	CH2 ZF	1.7E-11	3.6E-10
GGAGGCGAGAGTGGTGGGG	PAX5	Paired box	2.3E-11	4.9E-10
TGGCAGCAGGGGGGGC	CTCF	CTCF	5.6E-11	1.3E-09
CACGTG	TCFL5	BHLH	6.7E-11	1.3E-09
TGTCAGGGGGC	INSM1	CH2 ZF	7.4E-11	1.3E-09
GGGGGGGGGGTTGGGG	RREB1	CH2 ZF	1.4E-10	2.4E-09
CCCTGGGGC	TFAP2C	AP 2	1.5E-10	2.3E-09
GGAGGAGGAGGGGAGGAGG	ZNF263	CH2 ZF	3.0E-10	4.8E-09
GGGGGG	MEF1	CH2 ZF	9.4E-10	1.4E-08
CCCTGGGG	TFAP2E	AP 2	3.2E-09	4.8E-08
TCCCTGGGGG	SBF1	BHLH	3.9E-09	5.3E-08
GGCGGGAA	E2F4	E2F	1.4E-08	1.9E-07

Upstream 600bp as promoter

Enriched motif in promoter	TF	TF family	P val	FDR
GGGGGGGGCC	PATZ1	CH2 ZF	1.1E-16	6.6E-15
CCCGAGGCT	ZFY	CH2 ZF	4.4E-16	2.3E-14
GGGGGGT	ZIC5	CH2 ZF	5.6E-16	2.6E-14
GGGGGGGGGG	E2F1	CH2 ZF	3.6E-15	1.3E-13
GGGGGGGGC	KLF5	CH2 ZF	4.7E-15	1.6E-13
GGGAGTGG	ZNF281	CH2 ZF	5.8E-15	1.6E-13
GGGGA	MEF1	CH2 ZF	1.8E-14	5.0E-13
GGGGGGGGGGTGGGG	RREB1	CH2 ZF	1.9E-14	5.0E-13
CGAGGGCGAGTGAATGGAGGCGC	PAX5	Paired box	3.0E-14	7.7E-13
GGGGTGG	KLF7	CH2 ZF	6.7E-14	1.6E-12
TCCGGG	GCM1	GCM	4.0E-13	9.1E-12
GGGGGGCAGTGGCTCC	NHLH1	BHLH	4.3E-13	9.3E-12
GGGGGGAG	E2F1	E2F	5.4E-13	1.1E-11
TCCGGG	ZBTB1	CH2 ZF	1.5E-12	2.9E-11
TCC TGGGGA	E2F1	BHLH	2.0E-12	3.7E-11
CCCTGGGGC	TFAP2C	AP 2	6.4E-12	1.1E-10
GGGGGGCAGTGGGGC	NHLH1	BHLH	1.4E-11	2.3E-10
TGTCAGGGGGC	INSM1	CH2 ZF	2.3E-11	3.6E-10
GGGGCTCCGGTGGTGGAA	REST	CH2 ZF	2.4E-11	3.7E-10
GGGAGTCC	NFKB1	Rel	3.0E-11	4.6E-10
CCCTGGGG	TFAP2E	AP 2	3.3E-11	4.6E-10
CACGTG	TCFL5	BHLH	3.3E-11	4.6E-10

B





**Figure 3.7 – Résumé des motifs de sites de fixation des facteurs de transcription retrouvés au niveau des promoteurs (-300 ou -600 pb en amont du TSS) des gènes plus (A) ou moins (C) exprimés dans le modèle *KM4* comparativement au modèle *KM3*.**

L'analyse a été faite avec l'algorithme ShinyGO v0.61. Les niveaux d'expression des gènes codant pour les différents facteurs de transcription listés en (A) et (C) sont représentés respectivement dans les panels (B) et (D) : les couleurs rouges et bleues correspondent respectivement aux 15 réplicats biologiques du modèle *KM3* (MLLT3) ou *KM4* (MLLT4) et l'échelle représente le nombre normalisé de *reads* pour chacun des gènes.

### **3.4.3 L'intégration des données transcriptomiques et épigénétiques permet de révéler la présence de potentiels biomarqueurs spécifiques au modèle *KM4***

Les analyses précédentes ont permis de cartographier les différences globales entre les modèles *KM3* et *KM4*. Dans l'optique d'identifier des gènes dérégulés spécifiquement dans le modèle *KM4* pouvant servir de biomarqueurs voire de cibles thérapeutiques en clinique, nous avons intégré les données de RNA-seq, de ChIP-seq et d'ATAC-seq pour affiner l'analyse.

#### **3.4.3.1 *PROM1*, potentiel biomarqueur de la LMA médiée par la fusion *KM4***

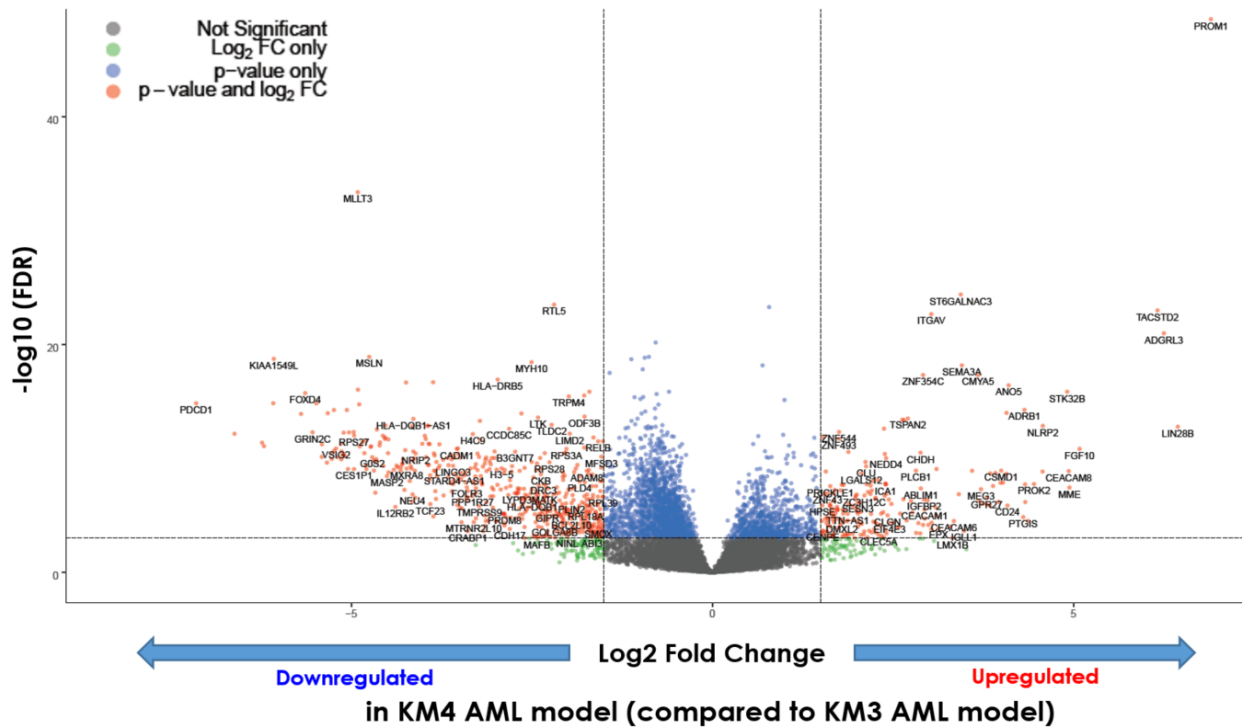
La représentation par *volcano plot* (**Figure 3.8**) issue de l'analyse de DESeq2 met en évidence les différences d'expression des gènes entre les deux modèles et d'y associer des valeurs statistiques. Peu de gènes passent les seuils statistiques et la vaste majorité d'entre eux ont une probabilité qui suggère une forte hétérogénéité entre les réplicats biologiques. Par ailleurs, les ratios d'expression entre les deux modèles sont faibles ce qui laisse à penser que la présence de la fusion *KM4* induit une modification minime dans les niveaux d'expression mais suffisante pour induire une leucémie plus agressive.

À noter que deux gènes se différencient des autres :

- *MLLT3* est plus fortement exprimé dans le modèle *KM3* par rapport au modèle *KM4*. Cette observation pourrait s'expliquer par un biais dans le séquençage et un nombre plus élevé de *reads* en raison de la présence de la fusion *KMT2A-MLLT3* dans le modèle en plus de la protéine sauvage.
- *PROM1* est significativement surexprimé dans le modèle *KM4*. Il a été montré qu'en clinique, une forte expression du gène codant pour *PROM1* corrèle avec une faible probabilité de survie chez les patients atteints de LMA (Tolba, Foda et al. 2013). Son potentiel rôle dans la leucémogénèse est discuté dans la partie suivante.

Par ailleurs, la différence d'expression des autres gènes présentés en **Figure 3.8** semblent être spécifiques du modèle car cette même tendance ne se retrouve pas dans les données cliniques en notre possession.





**Figure 3.8 –** Volcano plot représentant les différences d’expression des gènes entre les deux modèles leucémiques.

Chaque point représente un gène. À droite du graphique, l’expression des gènes dans le modèle *KM4* est plus élevée que dans le modèle *KM3*. À gauche du graphique, les gènes dans le modèle *KM4* sont moins exprimés comparativement au modèle *KM3*. La couleur attribuée à chacun des points fait référence aux différents seuils statistiques : le vert correspond aux valeurs d’expression pour lesquelles le  $\log_2 \text{fold change}$  est supérieur ou égal à 1 ; le bleu correspond à la statistique *p*-value supérieure à 0,05 ; le rouge correspond aux gènes remplissant ces deux critères.

### 3.4.3.2 *LPL* est un exemple de gènes régulés de manière épigénétique

Pour décrire les différences épigénétiques entre les deux modèles et permettre l'identification de biomarqueurs, nous avons réalisé des expériences en triplicata d'ATAC-seq et de ChIP-seq. Alors que les données transcriptomiques laissent à penser que les différences d'expression de gènes sont minimales et peu nombreuses, les mêmes observations peuvent être faites au niveau épigénétique.

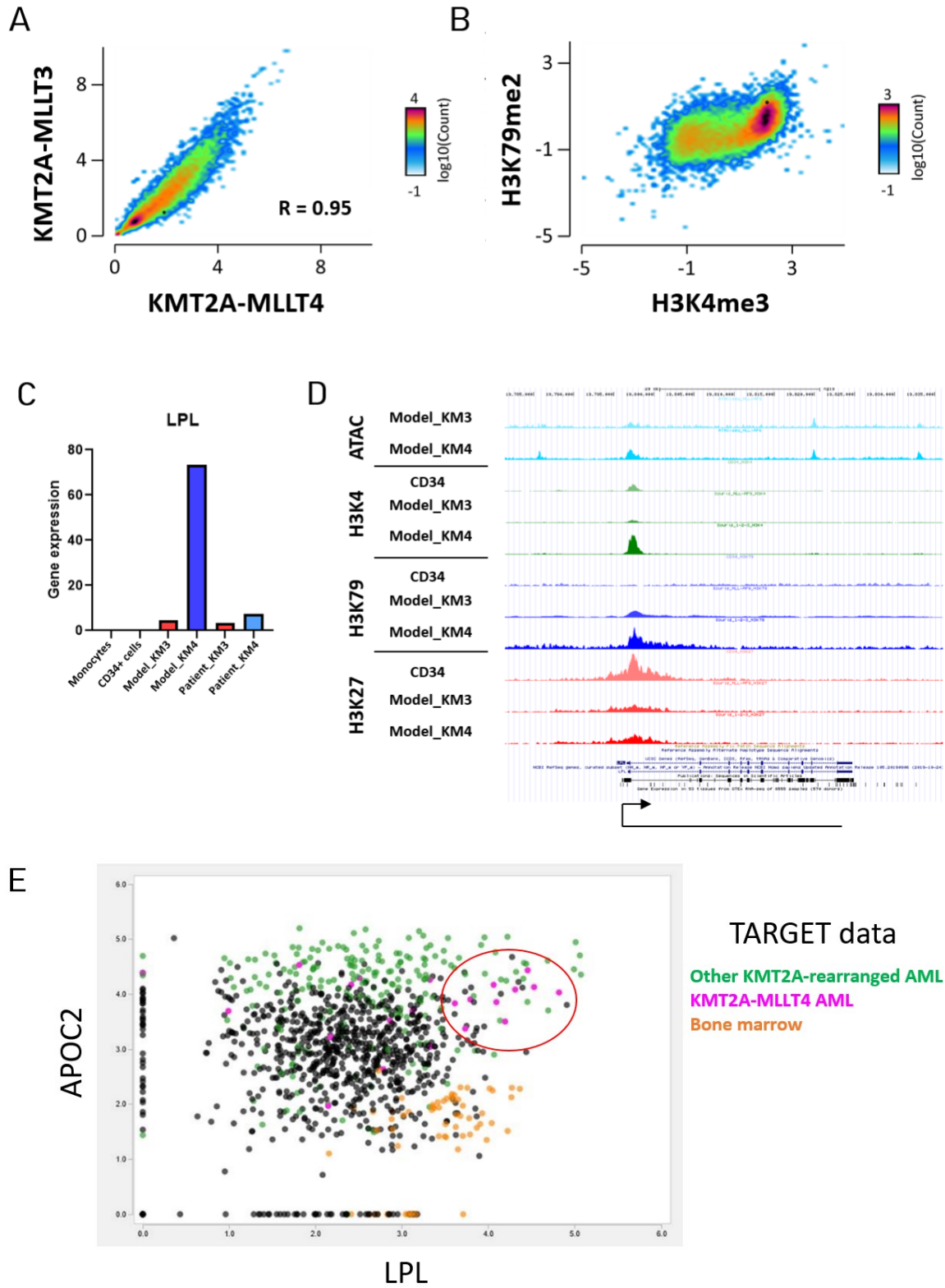
D'une part, l'approche par ATAC-seq met en évidence une très forte corrélation entre les deux modèles (coefficient de corrélation = 0.95) en termes d'accessibilité de la chromatine (**Figure 3.9a**). Cela corrobore l'idée qu'il y a peu de différences dans le mécanisme de régulation de l'expression des gènes entre les modèles, mais celles-ci seraient minimales et suffisantes pour induire une leucémie plus agressive. D'autre part, comparer la répartition de la marque d'histones H3K79me2 entre les deux modèles (**Figure 3.9b**) permet de mettre en évidence que la très grande majorité des gènes ne présentent pas ou peu de différences dans l'enrichissement de cette marque, alors que la répartition de la marque H3K4me3 entre les deux modèles semble légèrement affectée.

Ces approches permettent de sélectionner des groupes de gènes régulés de manière épigénétique et de les associer avec les données transcriptomiques. L'objectif est d'identifier des gènes différentiellement exprimés entre les deux modèles et régulés par des changements au niveau de la chromatine. Même si peu de gènes ressortent de cette analyse, cette approche épigénétique a permis d'identifier certains gènes spécifiques de la leucémie médiée par la fusion *KM4*. C'est le cas du gène codant pour la lipoprotéine lipase *LPL* dont le rôle est de catalyser l'hydrolyse des triglycérides. Son co-facteur *APOC2* joue un rôle critique dans le métabolisme lipidique et pourrait servir de potentielle cible thérapeutique dans le traitement contre la LMA (Zhang, Yang et al. 2020).

*LPL* est un gène fortement exprimé dans les échantillons de patients et du modèle *KM4* comparativement au modèle *KM3* (**Figure 3.9c**). L'approche par ATAC-seq permet de mettre en évidence une ouverture de la chromatine au niveau du promoteur du gène, plus marquée dans le modèle *KM4*. À cela s'ajoutent un enrichissement des marques d'histones activatrices H3K4me3 et H3K79me2 ainsi qu'une perte de la marque d'histones répressive H3K27me3

**(Figure 3.9d)**. En clinique, il est possible d'observer une co-expression des gènes *LPL* et *APOC2* spécifiquement chez les patients porteurs de la fusion *KM4* (**Figure 3.9e**).

Les données démontrent la difficulté à trouver des biomarqueurs spécifiques à la leucémie médiée par la fusion *KM4*. *PROM1* et *LPL* pourraient être deux bons candidats mais il reste encore à élucider leur spécificité et leur rôle dans ce type de leucémies.



**Figure 3.9 – *LPL* est un potentiel biomarqueur probablement régulé de manière épigénétique et spécifique de la LMA médiée par la fusion *KMT2A-MLLT4*.**

(A) Nuage de points (*scatterplot*) de l'enrichissement du signal d'ATAC-seq dans une région à +/- 2000 pb autour du TSS (n=3). Chaque point correspond à un gène et les couleurs correspondent à la densité des points. Les axes des ordonnées et des abscisses correspondent respectivement au modèle *KM3* et *KM4*.

(B) Nuage de points (*scatterplot*) de l'enrichissement des marques d'histones dans une région à +/- 2000 pb (H3K4me3) ou -1000/+3000 pb (H3K79me2) autour du TSS (n=3 pour les expériences de ChIP-seq sur chaque marque d'histones). Les valeurs des axes correspondent au ratio en  $\log_2$  de la valeur d'enrichissement des marques d'histones entre les modèles *KM4* et *KM3*. Chaque point correspond à un gène et les couleurs correspondent à la densité des points. Les graphiques (A) et (B) ont été produits en utilisant le programme EaSeq.

(C) Niveau moyen d'expression de *LPL* dans des monocytes, des cellules CD34<sup>+</sup> et des cellules de patients ou des modèles *KM3* ou *KM4*.

(D) Visualisation sur UCSC de données d'ATAC-seq et de ChIP-seq contre les marques d'histones H3K4me3, H3K79me2 et H3K27me3, dans les cellules CD34<sup>+</sup>, les modèles *KM3* et *KM4*, au niveau du gène *LPL*.

(E) Expression des gènes *LPL* et *APOC2* dans la cohorte de patients pédiatriques LMA (données TARGET). Chaque point correspond à un patient. Les couleurs vertes, roses et oranges correspondent respectivement à des données de patients avec une translocation de type *KMT2A*, avec une translocation *KMT2A-MLLT4* ou à des données de cellules non-leucémiques de moelle osseuse.

Model\_KM3 = Modèle *KMT2A-MLLT3*; Model\_KM4 = Modèle *KMT2A-MLLT4*.

### 3.5 Discussion

La translocation chromosomique *KMT2A-MLLT4* est à l'origine d'une LMA pédiatrique particulièrement agressive et pour laquelle les mécanismes génétiques ne sont pas encore compris. Notre étude se base sur un modèle précédemment utilisé pour étudier la fusion de gènes *KM3* en l'adaptant à la fusion *KM4*. La force du modèle est de cartographier les impacts directs de la présence de la translocation chromosomique sans l'ajout d'autres mutations récurrentes, contrairement aux études actuelles faites sur des lignées cellulaires leucémiques ou des échantillons de patients. Décrire et comparer les aspects génétiques et épigénétiques dans les LMA causées par les deux translocations chromosomiques permet d'apporter de nouveaux éléments de réponse pour comprendre pourquoi le pronostic vital chez les patients diffère autant.

L'une des premières observations concerne la perturbation des mécanismes traductionnelles. Les analyses par *GO-term* révèlent une baisse d'expression des gènes codant pour les ribosomes et les ARN ribosomiaux, ce qui corrobore avec l'idée que la traduction pourrait jouer un rôle dans la leucémogénèse (Bastide and David 2018, Pauli, Liu et al. 2020). Nous pouvons alors formuler l'hypothèse que le niveau global de protéines devrait être plus faible dans les cellules leucémiques porteuses de la fusion *KM4* par rapport à la fusion *KM3*. Se pose alors la question de déterminer l'avantage pour une cellule leucémique d'avoir une diminution de la concentration protéique, et surtout quels types de protéines sont concernés par cette baisse. Avec les données préliminaires à notre disposition, il est difficile d'émettre des hypothèses, mais cette variation protéique procurerait un avantage certain pour les cellules leucémiques, en termes de survie et de résistance aux traitements cytotoxiques, voire même de détection par le système immunitaire. Certaines protéines de surface pourraient être moins présentes ce qui empêcherait l'entrée des molécules thérapeutiques et leur action dans la cellule. La diminution de la présence d'autres protéines cytoplasmiques ou nucléaires pourrait venir empêcher le fonctionnement optimal voire même le transport des molécules thérapeutiques conventionnelles au sein de la cellule ou directement dans le noyau. Au-delà de la production protéique en elle-même, il est également probable que les modifications traductionnelles voire post-traductionnelles soient affectées dans la leucémogénèse médiée par la fusion *KM4*. Une analyse par spectrométrie de masse pourrait permettre d'étudier, de caractériser et de déterminer dans quelle mesure ce niveau de régulation est perturbé dans la leucémie causée par la fusion *KM4* comparativement à la

fusion *KM3*, à condition de connaître les cibles protéiques à analyser. Étudier les mécanismes traductionnels sous-entend également la difficulté à trouver une cible thérapeutique spécifique. La traduction faisant intervenir des centaines de protéines et d'ARN ribosomiaux, il est peu probable qu'un unique gène soit responsable d'une perturbation dans la traduction, ce qui complique la recherche de molécules thérapeutiques. Par ailleurs, l'exemple d'inhibiteurs de la traduction déjà étudiés pour traiter des maladies hématopoïétiques (Derenzini, Rossi et al. 2018) illustre l'enjeu des effets secondaires retrouvés en clinique (Laham-Karam, Pinto et al. 2020).

Une autre observation issue de l'analyse est l'augmentation du niveau d'expression des gènes codant pour le fonctionnement et la division cellulaire, pouvant potentiellement faciliter la prolifération des cellules leucémiques porteuses de la fusion *KM4*. La translocation chromosomique *KM4* pourrait également entraîner une recomposition des éléments du cytosquelette pour faciliter son expansion (via les projections cellulaires par exemple) et ses interactions avec le microenvironnement. Cela faciliterait la propagation des cellules leucémiques et la dissémination de métastases. Côté traitement, remodeler le cytosquelette pourrait également impacter l'efficacité et le mécanisme d'action des molécules thérapeutiques conventionnelles. Les gènes codant pour les vésicules sécrétoires semblent être également impactées par la présence de la fusion : elles pourraient servir à éliminer les molécules thérapeutiques en dehors de la cellule pour augmenter sa résistance aux traitements cytotoxiques conventionnels. Cela pourrait expliquer pourquoi les traitements classiques sont moins efficaces pour la LMA médiée par cette fusion *KM4*.

Ces tendances générales permettent d'avoir une vue globale des changements et des différences transcriptomiques induits directement par les deux translocations chromosomiques. De ces analyses, nous avons pu extraire deux gènes possédant un potentiel thérapeutique en clinique : *PROM1* et *LPL*.

Le gène *PROM1* semble être l'un des seuls gènes significativement plus exprimés dans le modèle *KM4*. Bien que son rôle biologique ne soit pas clairement défini, *PROM1* code pour la protéine membranaire CD133 et est exprimé dans les cellules hématopoïétiques progénitrices CD34<sup>+</sup>. C'est également un marqueur de cellules souches cancéreuses retrouvé dans une large variété de tumeurs solides et associé à un pronostic défavorable en clinique (Saha, Islam et al. 2020). L'expression élevée de *PROM1* est associée à une plus faible probabilité de survie chez

les patients atteints de LMA (Tolba, Foda et al. 2013). Il a été récemment montré que le gène *PROM1* est essentiel pour assurer la viabilité des cellules leucémiques lymphoïdes porteuses de la fusion *KMT2A-MLLT2* et sa régulation est assurée par la marque d'histones H3K79me2/3 au niveau de sa région amplificatrice *enhancer* (Godfrey, Crump et al. 2021). Dans nos données, aucune différence d'enrichissement de H3K79me2 entre les deux modèles n'a été observée, ce qui laisse à penser que le gène serait régulé par un autre mécanisme épigénétique. Par ailleurs, certains traitements en immunothérapie dirigés contre CD133 basés sur l'utilisation des *CAR-T cells* (*Chimeric Antigen Receptor T cells*) sont en développement dans des leucémies lymphoïdes aiguës causées par des réarrangements de type *KMT2A* (Li, Hu et al. 2018) mais se heurtent au fait qu'il soit également exprimé dans les CSH (Bueno, Velasco-Hernandez et al. 2019). En d'autres termes, même si *PROM1* semble être une cible intéressante dans les LMA, il ne semble pas être spécifique à la leucémie médiée par la fusion *KM4*.

Cependant, au niveau du modèle leucémique utilisé, *PROM1* pourrait être d'intérêt. En effet, *in vivo*, il est possible de produire des leucémies secondaires dans le modèle *KM4* en injectant des cellules leucémiques d'une souris dans une autre souris. L'objectif de ce type d'approche est de produire des leucémies qui proviennent d'un même clone et d'éviter ainsi l'hétérogénéité génétique des leucémies primaires causée par l'intégration aléatoire de la fusion dans le génome. Or, il est difficile voire impossible de produire des leucémies secondaires dans le modèle *KM3*. Cela nous amène à émettre l'hypothèse que *PROM1*, étant l'une des seules différences transcriptomiques marquées entre les deux modèles, pourrait jouer un rôle non seulement dans la production d'une leucémie secondaire, mais également en clinique, dans le maintien et le développement d'une LMA sur le long terme. Cette hypothèse pourrait être testée dans le modèle : forcer l'expression de *PROM1* dans un modèle *KM3* pour mesurer l'impact sur la production d'une leucémie secondaire. À l'inverse, il pourrait être intéressant d'inhiber l'expression du gène par une approche de CRISPR-Cas9 dans des cellules infectées par le rétrovirus porteur de la fusion *KM4* pour mesurer l'impact sur la leucémogénèse dans une souris. *In fine*, cette approche permettrait d'ajuster la façon de produire des leucémies dans un modèle murin, mais également d'utiliser potentiellement ce gène comme facteur de pronostic chez les patients atteints de LMA.



L'identification du gène *LPL*, régulé de manière épigénétique et essentiel pour assurer le stockage et l'utilisation des lipides dans l'organisme, laisse à penser que la perturbation du métabolisme lipidique jouerait un rôle dans le développement de la maladie. Dans la littérature, il a été montré que le gène *LPL* est impliqué dans les leucémies lymphoïdes chroniques pour lesquelles il représente un facteur de pronostic (Kaderi, Kanduri et al. 2011) et dont la régulation fonctionnerait via la méthylation de l'ADN (Abreu, Moreno et al. 2013). Côté mécanistique, la protéine LPL assure l'expansion et la survie des cellules leucémiques en les alimentant en acides gras et en se fixant aux cellules environnantes (Rombout, Verhasselt et al. 2016). L'augmentation de son expression pourrait donner un avantage aux cellules leucémiques en facilitant leur propagation et leur survie dans l'environnement médullaire. Quelques études mettent en avant l'importance du cofacteur APOC2 dans la LMA, notamment dans les lignées cellulaires porteuses de la fusion *KM3* (Zhang, Yang et al. 2020), mais aucune ne souligne son rôle dans les leucémies spécifiquement causées par la translocation chromosomique *KM4*. Il est probable que le rôle de *LPL* ne soit pas spécifique à ce type de leucémie mais valider sa fonction permettrait de proposer d'autres alternatives en termes de traitements thérapeutiques. Certains groupes ont développé des inhibiteurs ciblant la protéine LPL, notamment l'orlistat qui entraîne l'apoptose de cellules de leucémies lymphoïdes chroniques de type B (Pallasch, Schwamb et al. 2008). Il serait intéressant de tester l'efficacité de la molécule sur la viabilité des cellules leucémiques du modèle.

En conclusion, ce chapitre présente les données préliminaires permettant de mieux comprendre les mécanismes qui sous-tendent la leucémie médiée par la translocation chromosomique *KMT2A-MLLT4*. En exploitant les avantages du modèle, il a été possible de caractériser les conséquences directes de la présence de la fusion tout en évitant la présence de mutations récurrentes ou d'autres aberrations chromosomiques retrouvées dans le génome des patients ou dans des lignées cellulaires. L'identification des gènes *PROM1* et *LPL* dans ce type de leucémie ouvre la voie à de nouvelles potentielles avenues thérapeutiques. Par ailleurs, bien que ce soit encore des données préliminaires, nos résultats ont permis de soulever de nouvelles hypothèses et projets de recherche pour se concentrer davantage sur les mécanismes protéiques, aspect moins étudié dans la leucémogénèse.

### 3.6 Références

- Abreu, C., P. Moreno, F. Palacios, M. Borge, P. Morande, A. I. Landoni, R. Gabus, G. Dighiero, M. Giordano, R. Gamberale and P. Oppezzo (2013). "Methylation status regulates lipoprotein lipase expression in chronic lymphocytic leukemia." *Leuk Lymphoma* **54**(8): 1844-1848.
- Balgobind, B. V., S. C. Raimondi, J. Harbott, M. Zimmermann, T. A. Alonzo, A. Auvrignon, H. B. Beverloo, M. Chang, U. Creutzig, M. N. Dworzak, E. Forestier, B. Gibson, H. Hasle, C. J. Harrison, N. A. Heerema, G. J. Kaspers, A. Leszl, N. Litvinko, L. L. Nigro, A. Morimoto, C. Perot, R. Pieters, D. Reinhardt, J. E. Rubnitz, F. O. Smith, J. Stary, I. Stasevich, S. Strehl, T. Taga, D. Tomizawa, D. Webb, Z. Zemanova, C. M. Zwaan and M. M. van den Heuvel-Eibrink (2009). "Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study." *Blood* **114**(12): 2489-2496.
- Barabe, F., L. Gil, M. Celton, A. Bergeron, V. Lamontagne, E. Roques, K. Lagace, A. Forest, R. Johnson, L. Pecheux, J. Simard, J. Pelloux, A. Bellemare-Pelletier, E. Gagnon, J. Hebert, S. Cellot and B. T. Wilhelm (2017). "Modeling human MLL-AF9 translocated acute myeloid leukemia from single donors reveals RET as a potential therapeutic target." *Leukemia* **31**(5): 1166-1176.
- Bastide, A. and A. David (2018). "The ribosome, (slow) beating heart of cancer (stem) cell." *Oncogenesis* **7**(4): 34.
- Brix, D. M., K. K. Bundgaard Clemmensen and T. Kallunki (2020). "Zinc Finger Transcription Factor MZF1-A Specific Regulator of Cancer Invasion." *Cells* **9**(1).
- Buchert, M., S. Schneider, V. Meskenaite, M. T. Adams, E. Canaani, T. Baechi, K. Moelling and C. M. Hovens (1999). "The junction-associated protein AF-6 interacts and clusters with specific Eph receptor tyrosine kinases at specialized sites of cell-cell contact in the brain." *J Cell Biol* **144**(2): 361-371.
- Bueno, C., T. Velasco-Hernandez, F. Gutierrez-Aguera, S. R. Zanetti, M. L. Baroni, D. Sanchez-Martinez, O. Molina, A. Closa, A. Agraz-Doblas, P. Marin, E. Eyra, I. Varela and P. Menendez (2019). "CD133-directed CAR T-cells for MLL leukemia: on-target, off-tumor myeloablative toxicity." *Leukemia* **33**(8): 2090-2125.
- Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." *Curr Protoc Mol Biol* **109**: 21 29 21-21 29 29.
- Carmena, A., S. Speicher and M. Bayliss (2006). "The PDZ protein Canoe/AF-6 links Ras-MAPK, Notch and Wntless/Wnt signaling pathways by directly interacting with Ras, Notch and Dishevelled." *PLoS One* **1**: e66.
- Chen, L., Y. Sun, J. Wang, H. Jiang and A. G. Muntean (2016). "Differential regulation of the c-Myc/Lin28 axis discriminates subclasses of rearranged MLL leukemia." *Oncotarget* **7**(18): 25208-25223.
- Cho, J. H., M. J. Kim, K. J. Kim and J. R. Kim (2012). "POZ/BTB and AT-hook-containing zinc finger protein 1 (PATZ1) inhibits endothelial cell senescence through a p53 dependent pathway." *Cell Death Differ* **19**(4): 703-712.

- Coenen, E. A., C. M. Zwaan, J. Stary, A. Baruchel, V. de Haas, R. W. Stam, D. Reinhardt, G. J. Kaspers, S. T. Arentsen-Peters, C. Meyer, R. Marschalek, L. L. Nigro, M. Dworzak, R. Pieters and M. M. van den Heuvel-Eibrink (2014). "Unique BHLHB3 overexpression in pediatric acute myeloid leukemia with t(6;11)(q27;q23)." *Leukemia* **28**(7): 1564-1568.
- Concetti, J. and C. L. Wilson (2018). "NFkB1 and Cancer: Friend or Foe?" *Cells* **7**(9).
- Dasse, E., G. Volpe, D. S. Walton, N. Wilson, W. Del Pozzo, L. P. O'Neill, R. K. Slany, J. Frampton and S. Dumon (2012). "Distinct regulation of c-myb gene expression by HoxA9, Meis1 and Pbx proteins in normal hematopoietic progenitors and transformed myeloid cells." *Blood Cancer J* **2**(6): e76.
- Derenzini, E., A. Rossi and D. Trere (2018). "Treating hematological malignancies with drugs inhibiting ribosome biogenesis: when and why." *J Hematol Oncol* **11**(1): 75.
- Deshpande, A. J., L. Chen, M. Fazio, A. U. Sinha, K. M. Bernt, D. Banka, S. Dias, J. Chang, E. J. Olhava, S. R. Daigle, V. M. Richon, R. M. Pollock and S. A. Armstrong (2013). "Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l." *Blood* **121**(13): 2533-2541.
- DiMartino, J. F., T. Miller, P. M. Ayton, T. Landewe, J. L. Hess, M. L. Cleary and A. Shilatifard (2000). "A carboxy-terminal domain of ELL is required and sufficient for immortalization of myeloid progenitors by MLL-ELL." *Blood* **96**(12): 3887-3893.
- Fedele, M., E. Crescenzi and L. Cerchia (2017). "The POZ/BTB and AT-Hook Containing Zinc Finger 1 (PATZ1) Transcription Regulator: Physiological Functions and Disease Involvement." *Int J Mol Sci* **18**(12).
- Franco, R., G. Scognamiglio, E. Valentino, M. Vitiello, A. Luciano, G. Palma, C. Arra, E. La Mantia, L. Panico, V. Tenneriello, A. Pinto, F. Frigeri, G. Capobianco, G. Botti, L. Cerchia, A. De Chiara and M. Fedele (2016). "PATZ1 expression correlates positively with BAX and negatively with BCL6 and survival in human diffuse large B cell lymphomas." *Oncotarget* **7**(37): 59158-59172.
- Ge, S. X. and D. Jung (2018). "ShinyGO: a graphical enrichment tool for ani-mals and plants." *bioRxiv*: 315150.
- Ge, S. X., D. Jung and R. Yao (2020). "ShinyGO: a graphical gene-set enrichment tool for animals and plants." *Bioinformatics* **36**(8): 2628-2629.
- Godfrey, L., N. T. Crump, S. O'Byrne, I. J. Lau, S. Rice, J. R. Harman, T. Jackson, N. Elliott, G. Buck, C. Connor, R. Thorne, D. Knapp, O. Heidenreich, P. Vyas, P. Menendez, S. Inglott, P. Ancliff, H. Geng, I. Roberts, A. Roy and T. A. Milne (2021). "H3K79me2/3 controls enhancer-promoter interactions and activation of the pan-cancer stem cell marker PROM1/CD133 in MLL-AF4 leukemia cells." *Leukemia* **35**(1): 90-106.
- Kaderi, M. A., M. Kanduri, A. M. Buhl, M. Sevov, N. Cahill, R. Gunnarsson, M. Jansson, K. E. Smedby, H. Hjalgrim, J. Jurlander, G. Juliusson, L. Mansouri and R. Rosenquist (2011). "LPL is the strongest prognostic factor in a comparative analysis of RNA-based markers in early chronic lymphocytic leukemia." *Haematologica* **96**(8): 1153-1160.

- Kent, L. N. and G. Leone (2019). "The broken cycle: E2F dysfunction in cancer." Nat Rev Cancer **19**(6): 326-338.
- Kuriyama, M., N. Harada, S. Kuroda, T. Yamamoto, M. Nakafuku, A. Iwamatsu, D. Yamamoto, R. Prasad, C. Croce, E. Canaani and K. Kaibuchi (1996). "Identification of AF-6 and canoe as putative targets for Ras." J Biol Chem **271**(2): 607-610.
- Laham-Karam, N., G. P. Pinto, A. Poso and P. Kokkonen (2020). "Transcription and Translation Inhibitors in Cancer Treatment." Front Chem **8**: 276.
- Lavau, C., S. J. Szilvassy, R. Slany and M. L. Cleary (1997). "Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL." EMBO J **16**(14): 4226-4237.
- Li, D., Y. Hu, Z. Jin, Y. Zhai, Y. Tan, Y. Sun, S. Zhu, C. Zhao, B. Chen, J. Zhu, Z. Chen, S. Chen, J. Li and H. Liu (2018). "TanCAR T cells targeting CD19 and CD133 efficiently eliminate MLL leukemic cells." Leukemia **32**(9): 2012-2016.
- Liedtke, M., P. M. Ayton, T. C. Somerville, K. S. Smith and M. L. Cleary (2010). "Self-association mediated by the Ras association 1 domain of AF6 activates the oncogenic potential of MLL-AF6." Blood **116**(1): 63-70.
- Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.
- Manara, E., E. Baron, C. Tregnago, S. Aveic, V. Bisio, S. Bresolin, R. Masetti, F. Locatelli, G. Basso and M. Pigazzi (2014). "MLL-AF6 fusion oncogene sequesters AF6 into the nucleus to trigger RAS activation in myeloid leukemia." Blood **124**(2): 263-272.
- Martineau, M., R. Berger, D. M. Lillington, A. V. Moorman and L. M. Secker-Walker (1998). "The t(6;11)(q27;q23) translocation in acute leukemia: a laboratory and clinical study of 30 cases. EU Concerted Action 11q23 Workshop participants." Leukemia **12**(5): 788-791.
- Numata, A., H. S. Kwok, A. Kawasaki, J. Li, Q. L. Zhou, J. Kerry, T. Benoukraf, D. Bararia, F. Li, E. Ballabio, M. Tapia, A. J. Deshpande, R. S. Welner, R. Delwel, H. Yang, T. A. Milne, R. Taneja and D. G. Tenen (2018). "The basic helix-loop-helix transcription factor SHARP1 is an oncogenic driver in MLL-AF6 acute myelogenous leukemia." Nat Commun **9**(1): 1622.
- Pallasch, C. P., J. Schwamb, S. Konigs, A. Schulz, S. Debey, D. Kofler, J. L. Schultze, M. Hallek, A. Ultsch and C. M. Wendtner (2008). "Targeting lipid metabolism by the lipoprotein lipase inhibitor orlistat results in apoptosis of B-cell chronic lymphocytic leukemia cells." Leukemia **22**(3): 585-592.
- Pauli, C., Y. Liu, C. Rohde, C. Cui, D. Fijalkowska, D. Gerloff, C. Walter, J. Krijgsveld, M. Dugas, B. Edemir, C. Pabst, L. P. Muller, F. Zhou and C. Muller-Tidow (2020). "Site-specific methylation of 18S ribosomal RNA by SNORD42A is required for acute myeloid leukemia cell proliferation." Blood **135**(23): 2059-2070.
- Pigazzi, M., R. Masetti, S. Bresolin, A. Beghin, A. Di Meglio, S. Gelain, L. Trentin, E. Baron, M. Giordan, A. Zangrando, B. Buldini, A. Leszl, M. C. Putti, C. Rizzari, F. Locatelli, A. Pession, G. Te Kronnie and G. Basso (2011). "MLL partner genes drive distinct gene expression profiles

and genomic alterations in pediatric acute myeloid leukemia: an AIEOP study." Leukemia **25**(3): 560-563.

Prasad, R., Y. Gu, H. Alder, T. Nakamura, O. Canaani, H. Saito, K. Huebner, R. P. Gale, P. C. Nowell, K. Kuriyama and et al. (1993). "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in acute myeloid leukemias with the t(6;11) chromosome translocation." Cancer Res **53**(23): 5624-5628.

Radziwill, G., R. A. Erdmann, U. Margelisch and K. Moelling (2003). "The Bcr kinase downregulates Ras signaling by phosphorylating AF-6 and binding to its PDZ domain." Mol Cell Biol **23**(13): 4663-4672.

Radziwill, G., A. Weiss, J. Heinrich, M. Baumgartner, P. Boisguerin, K. Owada and K. Moelling (2007). "Regulation of c-Src by binding to the PDZ domain of AF-6." EMBO J **26**(11): 2633-2644.

Rombout, A., B. Verhasselt and J. Philippe (2016). "Lipoprotein lipase in chronic lymphocytic leukemia: function and prognostic implications." Eur J Haematol **97**(5): 409-415.

Saha, S. K., S. M. R. Islam, K. S. Kwak, M. S. Rahman and S. G. Cho (2020). "PROM1 and PROM2 expression differentially modulates clinical prognosis of cancer: a multiomics analysis." Cancer Gene Ther **27**(3-4): 147-167.

Sakaguchi, S., M. Hombauer, I. Bilic, Y. Naoe, A. Schebesta, I. Taniuchi and W. Ellmeier (2010). "The zinc-finger protein MAZR is part of the transcription factor network that controls the CD4 versus CD8 lineage fate of double-positive thymocytes." Nat Immunol **11**(5): 442-448.

Smith, M. J., E. Ottoni, N. Ishiyama, M. Goudreault, A. Haman, C. Meyer, M. Tucholska, G. Gasmi-Seabrook, S. Menezes, R. C. Laister, M. D. Minden, R. Marschalek, A. C. Gingras, T. Hoang and M. Ikura (2017). "Evolution of AF6-RAS association and its implications in mixed-lineage leukemia." Nat Commun **8**(1): 1099.

Takahashi, K., H. Nakanishi, M. Miyahara, K. Mandai, K. Satoh, A. Satoh, H. Nishioka, J. Aoki, A. Nomoto, A. Mizoguchi and Y. Takai (1999). "Nectin/PRR: an immunoglobulin-like cell adhesion molecule recruited to cadherin-based adherens junctions through interaction with Afadin, a PDZ domain-containing protein." J Cell Biol **145**(3): 539-549.

Takahashi, S. and A. Yokoyama (2020). "The molecular functions of common and atypical MLL fusion protein complexes." Biochim Biophys Acta Gene Regul Mech **1863**(7): 194548.

Thorsteinsdottir, U., E. Kroon, L. Jerome, F. Blasi and G. Sauvageau (2001). "Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia." Mol Cell Biol **21**(1): 224-234.

Tolba, F. M., M. E. Foda, H. M. Kamal and D. A. Elshabrawy (2013). "Expression of CD133 in acute leukemia." Med Oncol **30**(2): 527.

Tregnago, C., A. Da Ros, E. Porcu, M. Benetton, M. Simonato, L. Simula, G. Borella, K. Polato, S. Minuzzo, G. Borile, P. Cogo, S. Campello, A. Massi, R. Romagnoli, B. Buldini, F. Locatelli and M. Pigazzi (2020). "Thioridazine requires calcium influx to induce MLL-AF6-rearranged AML cell death." Blood Adv **4**(18): 4417-4429.

- Trikha, P., N. Sharma, R. Opavsky, A. Reyes, C. Pena, M. C. Ostrowski, M. F. Roussel and G. Leone (2011). "E2f1-3 are critical for myeloid development." J Biol Chem **286**(6): 4783-4795.
- Valentino, T., D. Palmieri, M. Vitiello, G. M. Pierantoni, A. Fusco and M. Fedele (2013). "PATZ1 interacts with p53 and regulates expression of p53-target genes enhancing apoptosis or cell survival based on the cellular context." Cell Death Dis **4**: e963.
- Yao, J., L. Zhong, P. Zhong, D. Liu, Z. Yuan, J. Liu, S. Yao, Y. Zhao, M. Chen, L. Li, L. Liu and B. Liu (2019). "RAS-Responsive Element-Binding Protein 1 Blocks the Granulocytic Differentiation of Myeloid Leukemia Cells." Oncol Res **27**(7): 809-818.
- Yokoyama, A., M. Lin, A. Naresh, I. Kitabayashi and M. L. Cleary (2010). "A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription." Cancer Cell **17**(2): 198-212.
- Yu, G. and Q. Y. He (2016). "ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization." Mol Biosyst **12**(2): 477-479.
- Zhang, T., J. Yang, V. P. Vaikari, J. S. Beckford, S. Wu, M. Akhtari and H. Alachkar (2020). "Apolipoprotein C2 - CD36 Promotes Leukemia Growth and Presents a Targetable Axis in Acute Myeloid Leukemia." Blood Cancer Discov **1**(2): 198-213.

# **CHAPITRE 4**

## **LEUCÉMIE PÉDIATRIQUE : VERS DES MODÈLES PLUS PRÉCIS**

## **4 Leucémie pédiatrique : vers des modèles plus précis**

### **4.1 Préface**

Le but de cette revue de littérature est de cataloguer les avantages et les limites des différents modèles actuels utilisés pour définir les mécanismes et exigences génétiques de la leucémogénèse. Les lignées cellulaires fournissent un modèle qui est, certes, peu coûteux, mais qui ne permet pas de récapituler les aspects critiques de la biologie du cancer. Les modèles murins de leucémie sont très informatifs mais ne reproduisent pas non plus entièrement la maladie humaine. Des avancées plus récentes dans le développement de xénogreffes dérivées de cellules leucémiques de patients (PDX) ou de modèles de leucémies humaines sont en passe de fournir une approche plus complète pour évaluer directement l'impact de l'environnement *in vivo* dans le développement de la maladie.

Pour élucider les mécanismes moléculaires sous-jacents à la biologie de la leucémie, des efforts ont été déployés pour développer des modèles expérimentaux plus pertinents. Dans le chapitre 5, nous discuterons du modèle utilisé dans cette thèse et des propositions pour l'améliorer et être plus proche de la réalité clinique.

### **4.2 Contributions**

J'ai contribué à l'écriture du manuscrit conjointement avec Hera V Canaj, Chloe Villeneuve et Brian T Wilhelm. Aditi Ghosh a réalisé l'analyse des données publiques pour la conception de la **Figure 4.1**. Tous les auteurs ont discuté de la structure de la revue de littérature et ont révisé le manuscrit.

### **4.3 Référence de la publication**

Thomas Milan, Hera V Canaj, Chloe Villeneuve, Aditi Ghosh, Frédéric Barabé, Sonia Cellot, Brian T Wilhelm. "Pediatric leukemia: Moving toward more accurate models." *Experimental Hematology* (2019). PMID: 31154068.



## 4.4 Article

### **Pediatric leukemia: Moving towards more accurate models.**

**Thomas Milan**<sup>1</sup>, Hera V Canaj<sup>1\*</sup>, Chloe Villeneuve<sup>1</sup>, Aditi Ghosh<sup>1\*\*</sup>, Frédéric Barabé<sup>3,4</sup>, Sonia Cellot<sup>5</sup>, Brian T Wilhelm<sup>1,6</sup>

<sup>1</sup>Laboratory for High Throughput Biology, Institute for Research in Immunology and Cancer

<sup>2</sup>Centre de recherche en infectiologie du CHUL, Centre de recherche du CHU de Québec, Quebec City, QC, Canada

<sup>3</sup>CHU de Québec Hôpital Enfant-Jésus; Quebec City, QC, Canada

<sup>4</sup>Department of Medicine, Université Laval, Quebec City, QC, Canada

<sup>5</sup>Department of pediatrics, division of Hematology, Ste-Justine Hospital, Montréal, QC, Canada, Université de Montréal, Montréal, QC, Canada

<sup>6</sup>Department of Medicine, Université de Montréal, Montréal, QC, Canada.

\* The Chan Zuckerberg Biohub, San Francisco, California, USA.

\*\* EECS, U. C. Berkeley, California, USA

\*Correspondence to:

Brian T Wilhelm

Institute for Research in Immunology and Cancer

Université de Montréal

P.O. Box 6128, Station Centre-Ville

Montreal, Québec, Canada H3C 3J7

Phone : +1 514 343.6111 ext 0923

Fax : +1 514 343.7780

E-mail : [brian.wilhelm@umontreal.ca](mailto:brian.wilhelm@umontreal.ca)

#### **4.4.1 Abstract**

Leukemia is a complex genetic disease caused by errors in differentiation, growth, and apoptosis of hematopoietic cells in either lymphoid or myeloid lineages. Large-scale genomic characterization of thousands of leukemia patients has produced a tremendous amount of data that have enabled a better understanding of the differences between adult and pediatric patients. For instance, although phenotypically similar, pediatric and adult myeloid leukemia patients differ in their mutational profiles, typically involving either chromosomal translocations or recurrent single-base-pair mutations, respectively. To elucidate the molecular mechanisms underlying the biology of this cancer, continual efforts have been made to develop more contextually and biologically relevant experimental models. Leukemic cell lines, for example, provide an inexpensive and tractable model but often fail to recapitulate critical aspects of tumor biology. Likewise, murine leukemia models of leukemia have been highly informative but also do not entirely reproduce the human disease. More recent advances in the development of patient-derived xenografts (PDXs) or human models of leukemias are poised to provide a more comprehensive, and biologically relevant, approach to directly assess the impact of the *in vivo* environment on human samples. In this review, the advantages and limitations of the various current models used to functionally define the genetic requirements of leukemogenesis are discussed.

#### ***Key words***

Leukemia – Translocations – Cell lines – Murine models – Patient-derived xenografts

## 4.4.2 Introduction

Leukemia is a cancer of the blood, originating from hematopoietic stem and progenitor cells that lose their capacity for proper self-renewal, differentiation and apoptosis. While typically diagnosed in older adults (median age of ~68 (Appelbaum, Gundacker et al. 2006, Estey and Dohner 2006, Phekoo, Richards et al. 2006, Juliusson, Antunovic et al. 2009)) this complex genetic disease still remains one of the most common cancer during childhood, representing almost one third of all cancer diagnoses in children under the age of 15 (Belson, Kingsley et al. 2006, Siegel, King et al. 2014, Steliarova-Foucher, Colombet et al. 2017). Large-scale genomics studies of adult Acute Myeloid Leukemia (AML) patient cohorts have shown that it is a genetically heterogeneous disease, with a complex mutational landscape that has complicated efforts to develop broadly applicable targeted therapies (Cancer Genome Atlas Research, Ley et al. 2013). Evidence suggests that in adult patients, the gradual acquisition of random mutations, some of which have oncogenic activity, over several decades can convert a normal hematopoietic stem cell into an unregulated leukemic blast (Welch, Ley et al. 2012). Recently, similar studies of large cohorts of pediatric acute myeloid leukemia patients have confirmed past observations that, in contrast to the gradual accumulation of mutations seen in adult AML, in younger AML patients, chromosomal translocations are most often responsible for the development of the disease (Bolouri, Farrar et al. 2018), in contrast to pediatric vs adult ALL (Liu, Wang et al. 2016). Even mutations that have been shown to be recurrent in both adult and pediatric AML patients occur with highly divergent frequencies (Bolouri, Farrar et al. 2018). Taken together, current evidence suggests that despite the common phenotypes of pediatric and adult AML, they largely represent distinct genetic diseases with chromosomal rearrangements likely representing the initiating (or driver) event in a large fraction of childhood leukemia development.

Patient outcomes for childhood leukemia have significantly improved in the last 40 years (Mussai, Yap et al. 2014), largely through improved treatment modulation, however they still remain lower in AML than for most other pediatric cancers (NCI 2017). In addition, patient outcomes are highly dependent on the specific subgroup of leukemia involved (Bolouri, Farrar et al. 2018), suggesting that distinct mutational profiles, and therefore therapeutic targets, exist. As noted above, individual genetic heterogeneity is observed in all leukemia subgroups,

confounding efforts to define the essential genetic elements required for leukemias with specific translocations or other genetic driver events. It is therefore critical to develop fully representative models of each leukemia type, such that the genetic background noise that comes with studying individual patient samples can be effectively removed.

Among the first experimental tools developed to model the biology underlying the disease were cell lines established from the *in vitro* culture of primary patient samples. After their initial establishment, leukemic cell lines represented a relatively simple and cost-effective way to study the disease, and have formed the backbone of much of the experimental work performed to date. More recently, large-scale studies of leukemic cell lines have also provided deeper insight into drug sensitivities (Barretina, Caponigro et al. 2012), and genetic networks that are active within cell lines (Yang, Li et al. 2018, Zhang, Wang et al. 2018). Despite their utility, all cell lines have undergone selection for their ability to grow in an *in vitro* environment, typically without growth factors or other signals primary leukemias typically receive in the bone marrow niche. As a result, the gene expression patterns and underlying molecular biology often are not representative of the disease seen in patients.

In parallel efforts, concerted studies aimed at better characterizing mouse hematopoietic stem/progenitor cells (HSPCs) have subsequently allowed the development of murine leukemic models to explore disease mechanisms *in vivo* (Okada, Nakauchi et al. 1992, Dykstra, Ramunas et al. 2006, Challen, Boles et al. 2009). Such models represent an extremely powerful tool, not only for defining the genetic dependencies of leukemias, but also for providing an experimental system in which to evaluate therapeutic approaches. Although a range of leukemic models have been developed, many of these involve the use of the potentially oncogenic gene fusions typically seen in pediatric leukemias (Corral, Lavenir et al. 1996, Forster, Pannell et al. 2003, Bijl, Sauvageau et al. 2005, Dang, Nance et al. 2017). While potentially closer to the human disease than cell lines because of the lack of selection for *in vitro* growth, differences in genome structure, life span and transformability of mouse cells, mean that some of these models still do not fully recapitulate the human disease. Clearly, the ability to generate human models of pediatric leukemia would represent an optimal experimental system; however, the practical development of such models has been hampered by technical challenges in isolating and transforming the correct HSPC population and identifying the correctly optimized culture

conditions, and the availability of suitable mouse strains. Once again, these road blocks have been overcome by transducing HSPCs with potentially oncogenic fusion genes such as KMT2A-MLLT3 (MLL-AF9) allowing engineered human leukemias to be produced (Barabe, Kennedy et al. 2007).

In this review, we discuss the various benefits and drawbacks of the various different model systems available to study pediatric leukemias, including cell lines and human models. Additionally, we present descriptions of the current public patient-derived xenograft (PDX) resources available, and describe their impact on the development of novel therapeutics.

### 4.4.3 Mining patient cohorts to understand the disease

In the past decade, rapid advances in DNA sequencing technology have fueled the field of cancer genomics, enabling the comprehensive sequencing of large cohorts of patient tumors and matched normal tissue (Stratton, Campbell et al. 2009, International Cancer Genome, Hudson et al. 2010). The vast data sets created through these efforts has opened a powerful window to understanding the genetics of cancer biology including recurrent mutations, gene expression and chromatin accessibility (Hoadley, Yau et al. 2018). Comparisons of the mutational burden in different cancer types have shown that acute myeloid leukemias have significantly fewer somatic mutations present relative to other cancers (Alexandrov, Nik-Zainal et al. 2013), with pediatric AML having even lower levels than adult AMLs (Bolouri, Farrar et al. 2018). Despite this fact, high levels of genetic heterogeneity exist between all cancer patients, including pediatric leukemia patient samples. Such heterogeneity includes not only recurrent somatic mutations, but also the 200-300 (Genomes Project, Abecasis et al. 2010, Shen, Li et al. 2013) loss-of-function variants and thousands of structural variations (Sudmant, Rausch et al. 2015) estimated to be present in each human genome. Clearly this genetic variation has the potential to influence the development of leukemia, though likely through complex interactions of other variants or somatic mutations, which may be patient specific. Although studies of patient cohorts provide an important “average view” of a cancer, a comprehensive understanding of the contribution of genetic drivers in individual patients (i.e. how all of their somatic and germline variants contribute to the leukemia) is often missing.

Several large-scale genomics studies, focusing on pediatric cancers (Grobner, Worst et al. 2018) or specifically on leukemias (Roberts, Morin et al. 2012, Jaffe, Wang et al. 2013, Andersson, Ma et al. 2015, Hsu, Nguyen et al. 2015, Schwartz, Ma et al. 2017, Bolouri, Farrar et al. 2018), have been published in which the frequency of recurrent translocations is typically high. One landmark project, the TARGET (Therapeutically Applicable Research to Generate Effective Treatments) program, is managed by the National Cancer Institute, the Children’s Oncology group, St. Jude Children’s Hospital and others, and has the goal of using a multi-omics approach to comprehensively define all the molecular changes that are specifically implicated in the initiation and progression of childhood cancer. The data from this project, much of it publically available, has been of significant value in investigating common disease

mechanisms and identifying novel therapeutic strategies for hard-to-treat pediatric cancers. For instance, a pan-cancer genome analysis of 1,699 pediatric leukemias reported 142 potential driver genes, of which only 45% were also found in adult patients, highlighting the need for the development of distinct treatment strategies for young patients (Ma, Liu et al. 2018). Another more specifically focused study characterized nearly 1,000 young and adult pediatric patients and provided an outstanding molecular landscape of AML (Bolouri, Farrar et al. 2018). Supporting previous evidence on the differences between adult and pediatric leukemia, the authors provided clear evidence for age-specific differences in the spectrum of mutations and genetic variations present. Mutations that are typically found at high frequency in adult patients (e.g. *FLT3*, *NPM1*, *IDH1*, and *IDH2*) are rare in pediatric patients, whereas the frequency of mutations in specific signaling pathways (e.g. *KRAS*, *NRAS*, *KIT*, *WT1*) is significantly higher (Bolouri, Farrar et al. 2018). Moreover, the incidence of recurrent translocations in AML also follows an age-specific pattern, a fact likely linked to the early development of leukemia. Of all gene fusions, those involving the *KMT2A* (Mixed Lineage Leukemia, *MLL*) gene are the most common translocations seen in infants and older children (Bolouri, Farrar et al. 2018, Brunner and Graubert 2018). With respect to recurrent fusions in pediatric B-cell (B-ALL) and T-cell (T-ALL) acute lymphoblastic leukemia, many of these have been identified and functionally characterized as well. For instance, the *ETV6-RUNX1* (*TEL/AML1*) fusion resulting from the often cryptic t(12;21) translocation is one of the most common fusions and is found in 25% of pediatric ALL patients (Zelent, Greaves et al. 2004), where its occurrence during pregnancy may provide a “first hit” for the leukemia (Mori, Colman et al. 2002). The well-documented Philadelphia chromosome resulting from the fusion between *ABL1* and *BCR* (t(9;22)(q34;q11)), which is often observed in adult B-ALL, is much less frequent in pediatric cases B-ALL (~25% vs 4% of cases (Bernt and Hunger 2014)). Interestingly however, in Philadelphia chromosome-like B-ALL (representing ~20% of all pediatric patients (Reshmi, Harvey et al. 2017)), fusions of the *CRLF2* gene were seen in 43% of patients, whereas those without *CRLF2* fusions were frequently characterized by *JAK2/EPOR* fusions (~8% of cases) (Reshmi, Harvey et al. 2017). The most common gene fusion in pediatric T-ALL involves the *STIL-TAL1* genes leading to an overexpression of *TAL1* (along with frequent overexpression of *SLC17A9*) not seen to the same extent in adult T-ALL (Chen, Jiang et al. 2018), while a diverse collection of other recurrent fusions have been identified (Papenhausen, Kelly et al. 2019).

Another large-scale genomics effort, the Pediatric Cancer Genome Project (PCPG) cooperatively developed in 2010 by St. Jude Children's Research Hospital and Washington University analyzed the genomes of childhood cancer patients and identified somatic mutations that drive cancer (Downing, Wilson et al. 2012). With roughly 5,000 patient samples from more than a dozen cancer types, including more than 2,300 leukemic samples, these data represent another highly informative source for pediatric cancers, complementing others available through the US National Human Genome Research Institute (NHGRI), National Cancer Institute (NCI), The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC) and the TARGET project. Importantly, the findings from all of these studies have emphasized the challenges involved in analyzing the complex genetic data, where age-dependent differences complicate comparisons between adult and pediatric patients. For example, although the frequency of somatic mutation is lower in pediatric AML patients than in adult patients, there may be a greater role for germline variants that predispose pediatric patients to AML. Such predisposition variants have been identified in pediatric ALL (Churchman, Qian et al. 2018, Qian, Cao et al. 2018), AML (Phillips, Gerbing et al. 2010), and multiple myeloma (Wei, Calvo-Vidal et al. 2018) patients and are estimated to occur in ~10% of all pediatric cancers (Zhang, Walsh et al. 2015). Given that no comprehensive list of such variants exists for pediatric AML, and that the 10% frequency reported may be an underestimate (Brodeur, Nichols et al. 2017), comparing the genetics of pediatric AML and adult AML through recurrently mutated genes is problematic. This, in turn, has highlighted the critical need for better models of the disease to be able to functionally study the role of specific mutations in the disease and to develop genotype-specific treatments.

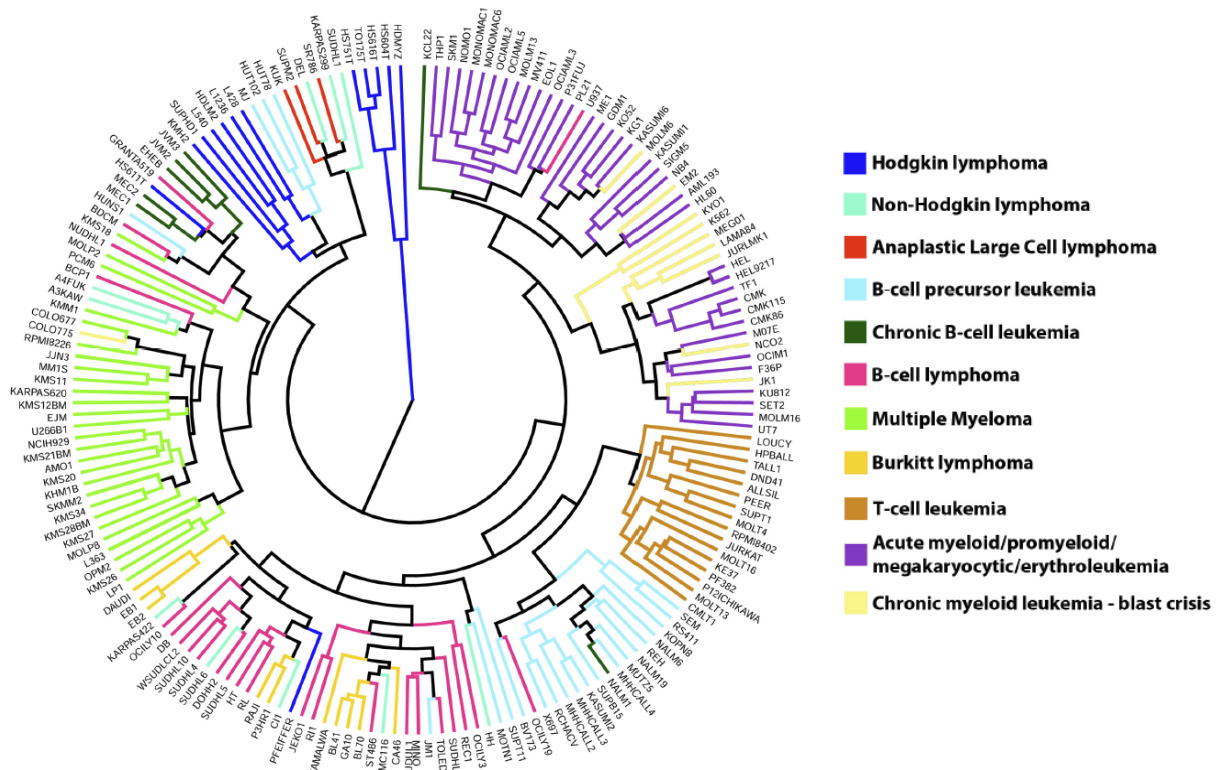


#### 4.4.4 Using cell lines as model systems for pediatric leukemias

Long before the advent of recombinant DNA technologies, immortal cell lines derived from patient tumors had already been developed as a model to try to better understand the biology of cancer. Over the years, hundreds of cell lines have been established from a range of different tumor types, and like current cancer patient samples, NGS approaches have recently been applied to provide a detailed molecular characterization of large sets of these cells (Barretina, Caponigro et al. 2012, Garnett, Edelman et al. 2012, Klijn, Durinck et al. 2015, Iorio, Knijnenburg et al. 2016). With respect to pediatric leukemias, a number of cell lines have been established from patients (Drexler, Quentmeier et al. 2004), many of which contain recurrent gene fusions seen in many patient leukemias. At the same time, this effort has been complicated by the fact that primary leukemic cells will rapidly differentiate or die when placed into *in vitro* culture. Even those cells which do not immediately die must often be cultured for months (Munker, Nordberg et al. 2009) before a proliferating clone emerges, resulting in a highly inefficient process for generating cell lines. Once established, such cell lines have, however, provided a useful, inexpensive and simple model to study defects in cell differentiation and self-renewal leading to the development of a specific leukemia. They also represent a powerful system for high-throughput chemical or genetic screens and for testing the impact of specific mutations. Moreover, the large number of different cell lines representing various tissues provides the possibility of not only assessing the consistency and reproducibility of results but also examining the tissue-specific aspects of experiments.

Despite their immense value, virtually all cell lines are in fact far removed, biologically speaking, from the actual human disease for a number of reasons. For instance, by virtue of how they are established, all tumor cells are selected for the “unnatural” characteristic of being able to grow *in vitro*. For leukemic cells, basic *in vitro* culture conditions typically lead to rapid cell differentiation and death without the addition of additional growth factors or stromal support cells (Pabst, Krosel et al. 2014). As a result, these stringent growth conditions can rapidly select for any subclones harboring mutations that allow them to survive, even if these would have no advantage *in vivo*. Moreover, the constant passage of cells also has the potential to apply selective pressure for faster growth, leading to the accumulation of additional mutations or genomic instability not seen in the primary disease and is often reflected in their highly variable

karyotypes. There is clear evidence of the dramatic effects of this process in other cancer cell lines (Ben-David, Siranosian et al. 2018) and although evidence for similar genetic heterogeneity has not yet been described for AML cell lines, an immortal and continually growing cell line with finite DNA replication fidelity must inevitably acquire novel mutations. In addition, the karyotypic abnormalities seen in large numbers of AML cell lines (Moy, Oleykowski et al. 2011) likely also reflects an inherent level of genomic instability that can affect the phenotype of the cell lines. For instance, analysis of the gene expression patterns of ~200 hematopoietic cell lines reveals that their morphological classification does not always correlate well with their clustering based on expression patterns of different leukemia subtypes (**Figure 4.1**). As to whether any specific AML cell line may be a more representative model of pediatric AML, it is difficult to say, and this likely depends on the presence of specific mutations or biological activities that are being examined.



**Figure 4.1 – Hierarchical clustering of hematopoietic cell lines.**

Published gene expression data from the Cancer Cell Line Encyclopedia analysis of (~200) established hematopoietic cell lines was used to perform hierarchical clustering based on the expression levels of the 1,000 most variable genes across all cell lines. Cell type annotation derived from published descriptions as well as information from commercial cell repositories was used to color the dendrogram branches for the cell types noted in the legend.

Attempts to overcome these inherent limitations of cell lines have largely focused on using large numbers of cells to observe the general trends despite the genetic noise present. Several large-scale studies, using hundreds of cell lines, have been performed to test tumor response to drugs to identify novel therapeutic targets (Barretina, Caponigro et al. 2012, Iorio, Knijnenburg et al. 2016). While similar caveats remain (e.g. that drugs showing therapeutic effects *in vitro* may not have the same consequences on cancerous cells because of differences in the *in vivo* microenvironment (Chang, Hernandez et al. 2019)), these data can still be informative for understanding the relevance of specific biological pathways. Human AML cells lines bearing gene fusions commonly seen in pediatric leukemias have been the focus of a number of high throughput studies aimed at establishing a catalog of genetic vulnerabilities. Approaches using both shRNA (Porter, Kim et al. 2012, Li, Mar et al. 2017, Heshmati, Turkoz et al. 2018) and CRISPR-Cas9-based (Tzelepis, Koike-Yusa et al. 2016, Wang, Yu et al. 2017) screening approaches have provided useful insight into the biology of pediatric leukemia despite clear differences caused by cell line-specific genetic heterogeneity (e.g. both differences between cell lines and subclonal heterogeneity with cell lines). Taken together, even though cell lines provide a useful resource to study the genetics of leukemias, they still present potential limitations for specific biological questions, and studies using individual cell lines may not provide accurate data for attempts to recapitulate the disease *in vivo*.

#### 4.4.5 Murine models of pediatric AML

Given the inherent limitations of cell line models of pediatric leukemia, the generation of novel models starting from primary cells has been of great interest. Such models have in turn been dependent on parallel technical advances in the ability to isolate and characterize murine progenitor cells that are competent to be transformed. Initial efforts to transform murine cells involved the prototypical *BCR-ABL* fusion seen in adult chronic myelogenous leukemia (CML) and used retroviral transduction of murine bone marrow (BM) cells (Daley, Van Etten et al. 1990, Kelliher, McLaughlin et al. 1990). These models illustrated that it was feasible to introduce specific oncogenes into murine cells to transform the cells into a state that can mimic the human disease. Subsequent models established through the homologous recombination of common pediatric gene fusions such as the *KMT2A* gene were performed using murine embryonic stem (ES) cells and indicated that, phenotypically, the transformed cells resembled human AML (Corral, Lavenir et al. 1996). The high frequency of *KMT2A* translocations in pediatric leukemias (Winters and Bernt 2017), its astonishing range of fusion partners (Meyer, Burmeister et al. 2018), and the success of initial studies all serve to explain the numerous subsequent efforts to study similar models with *KMT2A* fusions with *ENL* and *MLLT2*, among others. Interestingly, a number of oncogenic fusions seen in patient leukemias are insufficient by themselves to transform murine HSPCs (e.g. *ETV6-RUNX1*) or do so after long latency periods (e.g. *KMT2A-MLLT2*). This observation suggests that in some cases, the ability to accurately recapitulate certain human leukemias will likely require a better understanding and quantification of the contribution of typically heterogenic mutations seen in patient samples or isolation of relevant cell of origin for distinct leukemia subtypes.

The cell-of-origin used for the model leukemias that have been developed is also an important consideration for the accurate duplication of disease phenotypes across species. Although some earlier models using genetically manipulated embryonic stem cells from transgenic mice have been found to generate leukemias, most fusions in pediatric leukemias are typically tumor restricted along with recurrent somatic mutations. The continual refinement of cell surface markers for murine HSPCs has meant that many of the later models have specifically targeted these cells for transformation. Despite this more precise targeting, even relatively pure HSPC populations remain heterogeneous with a range of differentiation potential (Dykstra, Kent

et al. 2007) whose impact on leukemia development (e.g. lineage phenotype) remains undefined. In the case of particularly potent oncogenes such as *KMT2A* fusions, the cell-of-origin (or even the promoter driving the fusion (Metzler, Forster et al. 2006)) may have a negligible impact on leukemia development. This may not be the case, however, for other oncogenic fusions that have been used to model pediatric leukemias, including *AML1-ETO* (de Guzman, Warren et al. 2002), *E2A-PBX1* (Duque-Afonso, Feng et al. 2015), and, more recently, models looking at higher risk groups such as acute megakaryoblastic leukemia (AMKL) (Dang, Nance et al. 2017).

The creation of murine models of pediatric leukemia has been particularly informative for performing functional studies to define the essential genetic requirements for leukemia development from normal progenitor cells. The ability to perform these experiments iteratively, under a range of different conditions of genetic perturbation, provides significant flexibility to investigate specific genes. Although similar experiments can be performed in individual leukemia cell lines, in this case the leukemia is already maintained by a specific collection of genetic lesions that may or may not be relevant for a given patient subgroup, limiting the scope of experiments. At the same time, because mouse models are typically designed to address the role of specific proteins or mutations, testing complex combinations of the 4-10 driver mutations found in adult AMLs (Metzeler, Herold et al. 2016) is not feasible (**Table 4.1, Table 4.2**)

Year	Driver used	Phenotype of model	Notes and reference
2018	<i>GATA2-HOX49, MNI-FLI, NIPBL-HOXB, (CBFA2T3-GLIS2)</i>	AMKL	<i>CBFA2T3-GLIS2</i> fusion alone is insufficient to generate leukemia (Dang, Nance et al. 2017)
2017	<i>ETV6-RUNX1</i>	B-Cell precursor ALL	Loss of function in <i>KDM</i> family genes potentially relevant for disease penetrance (Rodriguez-Hernandez, Hauer et al. 2017)
2015	<i>E2A-PBX1</i>	B-Cell precursor ALL	Pre-BCR signaling/JAK kinases are potential therapeutic targets (Duque-Afonso, Feng et al. 2015)
2013	<i>KMT2A-MLLT4</i>	AML	Selective sensitivity to Dot11 inhibitor (EPZ0004777) (Deshpande, Chen et al. 2013)
2009	<i>ETV6-RUNX1</i>	Increased HSPC frequency but no leukemia	Chemical mutagenesis required for leukemia development (Bernardin, Yang et al. 2002)
	<i>RBM15-MKLI</i>	AMKL	Notch signalling (RBPJ) and cytokine (MPL) signaling are required for AMKL generation (Mercher, Raffel et al. 2009)
2008	<i>KMT2A-MLLT2</i>	B-ALL/AML	H3K79 methylation patterns correlate with gene expression patterns and are conserved between human and mouse samples (Krivtsov, Feng et al. 2008)
2006	<i>KMT2A-MLLT2</i>	B-ALL	Expression of <i>KMT2A-MLLT2</i> restricted to lymphoid cells, drives B-ALL development (Metzler, Forster et al. 2006)
	<i>KMT2A-MLLT2</i>	B-ALL/AML	Both <i>KMT2A-MLLT2/KMT2A-MLLT3</i> transform BM cells, but <i>KMT2A-MLLT3</i> has a shorter latency and favors AML vs. B-ALL/MPAL for <i>KMT2A-MLLT2</i> (Chen, Li et al. 2006)

2005	<i>AML1-ETO</i>	AML	Demonstrated that <i>AML1-ETO</i> fusions collaborate with FLT3 internal tandem duplications to generate AML (Schessl, Rawat et al. 2005)
2003	<i>KMT2A-MLLT1</i>	AML	Rapid highly penetrant leukemias result from Cre-loxP-mediated reciprocal chromosomal translocations (Forster, Pannell et al. 2003)
	<i>NUP98-HOXD13</i>	Myeloproliferation/ AML	Full transformation of BM cells into AML was dependent on co-expression of <i>MEIS1</i> (Pineault, Buske et al. 2003)
2002	<i>ETV6-RUNX1</i>	B-ALL/T-ALL	Development of leukemia is potentiated by loss of p16/p19 (Bernardin, Yang et al. 2002)
	<i>AML1-ETO</i>	AML	Altered HSC compartment size but without leukemia of disseminated disease (de Guzman, Warren et al. 2002)
2001	<i>NUP98-HOX49</i>	AML	Induces a polyclonal AML and defined a role for MEIS1 cofactor in accelerating the disease (Kroon, Thorsteinsdottir et al. 2001)
1996	<i>KMT2A-MLLT3</i>	AML	First mouse model of AML generated through Cre-loxP mediated fusion in ES cells (Corral, Lavenir et al. 1996)

**Table 4.1 – Mouse models of leukemia.**



Year	Driver used	Phenotype of model	Notes and reference
2018	<i>KMT2A-MLLT3</i>	AML	CRISPR-based editing can be used in human hematopoietic stem and progenitor cells to generate leukemias (Schneidawind, Jeong et al. 2018)
2017	<i>RBM15-MKL1</i>	HSPCs with AMKL-like expression	Two HSPC cell lines expressing <i>RBM15-MKL1</i> oncogenic fusion seen exhibit some gene expression patterns similar to that of AMKL (Ayllon, Vogel-Gonzalez et al. 2017)
	<i>KMT2A-MLLT3</i>	AML/B-ALL	RET highlighted as a therapeutic target and sequencing data showing the <i>KMT2A-MLLT3</i> fusion alone is sufficient to generate leukemias (Barabe, Gil et al. 2017)
2015	<i>KMT2A-MLLT3/MLLT1</i>	ALL/AML/MPAL	TALEN induced gene editing of the endogenous <i>KMT2A</i> gene in CD34 cells (Buechele, Breese et al. 2015)
2014	<i>NUP98-HOXD13</i>	AML	Forced expression of <i>MNI</i> is insufficient to generate leukemias in the absence of <i>HOX</i> gene fusion (Imren, Heuser et al. 2014)
2012	<i>KMT2A-MLLT2</i>	Block in HSC commitment	Enforced expression of <i>KMT2A-MLLT2</i> in hESCs impairs hematopoietic development and is insufficient to transform cells (Bueno, Montes et al. 2012)
	<i>KMT2A-MLLT3</i>	AML	Intrinsic properties of the cell of origin impact the efficiency of transformation in human model leukemias (Horton, Jaques et al. 2013)
	<i>KMT2A-MLLT10</i>	HSPCs with <i>KMT2A-MLLT10</i> and activated K-ras develop AML	<i>KMT2A-MLLT10</i> alone is insufficient to generate AML in humanized mice (Moriya, Suzuki et al. 2012)
2011	<i>KMT2A-MLLT2</i>	HSPCs with increased proliferation and clonogenic potential	<i>KMT2A-MLLT2</i> enhances proliferation but is insufficient to generate a leukemia in human cells (Montes, Ayllon et al. 2011)
2010	<i>BCR-ABL</i>	CML	<i>BMI1</i> collaborates with <i>BCR-ABL</i> in leukemia development; lymphoid phenotype <i>in vivo</i> with myeloid/lymphoid cells established in culture (Rizo, Horton et al. 2010)

2008	<i>KMT2A- MLLT3</i>	AML/ALL/MPAL	Phenotype of transformed CB cells is impacted by growth factors, microenvironment and mouse strain used (Wei, Wunderlich et al. 2008)
2007	<i>KMT2A- MLLT3/ MLLT1</i>	AML/B-ALL	First human model of AML developed using CB cells (Barabe, Kennedy et al. 2007)

**Table 4.2 – Human models of leukemia.**

#### 4.4.6 Human models of pediatric leukemia

One of the most significant constraints on the functional study of pediatric leukemias, aside from their rarity, is the fact that there is typically very little material available for studies. To overcome this challenge, immunocompromised (NOD-SCID, NSG (NOD.Cg-*Prkdc*<sup>scid</sup>*Il2rg*<sup>tm1Wjl</sup>/SzJ)) mice have been used to expand these limited tumor cells through xenografting of primary patient cells. In the case of both ALL (Baersch, Mollers et al. 1997, Lock, Liem et al. 2002) and AML (Krivtsov, Wang et al. 2014), a leukemia similar to the patient sample can be expanded. These PDXs have differing success rates (60-75% (Xujun, Farnoud et al. 2013, Vick, Rothenberg et al. 2015, Wang, Sanchez-Martin et al. 2017)) based on the permissibility of the recipient mice (Goyama, Wunderlich et al. 2015, Byrne, Alferez et al. 2017), the specific leukemia sample used, and the presence of cytokines and growth factors that can influence engraftment success (Wunderlich, Chou et al. 2010). Such PDX models present multiple unique advantages, both in fundamental research and in a clinical setting, including the ability to preserve stromal components in the context of orthotopic engraftment and the ability to be used in *in vivo* functional studies including screening of drugs capable of targeting the patient's tumor (Weroha, Becker et al. 2014, Pompili, Porru et al. 2016, Byrne, Alferez et al. 2017). Given the potential value of PDX samples, it is not surprising that numerous academic and commercial repositories have been established, including ProXe (Townsend, Murakami et al. 2016), HuBase (CrownBio), EurOPDX (Hidalgo, Amant et al. 2014), PROPEL (St Jude Children's hospital) along with meta-search services such as PDXFinder (Conte, Mason et al. 2018) (covering almost 2,000 available PDXs in eight repositories). These repositories, some of which include pediatric leukemia PDXs, allow researchers to easily obtain viable tumor samples, many of which have also undergone significant initial molecular or genomic characterization. These collections of PDX tumors, searchable by the tumor type, specific mutations present, and drug dosing information, give researchers unprecedented power in selecting relevant tumor samples for functional tests.

Although PDXs represent a powerful resource, questions remain as to the extent to which selective conditions within the mouse alter the phenotype of the original tumor and how stable the tumor genotype remains. For example, recent large-scale studies have reported a high frequency of copy number alternations within four cell passages, along with expansion of minor

clones within the initial tumor (Ben-David, Ha et al. 2017). And although in AML PDXs, a majority of driver mutations remain concordant with the primary tumor (even when serial transplantations are performed (Vick, Rothenberg et al. 2015)), some somatic variants present at a low frequency in the primary tumors (e.g., variant allele frequency: 10-30%) exhibit large changes (> two fold) in PDXs (Wang, Sanchez-Martin et al. 2017). Despite these observations, PDXs clearly provide a complementary approach to functional studies of patient tumors, and it remains to be seen how such alterations relate to similar changes seen during tumor evolution, driven by the selective pressure of patient treatment.

Patient-derived xenograft models such as those described above have been useful in removing some experimental constraints; however, the use of xenografts still presents a limitation with respect to accurately reproducing the human bone marrow niche and immune environment, which can play a critical role in tumor development and progression (Lane, Scadden et al. 2009, Kumar, Garcia et al. 2018, Lambie and Lind 2018). Moreover, although the engraftment of human peripheral blood mononuclear cells (PBMCs) or reactive T cells can provide some immune activity against the human tumor, the lack of human antigen presenting cells (APCs) and development of host-versus-graft responses limit the value of this approach. This limitation has been addressed by the development of a humanized mouse model, where sublethally irradiated immunodeficient mice received transplanted fetal thymic tissue (FTHY) and CD34<sup>+</sup> fetal liver cells (FLCs) (Brodeur, Nichols et al. 2017). Such humanized mice, when challenged by *KMT2A-MLLT3*-bearing leukemia cells, exhibited an immune response similar to that seen in patients while eliciting a more representative immunological response to the disease (Iorio, Knijnenburg et al. 2016).

Given the species differences between murine and human cells and the potential genetic differences induced through the generation of human PDXs, the ability to generate models of pediatric leukemia starting from normal human HSPCs would represent an attractive solution. Interestingly, similar genetic approaches using human hematopoietic progenitor cells have been slow to develop because of the technical challenges involved in achieving the levels of transformation seen in mouse progenitor cells. Although the differences responsible have not been completely defined, studies looking at different HSPC populations indicated that HSPC transformational competency decreases in cells with age (cord blood vs. bone marrow) (Horton,

Jaques et al. 2013) and location (bone marrow vs. peripheral blood) (Horton, Jaques et al. 2013). This observation also agrees with age-dependent changes in the competence of specific progenitor cells to be transformed in the case of pediatric leukemias that are extremely rare in adults (e.g. AMKL (Gruber and Downing 2015)).

One of the first successful demonstrations of a human model leukemia was performed using potent *KMT2A* fusions such as *MLLT3* (Barabe, Kennedy et al. 2007). The derived leukemias phenotypically resembled patient samples and were transplantable but, unlike primary AMLs, were also able to be maintained in *in vitro* culture for extended periods. Subsequent detailed genomics studies on similar single donor-derived leukemias also confirmed the generally held hypothesis that the fusion gene itself was sufficient to generate the leukemia and that no recurrent secondary mutations were required (Barabe, Gil et al. 2017). Other human leukemia model fusions have not only illustrated the importance of the fusion partner in the leukemia (Bueno, Montes et al. 2012, Buechele, Breese et al. 2015), but have also clearly highlighted differences between murine and human models with the same oncogene (Lin, Luo et al. 2016), underlining the importance of human model systems. As with the murine models, many of the oncogenic fusions used in initial human models involved the *KMT2A* gene, which has strong oncogenic potential. Although other fusions have been tested in human models, such as the *NUP98-HOXD13* fusion gene (Imren, Heuser et al. 2014) for AML and the *RBM15-MKLI* fusion gene (Ayllon, Vogel-Gonzalez et al. 2017) in the case of AMKL, these models have exhibited partial or weak leukemia penetrance or dependence on other genetic alterations to generate a leukemia. Given the fact that human HSPCs seem to be more resistant to transformation compared with their murine equivalent, it remains unclear how straightforward the development of other human models will be, even given the technical advances demonstrated through direct genome editing. Nevertheless, the existing models of pediatric leukemia indicate that with optimized constructs and culture conditions, it is possible to generate a human model of the disease that, importantly, closely recapitulates its behaviour and phenotype.

With the development of both murine and human models of pediatric leukemia, there has been a great interest in using these models for functional studies to uncover the genetic underpinnings of the disease. Although space constraints preclude including an exhaustive list

of these studies, a representative sample is presented in **Table 4.3** and **Table 4.4**. Overall, these efforts have provided highly valuable insight into the molecular mechanisms involved in specific models of leukemia and have resulted in the advancement of several specific compounds to clinical trials (Berthon, Raffoux et al. 2016, Shukla, Wetmore et al. 2016, Stein, Garcia-Manero et al. 2018). At the same time, given the variety of screening approaches applied to the models (including siRNA and shRNA knockdown, or CRISPR-based screens), it is perhaps not surprising that the consistency of essential targets identified is relatively low. In fact, this diversity recapitulates the range of inhibitory responses seen in adult primary AML samples when challenged with a diverse collection of small molecules (Baccelli, Krosi et al. 2017). If the underlying heterogeneity of somatic and germline variants in AML patients does explain the diversity of response seen, it suggests that although models of pediatric leukemia will be useful for dissecting specific pathways, their ability to identify “universal” therapeutic targets may be inherently limited. Moreover, the model leukemias used for such functional studies are heavily biased toward the subset in which strong oncogenic drivers have been validated. As a result, their ability to uncover relevant therapeutic targets for the range of other pediatric subgroups (that are perhaps less frequent but have clinical outcomes that are just as poor) remains unclear. These challenges notwithstanding, the ability to reproducibly generate a range of experimentally tractable human model leukemias, from multiple HSPC sources that could potentially be genetically manipulated beforehand (Bak, Dever et al. 2018), will provide an essential resource that may provide more flexibility than the generation of PDXs in which the engraftment is variable.

Year	Format	Method	Phenotype of cells screened	Screen hits
2018	<i>in vitro/in vivo</i>	CRISPR	AML	<i>DCPS</i> (Yamauchi, Masuda et al. 2018)
2017	<i>in vivo</i>	CRISPR	<i>KMT2A-MLLT3, HOXA9, Meis1</i>	<i>B4galt1</i> (Mercier, Shi et al. 2017)
2015	<i>in vitro/in vivo</i>	shRNA	B-ALL	<i>Phf6</i> (Meacham, Lawton et al. 2015)
2014	<i>in vitro</i>	shRNA	<i>KMT2A-MLLT3, HOXA9, Meis1</i>	<i>Jmjd1c</i> (Sroczynska, Cruickshank et al. 2014)
2013	<i>in vivo</i>	Small Molecule	<i>KMT2A-MLLT3</i>	Lovastatin (Hartwell, Miller et al. 2013)
	<i>in vivo</i>	shRNA	<i>KMT2A-MLLT3</i>	<i>Itgb3</i> (Miller, Al-Shahrour et al. 2013)
2011	<i>in vitro/in vivo</i>	Small Molecule	<i>KMT2A-MLLT3/MLLT1</i>	<i>Brd3/4</i> (Dawson, Prinjha et al. 2011)
	<i>in vitro</i>	shRNA	<i>KMT2A-MLLT3</i>	<i>Brd4</i> (Zuber, Shi et al. 2011)

**Table 4.3 – Screens performed using mouse leukemia models.**

Year	Format	Method	Phenotype of cells screened	Screen hits
2018	<i>in vitro</i>	shRNA	<i>KMT2A-MLLT3</i>	<i>CHD4</i> (Heshmati, Turkoz et al. 2018)
2017	<i>in vitro</i>	shRNA	<i>KMT2A-MLLT3</i>	<i>ZEB2</i> (Li, Mar et al. 2017)
	<i>in vitro</i>	Small Molecule	AML	Diverse (Baccelli, Krosi et al. 2017)
	<i>in vitro/in vivo</i>	shRNA	AML	<i>RET</i> (Barabe, Gil et al. 2017)
	<i>in vitro</i>	CRISPR	AML	<i>PREX1</i> (Wang, Yu et al. 2017)
2016	<i>in vitro</i>	CRISPR	5 leukemia cell lines	<i>DOT1L, BCL2, Men1</i> (Tzelepis, Koike-Yusa et al. 2016)
2015	<i>in vitro</i>	RNAi	AML	<i>BNIPL1, ROCK1, RPS13, STK3, SNX27, WDHD1</i> (Wermke, Camgoz et al. 2015)
2012	<i>in vitro</i>	shRNA	AML	<i>WEE1</i> (Porter, Kim et al. 2012)
	<i>in vitro</i>	shRNA	AML	<i>GSK-3<math>\alpha</math></i> (Banerji, Frumm et al. 2012)

**Table 4.4 – Screens performed using human leukemia models.**



#### **4.4.7 Conclusions**

Given the limits of either the static information from patient cohorts, or murine models that, in some cases, do not recapitulate all aspects of the human disease, the need to create new models that more closely reflect human cancer biology is clear. Such tools will be essential in understanding the stepwise requirements for leukemogenic transformation that will underpin the next generation of rationally designed, and genotype-specific, treatments. Many technical challenges remain, however, in identifying the genetic requirements of individual fusions to transform human progenitor cells to the correct phenotype. For this, the availability of both patient-derived xenograft models and engineered human model leukemias provides a range of biological tools to explore the functional dependencies. Whether PDX models or engineered leukemias are “superior” remains something of an open question, and so for the moment, it is most useful to view them simply as complementary systems, each well suited to solving different problems related to our incomplete understanding of leukemia biology and novel therapeutic avenues.

#### **4.4.8 Acknowledgements**

This work was supported through grant funding from the Canadian Cancer Society Research Institute Grant IMP-17 (BTW, SC, FB), the Cole Foundation (BTW, SC), La Fondation du Centre de Cancérologie Charles Bruneau and La Fondation CHU Sainte-Justine (SC).

We acknowledge members of the Wilhelm and Cellot labs for critical comments on the manuscript during revision.

#### 4.4.9 References

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, I. Australian Pancreatic Cancer Genome, I. B. C. Consortium, I. M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell and M. R. Stratton (2013). "Signatures of mutational processes in human cancer." Nature **500**(7463): 415-421.

Andersson, A. K., J. Ma, J. Wang, X. Chen, A. L. Gedman, J. Dang, J. Nakitandwe, L. Holmfeldt, M. Parker, J. Easton, R. Huether, R. Kriwacki, M. Rusch, G. Wu, Y. Li, H. Mulder, S. Raimondi, S. Pounds, G. Kang, L. Shi, J. Becksfort, P. Gupta, D. Payne-Turner, B. Vadodaria, K. Boggs, D. Yergeau, J. Manne, G. Song, M. Edmonson, P. Nagahawatte, L. Wei, C. Cheng, D. Pei, R. Sutton, N. C. Venn, A. Chetcuti, A. Rush, D. Catchpoole, J. Heldrup, T. Fioretos, C. Lu, L. Ding, C. H. Pui, S. Shurtleff, C. G. Mullighan, E. R. Mardis, R. K. Wilson, T. A. Gruber, J. Zhang, J. R. Downing and P. St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome (2015). "The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias." Nat Genet **47**(4): 330-337.

Appelbaum, F. R., H. Gundacker, D. R. Head, M. L. Slovak, C. L. Willman, J. E. Godwin, J. E. Anderson and S. H. Petersdorf (2006). "Age and acute myeloid leukemia." Blood **107**(9): 3481-3485.

Ayllon, V., M. Vogel-Gonzalez, F. Gonzalez-Pozas, J. Domingo-Reines, R. Montes, L. Morales-Cacho and V. Ramos-Mejia (2017). "New hPSC-based human models to study pediatric Acute Megakaryoblastic Leukemia harboring the fusion oncogene RBM15-MKL1." Stem Cell Res **19**: 1-5.

Baccelli, I., J. Kros, G. Boucher, I. Boivin, V. P. Lavalley, J. Hebert, S. Lemieux, A. Marinier and G. Sauvageau (2017). "A novel approach for the identification of efficient combination therapies in primary human acute myeloid leukemia specimens." Blood Cancer J **7**(2): e529.

Baersch, G., T. Mollers, A. Hotte, B. Dockhorn-Dworniczak, C. Rube, J. Ritter, H. Jurgens and J. Vormoor (1997). "Good engraftment of B-cell precursor ALL in NOD-SCID mice." Klin Padiatr **209**(4): 178-185.

Bak, R. O., D. P. Dever and M. H. Porteus (2018). "CRISPR/Cas9 genome editing in human hematopoietic stem cells." Nat Protoc **13**(2): 358-376.

Banerji, V., S. M. Frumm, K. N. Ross, L. S. Li, A. C. Schinzel, C. K. Hahn, R. M. Kakoza, K. T. Chow, L. Ross, G. Alexe, N. Tolliday, H. Inguilizian, I. Galinsky, R. M. Stone, D. J. DeAngelo, G. Roti, J. C. Aster, W. C. Hahn, A. L. Kung and K. Stegmaier (2012). "The

intersection of genetic and chemical genomic screens identifies GSK-3alpha as a target in human acute myeloid leukemia." *J Clin Invest* **122**(3): 935-947.

Barabe, F., L. Gil, M. Celton, A. Bergeron, V. Lamontagne, E. Roques, K. Lagace, A. Forest, R. Johnson, L. Pecheux, J. Simard, J. Pelloux, A. Bellemare-Pelletier, E. Gagnon, J. Hebert, S. Cellot and B. T. Wilhelm (2017). "Modeling human MLL-AF9 translocated acute myeloid leukemia from single donors reveals RET as a potential therapeutic target." *Leukemia* **31**(5): 1166-1176.

Barabe, F., J. A. Kennedy, K. J. Hope and J. E. Dick (2007). "Modeling the initiation and progression of human acute leukemia in mice." *Science* **316**(5824): 600-604.

Barretina, J., G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, Jr., M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palesscandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel and L. A. Garraway (2012). "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." *Nature* **483**(7391): 603-607.

Belson, M., B. Kingsley and A. Holmes (2006). "Risk Factors for Acute Leukemia in Children: A Review." *Environmental Health Perspectives* **115**(1): 138-145.

Ben-David, U., G. Ha, Y. Y. Tseng, N. F. Greenwald, C. Oh, J. Shih, J. M. McFarland, B. Wong, J. S. Boehm, R. Beroukhim and T. R. Golub (2017). "Patient-derived xenografts undergo mouse-specific tumor evolution." *Nat Genet* **49**(11): 1567-1575.

Ben-David, U., B. Siranosian, G. Ha, H. Tang, Y. Oren, K. Hinohara, C. A. Strathdee, J. Dempster, N. J. Lyons, R. Burns, A. Nag, G. Kugener, B. Cimini, P. Tsvetkov, Y. E. Maruvka, R. O'Rourke, A. Garrity, A. A. Tubelli, P. Bandopadhyay, A. Tsherniak, F. Vazquez, B. Wong, C. Birger, M. Ghandi, A. R. Thorner, J. A. Bittker, M. Meyerson, G. Getz, R. Beroukhim and T. R. Golub (2018). "Genetic and transcriptional evolution alters cancer cell line drug response." *Nature* **560**(7718): 325-330.

Bernardin, F., Y. Yang, R. Cleaves, M. Zahurak, L. Cheng, C. I. Civin and A. D. Friedman (2002). "TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice." *Cancer Res* **62**(14): 3904-3908.

Bernt, K. M. and S. P. Hunger (2014). "Current concepts in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia." *Front Oncol* **4**: 54.

Berthon, C., E. Raffoux, X. Thomas, N. Vey, C. Gomez-Roca, K. Yee, D. C. Taussig, K. Rezai, C. Roumier, P. Herait, C. Kahatt, B. Quesnel, M. Michallet, C. Recher, F. Lokiec, C. Preudhomme and H. Dombret (2016). "Bromodomain inhibitor OTX015 in patients with acute leukaemia: a dose-escalation, phase 1 study." *Lancet Haematol* **3**(4): e186-195.

Bijl, J., M. Sauvageau, A. Thompson and G. Sauvageau (2005). "High incidence of proviral integrations in the Hoxa locus in a new model of E2a-PBX1-induced B-cell leukemia." Genes Dev **19**(2): 224-233.

Bolouri, H., J. E. Farrar, T. Triche, Jr., R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvin, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard and S. Meshinchi (2018). "The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions." Nat Med **24**(1): 103-112.

Brodeur, G. M., K. E. Nichols, S. E. Plon, J. D. Schiffman and D. Malkin (2017). "Pediatric Cancer Predisposition and Surveillance: An Overview, and a Tribute to Alfred G. Knudson Jr." Clin Cancer Res **23**(11): e1-e5.

Brunner, A. M. and T. A. Graubert (2018). "Genomics in childhood acute myeloid leukemia comes of age." Nat Med **24**(1): 7-9.

Buechele, C., E. H. Breese, D. Schneidawind, C. H. Lin, J. Jeong, J. Duque-Afonso, S. H. Wong, K. S. Smith, R. S. Negrin, M. Porteus and M. L. Cleary (2015). "MLL leukemia induction by genome editing of human CD34+ hematopoietic cells." Blood **126**(14): 1683-1694.

Bueno, C., R. Montes, G. J. Melen, V. Ramos-Mejia, P. J. Real, V. Ayllon, L. Sanchez, G. Ligerio, I. Gutierrez-Aranda, A. F. Fernandez, M. F. Fraga, I. Moreno-Gimeno, D. Burks, C. Plaza-Calonge Mdel, J. C. Rodriguez-Manzanique and P. Menendez (2012). "A human ESC model for MLL-AF4 leukemic fusion gene reveals an impaired early hematopoietic-endothelial specification." Cell Res **22**(6): 986-1002.

Byrne, A. T., D. G. Alferez, F. Amant, D. Annibali, J. Arribas, A. V. Biankin, A. Bruna, E. Budinska, C. Caldas, D. K. Chang, R. B. Clarke, H. Clevers, G. Coukos, V. Dangles-Marie, S. G. Eckhardt, E. Gonzalez-Suarez, E. Hermans, M. Hidalgo, M. A. Jarzabek, S. de Jong, J. Jonkers, K. Kemper, L. Lanfrancone, G. M. Maeldansmo, E. Marangoni, J. C. Marine, E. Medico, J. H. Norum, H. G. Palmer, D. S. Peeper, P. G. Pelicci, A. Piris-Gimenez, S. Roman-Roman, O. M. Rueda, J. Seoane, V. Serra, L. Soucek, D. Vanhecke, A. Villanueva, E. Vinolo, A. Bertotti and L. Trusolino (2017). "Interrogating open issues in cancer precision medicine with patient-derived xenografts." Nat Rev Cancer **17**(4): 254-268.

Cancer Genome Atlas Research, N., T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J.

Sofia, L. Yang, J. R. Downing and G. Eley (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." N Engl J Med **368**(22): 2059-2074.

Challen, G. A., N. Boles, K. K. Lin and M. A. Goodell (2009). "Mouse hematopoietic stem cell identification and analysis." Cytometry A **75**(1): 14-24.

Chang, Y. T., D. Hernandez, S. Alonso, M. Gao, M. Su, G. Ghiaur, M. J. Levis and R. J. Jones (2019). "Role of CYP3A4 in bone marrow microenvironment-mediated protection of FLT3/ITD AML from tyrosine kinase inhibitors." Blood Adv **3**(6): 908-916.

Chen, B., L. Jiang, M. L. Zhong, J. F. Li, B. S. Li, L. J. Peng, Y. T. Dai, B. W. Cui, T. Q. Yan, W. N. Zhang, X. Q. Weng, Y. Y. Xie, J. Lu, R. B. Ren, S. N. Chen, J. D. Hu, D. P. Wu, Z. Chen, J. Y. Tang, J. Y. Huang, J. Q. Mi and S. J. Chen (2018). "Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia." Proc Natl Acad Sci U S A **115**(2): 373-378.

Chen, W., Q. Li, W. A. Hudson, A. Kumar, N. Kirchhof and J. H. Kersey (2006). "A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy." Blood **108**(2): 669-677.

Churchman, M. L., M. Qian, G. Te Kronnie, R. Zhang, W. Yang, H. Zhang, T. Lana, P. Tedrick, R. Baskin, K. Verbist, J. L. Peters, M. Devidas, E. Larsen, I. M. Moore, Z. Gu, C. Qu, H. Yoshihara, S. N. Porter, S. M. Pruett-Miller, G. Wu, E. Raetz, P. L. Martin, W. P. Bowman, N. Winick, E. Mardis, R. Fulton, M. Stanulla, W. E. Evans, M. V. Relling, C. H. Pui, S. P. Hunger, M. L. Loh, R. Handgretinger, K. E. Nichols, J. J. Yang and C. G. Mullighan (2018). "Germline Genetic IKZF1 Variation and Predisposition to Childhood Acute Lymphoblastic Leukemia." Cancer Cell **33**(5): 937-948 e938.

Conte, N., J. Mason, C. Halmagyi, S. B. Neuhauser, A. Mosaku, D. A. Begley, D. M. Krupke, H. Parkinson, T. Meehan and C. J. Bult (2018). "PDX Finder: A Portal for Patient-Derived tumor Xenograft Model Discovery." bioRxiv.

Corral, J., I. Lavenir, H. Impey, A. J. Warren, A. Forster, T. A. Larson, S. Bell, A. N. McKenzie, G. King and T. H. Rabbitts (1996). "An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes." Cell **85**(6): 853-861.

Daley, G. Q., R. A. Van Etten and D. Baltimore (1990). "Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome." Science **247**(4944): 824-830.

Dang, J., S. Nance, J. Ma, J. Cheng, M. P. Walsh, P. Vogel, J. Easton, G. Song, M. Rusch, A. L. Gedman, C. Koss, J. R. Downing and T. A. Gruber (2017). "AMKL chimeric transcription factors are potent inducers of leukemia." Leukemia **31**(10): 2228-2234.

Dawson, M. A., R. K. Prinjha, A. Dittmann, G. Giotopoulos, M. Bantscheff, W. I. Chan, S. C. Robson, C. W. Chung, C. Hopf, M. M. Savitski, C. Huthmacher, E. Gudgin, D. Lugo, S. Beinke, T. D. Chapman, E. J. Roberts, P. E. Soden, K. R. Auger, O. Mirguet, K. Doehner, R. Delwel, A. K. Burnett, P. Jeffrey, G. Drewes, K. Lee, B. J. Huntly and T. Kouzarides (2011). "Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia." Nature **478**(7370): 529-533.

de Guzman, C. G., A. J. Warren, Z. Zhang, L. Gartland, P. Erickson, H. Drabkin, S. W. Hiebert and C. A. Klug (2002). "Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation." Mol Cell Biol **22**(15): 5506-5517.

Deshpande, A. J., L. Chen, M. Fazio, A. U. Sinha, K. M. Bernt, D. Banka, S. Dias, J. Chang, E. J. Olhava, S. R. Daigle, V. M. Richon, R. M. Pollock and S. A. Armstrong (2013). "Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l." Blood **121**(13): 2533-2541.

Downing, J. R., R. K. Wilson, J. Zhang, E. R. Mardis, C. H. Pui, L. Ding, T. J. Ley and W. E. Evans (2012). "The Pediatric Cancer Genome Project." Nat Genet **44**(6): 619-622.

Drexler, H. G., H. Quentmeier and R. A. MacLeod (2004). "Malignant hematopoietic cell lines: in vitro models for the study of MLL gene alterations." Leukemia **18**(2): 227-232.

Duque-Afonso, J., J. Feng, F. Scherer, C. H. Lin, S. H. Wong, Z. Wang, M. Iwasaki and M. L. Cleary (2015). "Comparative genomics reveals multistep pathogenesis of E2A-PBX1 acute lymphoblastic leukemia." J Clin Invest **125**(9): 3667-3680.

Dykstra, B., D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S. J. Lee, R. Brinkman and C. Eaves (2007). "Long-term propagation of distinct hematopoietic differentiation programs in vivo." Cell Stem Cell **1**(2): 218-229.

Dykstra, B., J. Ramunas, D. Kent, L. McCaffrey, E. Szumsky, L. Kelly, K. Farn, A. Blaylock, C. Eaves and E. Jervis (2006). "High-resolution video monitoring of hematopoietic stem cells cultured in single-cell arrays identifies new features of self-renewal." Proc Natl Acad Sci U S A **103**(21): 8185-8190.

Estey, E. and H. Dohner (2006). "Acute myeloid leukaemia." Lancet **368**(9550): 1894-1907.

Forster, A., R. Pannell, L. F. Drynan, M. McCormack, E. C. Collins, A. Daser and T. H. Rabbitts (2003). "Engineering de novo reciprocal chromosomal translocations associated with MLL to replicate primary events of human cancer." Cancer Cell **3**(5): 449-458.

Garnett, M. J., E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P. Greninger, I. R. Thompson, X. Luo, J. Soares, Q. Liu, F. Iorio, D. Surdez, L. Chen, R. J. Milano, G. R. Bignell, A. T. Tam, H. Davies, J. A. Stevenson, S. Barthorpe, S. R. Lutz, F. Kogera, K. Lawrence, A. McLaren-Douglas, X. Mitropoulos, T. Mironenko, H. Thi, L. Richardson, W. Zhou, F. Jewitt, T. Zhang, P. O'Brien, J. L. Boisvert, S. Price, W. Hur, W. Yang, X. Deng, A. Butler, H. G. Choi, J. W. Chang, J. Baselga, I. Stamenkovic, J. A. Engelman, S. V. Sharma, O. Delattre, J. Saez-Rodriguez, N. S. Gray, J. Settleman, P. A. Futreal, D. A. Haber, M. R. Stratton, S. Ramaswamy, U. McDermott and C. H. Benes (2012). "Systematic identification of genomic markers of drug sensitivity in cancer cells." Nature **483**(7391): 570-575.

Genomes Project, C., G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Goyama, S., M. Wunderlich and J. C. Mulloy (2015). "Xenograft models for normal and malignant stem cells." Blood **125**(17): 2630-2640.

Grobner, S. N., B. C. Worst, J. Weischenfeldt, I. Buchhalter, K. Kleinheinz, V. A. Rudneva, P. D. Johann, G. P. Balasubramanian, M. Segura-Wang, S. Brabetz, S. Bender, B. Hutter, D. Sturm, E. Pfaff, D. Hubschmann, G. Zipprich, M. Heinold, J. Eils, C. Lawrenz, S. Erkek, S. Lambo, S. Waszak, C. Blattmann, A. Borkhardt, M. Kuhlen, A. Eggert, S. Fulda, M. Gessler, J. Wegert, R. Kappler, D. Baumhoer, S. Burdach, R. Kirschner-Schwabe, U. Kontny, A. E. Kulozik, D. Lohmann, S. Hettmer, C. Eckert, S. Bielack, M. Nathrath, C. Niemeyer, G. H. Richter, J. Schulte, R. Siebert, F. Westermann, J. J. Molenaar, G. Vassal, H. Witt, I. P.-S. Project, I. M.-S. Project, B. Burkhardt, C. P. Kratz, O. Witt, C. M. van Tilburg, C. M. Kramm, G. Fleischhack, U. Dirksen, S. Rutkowski, M. Fruhwald, K. von Hoff, S. Wolf, T. Klingebiel, E. Koscielniak, P. Landgraf, J. Koster, A. C. Resnick, J. Zhang, Y. Liu, X. Zhou, A. J. Waanders, D. A. Zwijnenburg, P. Raman, B. Brors, U. D. Weber, P. A. Northcott, K. W. Pajtler, M. Kool, R. M. Piro, J. O. Korbel, M. Schlesner, R. Eils, D. T. W. Jones, P. Lichter, L. Chavez, M. Zapatka and S. M. Pfister (2018). "The landscape of genomic alterations across childhood cancers." *Nature* **555**(7696): 321-327.

Gruber, T. A. and J. R. Downing (2015). "The biology of pediatric acute megakaryoblastic leukemia." *Blood* **126**(8): 943-949.

Hartwell, K. A., P. G. Miller, S. Mukherjee, A. R. Kahn, A. L. Stewart, D. J. Logan, J. M. Negri, M. Duvet, M. Jaras, R. Puram, V. Dancik, F. Al-Shahrour, T. Kindler, Z. Tothova, S. Chattopadhyay, T. Hasaka, R. Narayan, M. Dai, C. Huang, S. Shterental, L. P. Chu, J. E. Haydu, J. H. Shieh, D. P. Steensma, B. Munoz, J. A. Bittker, A. F. Shamji, P. A. Clemons, N. J. Tolliday, A. E. Carpenter, D. G. Gilliland, A. M. Stern, M. A. S. Moore, D. T. Scadden, S. L. Schreiber, B. L. Ebert and T. R. Golub (2013). "Niche-based screening identifies small-molecule inhibitors of leukemia stem cells." *Nat Chem Biol* **9**(12): 840-848.

Heshmati, Y., G. Turkoz, A. Harisankar, S. Kharazi, J. Bostrom, E. K. Dolatabadi, A. Krstic, D. Chang, R. Mansson, M. Altun, H. Qian and J. Walfridsson (2018). "The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia." *Haematologica* **103**(7): 1169-1181.

Hidalgo, M., F. Amant, A. V. Biankin, E. Budinska, A. T. Byrne, C. Caldas, R. B. Clarke, S. de Jong, J. Jonkers, G. M. Maeldansmo, S. Roman-Roman, J. Seoane, L. Trusolino and A. Villanueva (2014). "Patient-derived xenograft models: an emerging platform for translational cancer research." *Cancer Discov* **4**(9): 998-1013.

Hoadley, K. A., C. Yau, T. Hinoue, D. M. Wolf, A. J. Lazar, E. Drill, R. Shen, A. M. Taylor, A. D. Cherniack, V. Thorsson, R. Akbani, R. Bowlby, C. K. Wong, M. Wiznerowicz, F. Sanchez-Vega, A. G. Robertson, B. G. Schneider, M. S. Lawrence, H. Noushmehr, T. M. Malta, N. Cancer Genome Atlas, J. M. Stuart, C. C. Benz and P. W. Laird (2018). "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer." *Cell* **173**(2): 291-304 e296.

Horton, S. J., J. Jaques, C. Woolthuis, J. van Dijk, M. Mesuraca, G. Huls, G. Morrone, E. Vellenga and J. J. Schuringa (2013). "MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny." *Leukemia* **27**(5): 1116-1126.



Hsu, C. H., C. Nguyen, C. Yan, R. E. Ries, Q. R. Chen, Y. Hu, F. Ostronoff, D. L. Stirewalt, G. Komatsoulis, S. Levy, D. Meerzaman and S. Meshinchi (2015). "Transcriptome Profiling of Pediatric Core Binding Factor AML." *PLoS One* **10**(9): e0138782.

Imren, S., M. Heuser, M. Gasparetto, P. A. Beer, G. L. Norddahl, P. Xiang, L. Chen, T. Berg, G. W. Rhyasen, P. Rosten, G. Park, Y. Moon, A. P. Weng, C. J. Eaves and R. K. Humphries (2014). "Modeling de novo leukemogenesis from human cord blood with MN1 and NUP98HOXD13." *Blood* **124**(24): 3608-3612.

International Cancer Genome, C., T. J. Hudson, W. Anderson, A. Artez, A. D. Barker, C. Bell, R. R. Bernabe, M. K. Bhan, F. Calvo, I. Eerola, D. S. Gerhard, A. Guttmacher, M. Guyer, F. M. Hemsley, J. L. Jennings, D. Kerr, P. Klatt, P. Kolar, J. Kusada, D. P. Lane, F. Laplace, L. Youyong, G. Nettekoven, B. Ozenberger, J. Peterson, T. S. Rao, J. Remacle, A. J. Schafer, T. Shibata, M. R. Stratton, J. G. Vockley, K. Watanabe, H. Yang, M. M. Yuen, B. M. Knoppers, M. Bobrow, A. Cambon-Thomsen, L. G. Dressler, S. O. Dyke, Y. Joly, K. Kato, K. L. Kennedy, P. Nicolas, M. J. Parker, E. Rial-Sebbag, C. M. Romeo-Casabona, K. M. Shaw, S. Wallace, G. L. Wiesner, N. Zeps, P. Lichter, A. V. Biankin, C. Chabannon, L. Chin, B. Clement, E. de Alava, F. Degos, M. L. Ferguson, P. Geary, D. N. Hayes, T. J. Hudson, A. L. Johns, A. Kasprzyk, H. Nakagawa, R. Penny, M. A. Piris, R. Sarin, A. Scarpa, T. Shibata, M. van de Vijver, P. A. Futreal, H. Aburatani, M. Bayes, D. D. Botwell, P. J. Campbell, X. Estivill, D. S. Gerhard, S. M. Grimmond, I. Gut, M. Hirst, C. Lopez-Otin, P. Majumder, M. Marra, J. D. McPherson, H. Nakagawa, Z. Ning, X. S. Puente, Y. Ruan, T. Shibata, M. R. Stratton, H. G. Stunnenberg, H. Swerdlow, V. E. Velculescu, R. K. Wilson, H. H. Xue, L. Yang, P. T. Spellman, G. D. Bader, P. C. Boutros, P. J. Campbell, P. Flicek, G. Getz, R. Guigo, G. Guo, D. Haussler, S. Heath, T. J. Hubbard, T. Jiang, S. M. Jones, Q. Li, N. Lopez-Bigas, R. Luo, L. Muthuswamy, B. F. Ouellette, J. V. Pearson, X. S. Puente, V. Quesada, B. J. Raphael, C. Sander, T. Shibata, T. P. Speed, L. D. Stein, J. M. Stuart, J. W. Teague, Y. Totoki, T. Tsunoda, A. Valencia, D. A. Wheeler, H. Wu, S. Zhao, G. Zhou, L. D. Stein, R. Guigo, T. J. Hubbard, Y. Joly, S. M. Jones, A. Kasprzyk, M. Lathrop, N. Lopez-Bigas, B. F. Ouellette, P. T. Spellman, J. W. Teague, G. Thomas, A. Valencia, T. Yoshida, K. L. Kennedy, M. Axton, S. O. Dyke, P. A. Futreal, D. S. Gerhard, C. Gunter, M. Guyer, T. J. Hudson, J. D. McPherson, L. J. Miller, B. Ozenberger, K. M. Shaw, A. Kasprzyk, L. D. Stein, J. Zhang, S. A. Haider, J. Wang, C. K. Yung, A. Cros, Y. Liang, S. Gnaneshan, J. Guberman, J. Hsu, M. Bobrow, D. R. Chalmers, K. W. Hasel, Y. Joly, T. S. Kaan, K. L. Kennedy, B. M. Knoppers, W. W. Lowrance, T. Masui, P. Nicolas, E. Rial-Sebbag, L. L. Rodriguez, C. Vergely, T. Yoshida, S. M. Grimmond, A. V. Biankin, D. D. Bowtell, N. Cloonan, A. deFazio, J. R. Eshleman, D. Etemadmoghadam, B. B. Gardiner, J. G. Kench, A. Scarpa, R. L. Sutherland, M. A. Tempero, N. J. Waddell, P. J. Wilson, J. D. McPherson, S. Gallinger, M. S. Tsao, P. A. Shaw, G. M. Petersen, D. Mukhopadhyay, L. Chin, R. A. DePinho, S. Thayer, L. Muthuswamy, K. Shazand, T. Beck, M. Sam, L. Timms, V. Ballin, Y. Lu, J. Ji, X. Zhang, F. Chen, X. Hu, G. Zhou, Q. Yang, G. Tian, L. Zhang, X. Xing, X. Li, Z. Zhu, Y. Yu, J. Yu, H. Yang, M. Lathrop, J. Tost, P. Brennan, I. Holcatova, D. Zaridze, A. Brazma, L. Egevard, E. Prokhortchouk, R. E. Banks, M. Uhlen, A. Cambon-Thomsen, J. Viksna, F. Ponten, K. Skryabin, M. R. Stratton, P. A. Futreal, E. Birney, A. Borg, A. L. Borresen-Dale, C. Caldas, J. A. Foekens, S. Martin, J. S. Reis-Filho, A. L. Richardson, C. Sotiriou, H. G. Stunnenberg, G. Thoms, M. van de Vijver, L. van't Veer, F. Calvo, D. Birnbaum, H. Blanche, P. Boucher, S. Boyault, C. Chabannon, I. Gut, J. D. Masson-Jacquemier, M. Lathrop, I. Pauporte, X. Pivot, A. Vincent-Salomon, E. Tabone, C. Theillet, G. Thomas, J. Tost,

I. Treilleux, F. Calvo, P. Bioulac-Sage, B. Clement, T. Decaens, F. Degos, D. Franco, I. Gut, M. Gut, S. Heath, M. Lathrop, D. Samuel, G. Thomas, J. Zucman-Rossi, P. Lichter, R. Eils, B. Brors, J. O. Korbel, A. Korshunov, P. Landgraf, H. Lehrach, S. Pfister, B. Radlwimmer, G. Reifenger, M. D. Taylor, C. von Kalle, P. P. Majumder, R. Sarin, T. S. Rao, M. K. Bhan, A. Scarpa, P. Pederzoli, R. A. Lawlor, M. Delledonne, A. Bardelli, A. V. Biankin, S. M. Grimmond, T. Gress, D. Klimstra, G. Zamboni, T. Shibata, Y. Nakamura, H. Nakagawa, J. Kusada, T. Tsunoda, S. Miyano, H. Aburatani, K. Kato, A. Fujimoto, T. Yoshida, E. Campo, C. Lopez-Otin, X. Estivill, R. Guigo, S. de Sanjose, M. A. Piris, E. Montserrat, M. Gonzalez-Diaz, X. S. Puente, P. Jares, A. Valencia, H. Himmelbauer, V. Quesada, S. Bea, M. R. Stratton, P. A. Futreal, P. J. Campbell, A. Vincent-Salomon, A. L. Richardson, J. S. Reis-Filho, M. van de Vijver, G. Thomas, J. D. Masson-Jacquemier, S. Aparicio, A. Borg, A. L. Borresen-Dale, C. Caldas, J. A. Foekens, H. G. Stunnenberg, L. van't Veer, D. F. Easton, P. T. Spellman, S. Martin, A. D. Barker, L. Chin, F. S. Collins, C. C. Compton, M. L. Ferguson, D. S. Gerhard, G. Getz, C. Gunter, A. Guttmacher, M. Guyer, D. N. Hayes, E. S. Lander, B. Ozenberger, R. Penny, J. Peterson, C. Sander, K. M. Shaw, T. P. Speed, P. T. Spellman, J. G. Vockley, D. A. Wheeler, R. K. Wilson, T. J. Hudson, L. Chin, B. M. Knoppers, E. S. Lander, P. Lichter, L. D. Stein, M. R. Stratton, W. Anderson, A. D. Barker, C. Bell, M. Bobrow, W. Burke, F. S. Collins, C. C. Compton, R. A. DePinho, D. F. Easton, P. A. Futreal, D. S. Gerhard, A. R. Green, M. Guyer, S. R. Hamilton, T. J. Hubbard, O. P. Kallioniemi, K. L. Kennedy, T. J. Ley, E. T. Liu, Y. Lu, P. Majumder, M. Marra, B. Ozenberger, J. Peterson, A. J. Schafer, P. T. Spellman, H. G. Stunnenberg, B. J. Wainwright, R. K. Wilson and H. Yang (2010). "International network of cancer genome projects." *Nature* **464**(7291): 993-998.

Iorio, F., T. A. Knijnenburg, D. J. Vis, G. R. Bignell, M. P. Menden, M. Schubert, N. Aben, E. Goncalves, S. Barthorpe, H. Lightfoot, T. Cokelaer, P. Greninger, E. van Dyk, H. Chang, H. de Silva, H. Heyn, X. Deng, R. K. Egan, Q. Liu, T. Mironenko, X. Mitropoulos, L. Richardson, J. Wang, T. Zhang, S. Moran, S. Sayols, M. Soleimani, D. Tamborero, N. Lopez-Bigas, P. Ross-Macdonald, M. Esteller, N. S. Gray, D. A. Haber, M. R. Stratton, C. H. Benes, L. F. A. Wessels, J. Saez-Rodriguez, U. McDermott and M. J. Garnett (2016). "A Landscape of Pharmacogenomic Interactions in Cancer." *Cell* **166**(3): 740-754.

Jaffe, J. D., Y. Wang, H. M. Chan, J. Zhang, R. Huether, G. V. Kryukov, H. E. Bhang, J. E. Taylor, M. Hu, N. P. Englund, F. Yan, Z. Wang, E. Robert McDonald, 3rd, L. Wei, J. Ma, J. Easton, Z. Yu, R. deBeaumont, V. Gibaja, K. Venkatesan, R. Schlegel, W. R. Sellers, N. Keen, J. Liu, G. Caponigro, J. Barretina, V. G. Cooke, C. Mullighan, S. A. Carr, J. R. Downing, L. A. Garraway and F. Stegmeier (2013). "Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia." *Nat Genet* **45**(11): 1386-1391.

Juliusson, G., P. Antunovic, A. Derolf, S. Lehmann, L. Mollgard, D. Stockelberg, U. Tidefelt, A. Wahlin and M. Hoglund (2009). "Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry." *Blood* **113**(18): 4179-4187.

Kelliher, M. A., J. McLaughlin, O. N. Witte and N. Rosenberg (1990). "Induction of a chronic myelogenous leukemia-like syndrome in mice with v-abl and BCR/ABL." *Proc Natl Acad Sci U S A* **87**(17): 6649-6653.

Klijn, C., S. Durinck, E. W. Stawiski, P. M. Haverty, Z. Jiang, H. Liu, J. Degenhardt, O. Mayba, F. Gnad, J. Liu, G. Pau, J. Reeder, Y. Cao, K. Mukhyala, S. K. Selvaraj, M. Yu, G. J. Zynda, M. J. Brauer, T. D. Wu, R. C. Gentleman, G. Manning, R. L. Yauch, R. Bourgon, D. Stokoe, Z. Modrusan, R. M. Neve, F. J. de Sauvage, J. Settleman, S. Seshagiri and Z. Zhang (2015). "A comprehensive transcriptional portrait of human cancer cell lines." *Nat Biotechnol* **33**(3): 306-312.

Krivtsov, A. V., Z. Feng, M. E. Lemieux, J. Faber, S. Vempati, A. U. Sinha, X. Xia, J. Jesneck, A. P. Bracken, L. B. Silverman, J. L. Kutok, A. L. Kung and S. A. Armstrong (2008). "H3K79 methylation profiles define murine and human MLL-AF4 leukemias." *Cancer Cell* **14**(5): 355-368.

Krivtsov, A. V., X. Wang, N. R. Farnoud, M. Hadler, M. S. Martin, K. M. Knapp, E. M. Stein, D. Lipson, M. K. Nahas, P. J. Stephens, E. M. Sanford, K. Wang, G. M. Frampton, T. Brennan, G. A. Otto, R. Yelensky, V. Miller, R. L. Levine, A. A. Ferrando and S. A. Armstrong (2014). "Patient Derived Xenograft (PDX) Models Recapitulate the Genomic-Driver Composition of Acute Leukemia Samples." *Blood* **124**(21): 286.

Kroon, E., U. Thorsteinsdottir, N. Mayotte, T. Nakamura and G. Sauvageau (2001). "NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice." *EMBO J* **20**(3): 350-361.

Kumar, B., M. Garcia, L. Weng, X. Jung, J. L. Murakami, X. Hu, T. McDonald, A. Lin, A. R. Kumar, D. L. DiGiusto, A. S. Stein, V. A. Pullarkat, S. K. Hui, N. Carlesso, Y. H. Kuo, R. Bhatia, G. Marcucci and C. C. Chen (2018). "Acute myeloid leukemia transforms the bone marrow niche into a leukemia-permissive microenvironment through exosome secretion." *Leukemia* **32**(3): 575-587.

Lamble, A. J. and E. F. Lind (2018). "Targeting the Immune Microenvironment in Acute Myeloid Leukemia: A Focus on T Cell Immunity." *Front Oncol* **8**: 213.

Lane, S. W., D. T. Scadden and D. G. Gilliland (2009). "The leukemic stem cell niche: current concepts and therapeutic opportunities." *Blood* **114**(6): 1150-1157.

Li, H., B. G. Mar, H. Zhang, R. V. Puram, F. Vazquez, B. A. Weir, W. C. Hahn, B. Ebert and D. Pellman (2017). "The EMT regulator ZEB2 is a novel dependency of human and murine acute myeloid leukemia." *Blood* **129**(4): 497-508.

Lin, S., R. T. Luo, A. Ptasinska, J. Kerry, S. A. Assi, M. Wunderlich, T. Imamura, J. J. Kaberlein, A. Rayes, M. J. Althoff, J. Anastasi, M. M. O'Brien, A. R. Meetei, T. A. Milne, C. Bonifer, J. C. Mulloy and M. J. Thirman (2016). "Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia." *Cancer Cell* **30**(5): 737-749.

Liu, Y. F., B. Y. Wang, W. N. Zhang, J. Y. Huang, B. S. Li, M. Zhang, L. Jiang, J. F. Li, M. J. Wang, Y. J. Dai, Z. G. Zhang, Q. Wang, J. Kong, B. Chen, Y. M. Zhu, X. Q. Weng, Z. X. Shen, J. M. Li, J. Wang, X. J. Yan, Y. Li, Y. M. Liang, L. Liu, X. Q. Chen, W. G. Zhang, J. S. Yan, J. D. Hu, S. H. Shen, J. Chen, L. J. Gu, D. Pei, Y. Li, G. Wu, X. Zhou, R. B. Ren, C. Cheng, J. J. Yang, K. K. Wang, S. Y. Wang, J. Zhang, J. Q. Mi, C. H. Pui, J. Y. Tang, Z. Chen and S. J. Chen (2016). "Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia." *EBioMedicine* **8**: 173-183.

Lock, R. B., N. Liem, M. L. Farnsworth, C. G. Milross, C. Xue, M. Tajbakhsh, M. Haber, M. D. Norris, G. M. Marshall and A. M. Rice (2002). "The nonobese diabetic/severe combined immunodeficient (NOD/SCID) mouse model of childhood acute lymphoblastic leukemia reveals intrinsic differences in biologic characteristics at diagnosis and relapse." *Blood* **99**(11): 4100-4108.

Ma, X., Y. Liu, Y. Liu, L. B. Alexandrov, M. N. Edmonson, C. Gawad, X. Zhou, Y. Li, M. C. Rusch, J. Easton, R. Huether, V. Gonzalez-Pena, M. R. Wilkinson, L. C. Hermida, S. Davis, E. Sioson, S. Pounds, X. Cao, R. E. Ries, Z. Wang, X. Chen, L. Dong, S. J. Diskin, M. A. Smith, J. M. Guidry Auvil, P. S. Meltzer, C. C. Lau, E. J. Perlman, J. M. Maris, S. Meshinchi, S. P. Hunger, D. S. Gerhard and J. Zhang (2018). "Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours." *Nature* **555**(7696): 371-376.

Meacham, C. E., L. N. Lawton, Y. M. Soto-Feliciano, J. R. Pritchard, B. A. Joughin, T. Ehrenberger, N. Fenouille, J. Zuber, R. T. Williams, R. A. Young and M. T. Hemann (2015). "A genome-scale in vivo loss-of-function screen identifies Phf6 as a lineage-specific regulator of leukemia cell growth." *Genes Dev* **29**(5): 483-488.

Mercher, T., G. D. Raffel, S. A. Moore, M. G. Cornejo, D. Baudry-Bluteau, N. Cagnard, J. L. Jesneck, Y. Pikman, D. Cullen, I. R. Williams, K. Akashi, H. Shigematsu, J. P. Bourquin, M. Giovannini, W. Vainchenker, R. L. Levine, B. H. Lee, O. A. Bernard and D. G. Gilliland (2009). "The OTT-MAL fusion oncogene activates RBPJ-mediated transcription and induces acute megakaryoblastic leukemia in a knockin mouse model." *J Clin Invest* **119**(4): 852-864.

Mercier, F., J. Shi, D. B. Sykes, E. Miller, T. Oki, C.-H. Man, X.-J. Yan, Y. Kfoury, D. Lee, J. G. Doench, F. Michor and D. T. Scadden (2017). "A Genome-Wide, In Vivo, Dropout CRISPR Screen in Acute Myeloid Leukemia Identifies an Essential Role for Beta-Galactosylation in Leukemic Cell Homing." *Blood* **130**(Suppl 1): 2493.

Metzeler, K. H., T. Herold, M. Rothenberg-Thurley, S. Amler, M. C. Sauerland, D. Gorlich, S. Schneider, N. P. Konstandin, A. Dufour, K. Braundl, B. Ksienzyk, E. Zellmeier, L. Hartmann, P. A. Greif, M. Fiegl, M. Subklewe, S. K. Bohlander, U. Krug, A. Faldum, W. E. Berdel, B. Wormann, T. Buchner, W. Hiddemann, J. Braess, K. Spiekermann and A. S. Group (2016). "Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia." *Blood* **128**(5): 686-698.

Metzler, M., A. Forster, R. Pannell, M. J. Arends, A. Daser, M. N. Lobato and T. H. Rabbitts (2006). "A conditional model of MLL-AF4 B-cell tumorigenesis using invertor technology." *Oncogene* **25**(22): 3093-3103.

Meyer, C., T. Burmeister, D. Groger, G. Tsaur, L. Fechina, A. Renneville, R. Sutton, N. C. Venn, M. Emerenciano, M. S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal, V. H. J. van der Velden, E. Delabesse, T. S. Park, S. H. Oh, M. L. M. Silva, T. Lund-Aho, V. Juvonen, A. S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenkova, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. Zur Stadt, J. Lentjes, G. Gohring, A. Kustanovich, O. Aleinikova, B. W. Schafer, S. Kubetzko, H. O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J. M. Cayuela, E. Clappier, C. N. Alonso, C. M. Zwaan, M. M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Moricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O. A. Haas, R. Panzer-Grumayer,

L. Sedek, T. Szczepanski, A. Caye, L. Suarez, H. Cave and R. Marschalek (2018). "The MLL recombinome of acute leukemias in 2017." *Leukemia* **32**(2): 273-284.

Miller, P. G., F. Al-Shahrour, K. A. Hartwell, L. P. Chu, M. Jaras, R. V. Puram, A. Puissant, K. P. Callahan, J. Ashton, M. E. McConkey, L. P. Poveromo, G. S. Cowley, M. G. Kharas, M. Labelle, S. Shterental, J. Fujisaki, L. Silberstein, G. Alexe, M. A. Al-Hajj, C. A. Shelton, S. A. Armstrong, D. E. Root, D. T. Scadden, R. O. Hynes, S. Mukherjee, K. Stegmaier, C. T. Jordan and B. L. Ebert (2013). "In Vivo RNAi screening identifies a leukemia-specific dependence on integrin beta 3 signaling." *Cancer Cell* **24**(1): 45-58.

Montes, R., V. Ayllon, I. Gutierrez-Aranda, I. Prat, M. C. Hernandez-Lamas, L. Ponce, S. Bresolin, G. Te Kronnie, M. Greaves, C. Bueno and P. Menendez (2011). "Enforced expression of MLL-AF4 fusion in cord blood CD34+ cells enhances the hematopoietic repopulating cell function and clonogenic potential but is not sufficient to initiate leukemia." *Blood* **117**(18): 4746-4758.

Mori, H., S. M. Colman, Z. Xiao, A. M. Ford, L. E. Healy, C. Donaldson, J. M. Hows, C. Navarrete and M. Greaves (2002). "Chromosome translocations and covert leukemic clones are generated during normal fetal development." *Proc Natl Acad Sci U S A* **99**(12): 8242-8247.

Moriya, K., M. Suzuki, Y. Watanabe, T. Takahashi, Y. Aoki, T. Uchiyama, S. Kumaki, Y. Sasahara, M. Minegishi, S. Kure, S. Tsuchiya, K. Sugamura and N. Ishii (2012). "Development of a multi-step leukemogenesis model of MLL-rearranged leukemia using humanized mice." *PLoS One* **7**(6): e37892.

Moy, C., C. A. Oleykowski, R. Plant, J. Greshock, J. Jing, K. Bachman, M. A. Hardwicke, R. Wooster and Y. Degenhardt (2011). "High chromosome number in hematological cancer cell lines is a negative predictor of response to the inhibition of Aurora B and C by GSK1070916." *J Transl Med* **9**: 110.

Munker, R., M. L. Nordberg, D. Veillon, B. J. Williams, A. Roggero, W. Kern, F. Dicker and T. Haferlach (2009). "Characterization of a new myeloid leukemia cell line with normal cytogenetics (CG-SH)." *Leuk Res* **33**(10): 1405-1408.

Mussai, F. J., C. Yap, C. Mitchell and P. Kearns (2014). "Challenges of clinical trial design for targeted agents against pediatric leukemias." *Front Oncol* **4**: 374.

NCI, S.-. (2017). "SEER Cancer Statistics Review." from <http://seer.cancer.gov/>.

Okada, S., H. Nakauchi, K. Nagayoshi, S. Nishikawa, Y. Miura and T. Suda (1992). "In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells." *Blood* **80**(12): 3044-3050.

Pabst, C., J. Krosch, I. Fares, G. Boucher, R. Ruel, A. Marinier, S. Lemieux, J. Hebert and G. Sauvageau (2014). "Identification of small molecules that support human leukemia stem cell activity ex vivo." *Nat Methods* **11**(4): 436-442.

Papenhausen, P., C. A. Kelly, Z. Zhang, J. Tepperberg, R. D. Burnside and S. Schwartz (2019). "Multidisciplinary analysis of pediatric T-ALL: 9q34 gene fusions." *Cancer Genet* **231-232**: 1-13.

- Phekoo, K. J., M. A. Richards, H. Moller, S. A. Schey and C. South Thames Haematology Specialist (2006). "The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England." *Haematologica* **91**(10): 1400-1404.
- Phillips, C. L., R. Gerbing, T. Alonzo, J. P. Perentesis, I. T. Harley, S. Meshinchi, D. Bhatla, G. Radloff and S. M. Davies (2010). "MDM2 polymorphism increases susceptibility to childhood acute myeloid leukemia: a report from the Children's Oncology Group." *Pediatr Blood Cancer* **55**(2): 248-253.
- Pineault, N., C. Buske, M. Feuring-Buske, C. Abramovich, P. Rosten, D. E. Hogge, P. D. Aplan and R. K. Humphries (2003). "Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1." *Blood* **101**(11): 4529-4538.
- Pompili, L., M. Porru, C. Caruso, A. Biroccio and C. Leonetti (2016). "Patient-derived xenografts: a relevant preclinical model for drug development." *J Exp Clin Cancer Res* **35**(1): 189.
- Porter, C. C., J. Kim, S. Fosmire, C. M. Gearheart, A. van Linden, D. Baturin, V. Zaberezhnyy, P. R. Patel, D. Gao, A. C. Tan and J. DeGregori (2012). "Integrated genomic analyses identify WEE1 as a critical mediator of cell fate and a novel therapeutic target in acute myeloid leukemia." *Leukemia* **26**(6): 1266-1276.
- Qian, M., X. Cao, M. Devidas, W. Yang, C. Cheng, Y. Dai, A. Carroll, N. A. Heerema, H. Zhang, T. Moriyama, J. M. Gastier-Foster, H. Xu, E. Raetz, E. Larsen, N. Winick, W. P. Bowman, P. L. Martin, E. R. Mardis, R. Fulton, G. Zambetti, M. Borowitz, B. Wood, K. E. Nichols, W. L. Carroll, C. H. Pui, C. G. Mullighan, W. E. Evans, S. P. Hunger, M. V. Relling, M. L. Loh and J. J. Yang (2018). "TP53 Germline Variations Influence the Predisposition and Prognosis of B-Cell Acute Lymphoblastic Leukemia in Children." *J Clin Oncol* **36**(6): 591-599.
- Reshmi, S. C., R. C. Harvey, K. G. Roberts, E. Stonerock, A. Smith, H. Jenkins, I. M. Chen, M. Valentine, Y. Liu, Y. Li, Y. Shao, J. Easton, D. Payne-Turner, Z. Gu, T. H. Tran, J. V. Nguyen, M. Devidas, Y. Dai, N. A. Heerema, A. J. Carroll, 3rd, E. A. Raetz, M. J. Borowitz, B. L. Wood, A. L. Angiolillo, M. J. Burke, W. L. Salzer, P. A. Zweidler-McKay, K. R. Rabin, W. L. Carroll, J. Zhang, M. L. Loh, C. G. Mullighan, C. L. Willman, J. M. Gastier-Foster and S. P. Hunger (2017). "Targetable kinase gene fusions in high-risk B-ALL: a study from the Children's Oncology Group." *Blood* **129**(25): 3352-3361.
- Rizo, A., S. J. Horton, S. Olthof, B. Dontje, A. Ausema, R. van Os, V. van den Boom, E. Vellenga, G. de Haan and J. J. Schuringa (2010). "BMI1 collaborates with BCR-ABL in leukemic transformation of human CD34+ cells." *Blood* **116**(22): 4621-4630.
- Roberts, K. G., R. D. Morin, J. Zhang, M. Hirst, Y. Zhao, X. Su, S. C. Chen, D. Payne-Turner, M. L. Churchman, R. C. Harvey, X. Chen, C. Kasap, C. Yan, J. Becksfort, R. P. Finney, D. T. Teachey, S. L. Maude, K. Tse, R. Moore, S. Jones, K. Mungall, I. Birol, M. N. Edmonson, Y. Hu, K. E. Buetow, I. M. Chen, W. L. Carroll, L. Wei, J. Ma, M. Kleppe, R. L. Levine, G. Garcia-Manero, E. Larsen, N. P. Shah, M. Devidas, G. Reaman, M. Smith, S. W. Paugh, W. E. Evans, S. A. Grupp, S. Jeha, C. H. Pui, D. S. Gerhard, J. R. Downing, C. L. Willman, M. Loh, S. P. Hunger, M. A. Marra and C. G. Mullighan (2012). "Genetic alterations activating kinase and

cytokine receptor signaling in high-risk acute lymphoblastic leukemia." *Cancer Cell* **22**(2): 153-166.

Rodriguez-Hernandez, G., J. Hauer, A. Martin-Lorenzo, D. Schafer, C. Bartenhagen, I. Garcia-Ramirez, F. Auer, I. Gonzalez-Herrero, L. Ruiz-Roca, M. Gombert, V. Okpanyi, U. Fischer, C. Chen, M. Dugas, S. Bhatia, R. M. Linka, M. Garcia-Suquia, M. V. Rascon-Trincado, A. Garcia-Sanchez, O. Blanco, M. B. Garcia-Cenador, F. J. Garcia-Criado, C. Cobaleda, D. Alonso-Lopez, J. De Las Rivas, M. Muschen, C. Vicente-Duenas, I. Sanchez-Garcia and A. Borkhardt (2017). "Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 Demethylases." *Cancer Res* **77**(16): 4365-4377.

Schessl, C., V. P. Rawat, M. Cusan, A. Deshpande, T. M. Kohl, P. M. Rosten, K. Spiekermann, R. K. Humphries, S. Schnittger, W. Kern, W. Hiddemann, L. Quintanilla-Martinez, S. K. Bohlander, M. Feuring-Buske and C. Buske (2005). "The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice." *J Clin Invest* **115**(8): 2159-2168.

Schneidawind, C., J. Jeong, D. Schneidawind, I. S. Kim, J. Duque-Afonso, S. H. K. Wong, M. Iwasaki, E. H. Breese, J. L. Zehnder, M. Porteus and M. L. Cleary (2018). "MLL leukemia induction by t(9;11) chromosomal translocation in human hematopoietic stem cells using genome editing." *Blood Adv* **2**(8): 832-845.

Schwartz, J. R., J. Ma, T. Lamprecht, M. Walsh, S. Wang, V. Bryant, G. Song, G. Wu, J. Easton, C. Kesserwan, K. E. Nichols, C. G. Mullighan, R. C. Ribeiro and J. M. Kluco (2017). "The genomic landscape of pediatric myelodysplastic syndromes." *Nat Commun* **8**(1): 1557.

Shen, H., J. Li, J. Zhang, C. Xu, Y. Jiang, Z. Wu, F. Zhao, L. Liao, J. Chen, Y. Lin, Q. Tian, C. J. Papasian and H. W. Deng (2013). "Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians." *PLoS One* **8**(4): e59494.

Shukla, N., C. Wetmore, M. M. Brien, L. B. Silverman, P. Brown, T. M. Cooper, B. Thomson, S. J. Blakemore, S. Daigle, B. Suttle, N. J. Waters, A. V. Krivstov, S. A. Armstrong, P. T. Ho and L. Gore (2016). "Final Report of Phase 1 Study of the DOT1L Inhibitor, Pinometostat (EPZ-5676), in Children with Relapsed or Refractory MLL-r Acute Leukemia." *Blood* **128**(22): 2780.

Siegel, D. A., J. King, E. Tai, N. Buchanan, U. A. Ajani and J. Li (2014). "Cancer incidence rates and trends among children and adolescents in the United States, 2001-2009." *Pediatrics* **134**(4): e945-955.

Sroczyńska, P., V. A. Cruickshank, J. P. Bukowski, S. Miyagi, F. O. Bagger, J. Walfridsson, M. B. Schuster, B. Porse and K. Helin (2014). "shRNA screening identifies JMJD1C as being required for leukemia maintenance." *Blood* **123**(12): 1870-1882.

Stein, E. M., G. Garcia-Manero, D. A. Rizzieri, R. Tibes, J. G. Berdeja, M. R. Savona, M. Jongen-Lavrenic, J. K. Altman, B. Thomson, S. J. Blakemore, S. R. Daigle, N. J. Waters, A. B. Suttle, A. Clawson, R. Pollock, A. Krivtsov, S. A. Armstrong, J. DiMartino, E. Hedrick, B. Lowenberg and M. S. Tallman (2018). "The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia." *Blood* **131**(24): 2661-2669.

Steliarova-Foucher, E., M. Colombet, L. A. G. Ries, F. Moreno, A. Dolya, F. Bray, P. Hesselting, H. Y. Shin, C. A. Stiller and I.-. contributors (2017). "International incidence of childhood cancer, 2001-10: a population-based registry study." *Lancet Oncol* **18**(6): 719-731.

Stratton, M. R., P. J. Campbell and P. A. Futreal (2009). "The cancer genome." *Nature* **458**(7239): 719-724.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Fritz, M. K. Konkel, A. Malhotra, A. M. Stutz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E. W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalín, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, C. Genomes Project, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler and J. O. Korbel (2015). "An integrated map of structural variation in 2,504 human genomes." *Nature* **526**(7571): 75-81.

Townsend, E. C., M. A. Murakami, A. Christodoulou, A. L. Christie, J. Koster, T. A. DeSouza, E. A. Morgan, S. P. Kallgren, H. Liu, S. C. Wu, O. Plana, J. Montero, K. E. Stevenson, P. Rao, R. Vadhi, M. Andreeff, P. Armand, K. K. Ballen, P. Barzaghi-Rinaudo, S. Cahill, R. A. Clark, V. G. Cooke, M. S. Davids, D. J. DeAngelo, D. M. Dorfman, H. Eaton, B. L. Ebert, J. Etchin, B. Firestone, D. C. Fisher, A. S. Freedman, I. A. Galinsky, H. Gao, J. S. Garcia, F. Garnache-Ottou, T. A. Graubert, A. Gutierrez, E. Halilovic, M. H. Harris, Z. T. Herbert, S. M. Horwitz, G. Inghirami, A. M. Intlekofer, M. Ito, S. Izraeli, E. D. Jacobsen, C. A. Jacobson, S. Jeay, I. Jeremias, M. A. Kelliher, R. Koch, M. Konopleva, N. Kopp, S. M. Kornblau, A. L. Kung, T. S. Kupper, N. R. LeBoeuf, A. S. LaCasce, E. Lees, L. S. Li, A. T. Look, M. Murakami, M. Muschen, D. Neuberg, S. Y. Ng, O. O. Odejide, S. H. Orkin, R. R. Paquette, A. E. Place, J. E. Roderick, J. A. Ryan, S. E. Sallan, B. Shoji, L. B. Silverman, R. J. Soiffer, D. P. Steensma, K. Stegmaier, R. M. Stone, J. Tamburini, A. R. Thorner, P. van Hummelen, M. Wadleigh, M. Wiesmann, A. P. Weng, J. U. Wuerthner, D. A. Williams, B. M. Wollison, A. A. Lane, A. Letai, M. M. Bertagnolli, J. Ritz, M. Brown, H. Long, J. C. Aster, M. A. Shipp, J. D. Griffin and D. M. Weinstock (2016). "The Public Repository of Xenografts Enables Discovery and Randomized Phase II-like Trials in Mice." *Cancer Cell* **29**(4): 574-586.

Tzelepis, K., H. Koike-Yusa, E. De Braekeleer, Y. Li, E. Metzakopian, O. M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, M. Gozdecka, S. Ohnishi, J. Cooper, M. Patel, T. McKerrell, B. Chen, A. F. Domingues, P. Gallipoli, S. Teichmann, H. Ponstingl, U. McDermott, J. Saez-Rodriguez, B. J. P. Huntly, F. Iorio, C. Pina, G. S. Vassiliou and K. Yusa (2016). "A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia." *Cell Rep* **17**(4): 1193-1205.

Vick, B., M. Rothenberg, N. Sandhofer, M. Carlet, C. Finkenzeller, C. Krupka, M. Grunert, A. Trumpp, S. Corbacioglu, M. Ebinger, M. C. Andre, W. Hiddemann, S. Schneider, M. Subklewe, K. H. Metzeler, K. Spiekermann and I. Jeremias (2015). "An advanced preclinical mouse model



for acute myeloid leukemia using patients' cells of various genetic subgroups and in vivo bioluminescence imaging." *PLoS One* **10**(3): e0120925.

Wang, K., M. Sanchez-Martin, X. Wang, K. M. Knapp, R. Koche, L. Vu, M. K. Nahas, J. He, M. Hadler, E. M. Stein, M. S. Tallman, A. L. Donahue, G. M. Frampton, D. Lipson, S. Roels, P. J. Stephens, E. M. Sanford, T. Brennan, G. A. Otto, R. Yelensky, V. A. Miller, M. G. Kharas, R. L. Levine, A. Ferrando, S. A. Armstrong and A. V. Krivtsov (2017). "Patient-derived xenotransplants can recapitulate the genetic driver landscape of acute leukemias." *Leukemia* **31**(1): 151-158.

Wang, T., H. Yu, N. W. Hughes, B. Liu, A. Kendirli, K. Klein, W. W. Chen, E. S. Lander and D. M. Sabatini (2017). "Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras." *Cell* **168**(5): 890-903 e815.

Wei, J., M. Wunderlich, C. Fox, S. Alvarez, J. C. Cigudosa, J. S. Wilhelm, Y. Zheng, J. A. Cancelas, Y. Gu, M. Jansen, J. F. Dimartino and J. C. Mulloy (2008). "Microenvironment determines lineage fate in a human model of MLL-AF9 leukemia." *Cancer Cell* **13**(6): 483-495.

Wei, X., M. N. Calvo-Vidal, S. Chen, G. Wu, M. V. Revuelta, J. Sun, J. Zhang, M. F. Walsh, K. E. Nichols, V. Joseph, C. Snyder, C. M. Vachon, J. D. McKay, S. P. Wang, D. S. Jayabalan, L. M. Jacobs, D. Becirovic, R. G. Waller, M. Artomov, A. Viale, J. Patel, J. Phillip, S. Chen-Kiang, K. Curtin, M. Salama, D. Atanackovic, R. Niesvizky, O. Landgren, S. L. Slager, L. A. Godley, J. Churpek, J. E. Garber, K. C. Anderson, M. J. Daly, R. G. Roeder, C. Dumontet, H. T. Lynch, C. G. Mullighan, N. J. Camp, K. Offit, R. J. Klein, H. Yu, L. Cerchietti and S. M. Lipkin (2018). "Germline Lysine-Specific Demethylase 1 (LSD1/KDM1A) Mutations Confer Susceptibility to Multiple Myeloma." *Cancer Res* **78**(10): 2747-2759.

Welch, J. S., T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, T. L. Lamprecht, F. Liu, J. Xia, C. Kandoth, R. S. Fulton, M. D. McLellan, D. J. Dooling, J. W. Wallis, K. Chen, C. C. Harris, H. K. Schmidt, J. M. Kalicki-Veizer, C. Lu, Q. Zhang, L. Lin, M. D. O'Laughlin, J. F. McMichael, K. D. Delehaunty, L. A. Fulton, V. J. Magrini, S. D. McGrath, R. T. Demeter, T. L. Vickery, J. Hundal, L. L. Cook, G. W. Swift, J. P. Reed, P. A. Alldredge, T. N. Wylie, J. R. Walker, M. A. Watson, S. E. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, J. E. Payton, J. D. Baty, S. Kulkarni, J. M. Klco, M. H. Tomasson, P. Westervelt, M. J. Walter, T. A. Graubert, J. F. DiPersio, L. Ding, E. R. Mardis and R. K. Wilson (2012). "The origin and evolution of mutations in acute myeloid leukemia." *Cell* **150**(2): 264-278.

Wermke, M., A. Camgoz, M. Paszkowski-Rogacz, S. Thieme, M. von Bonin, A. Dahl, U. Platzbecker, M. Theis, G. Ehninger, S. Brenner, M. Bornhauser and F. Buchholz (2015). "RNAi profiling of primary human AML cells identifies ROCK1 as a therapeutic target and nominates fasudil as an antileukemic drug." *Blood* **125**(24): 3760-3768.

Weroha, S. J., M. A. Becker, S. Enderica-Gonzalez, S. C. Harrington, A. L. Oberg, M. J. Maurer, S. E. Perkins, M. AlHilli, K. A. Butler, S. McKinstry, S. Fink, R. B. Jenkins, X. Hou, K. R. Kalli, K. M. Goodman, J. N. Sarkaria, B. Y. Karlan, A. Kumar, S. H. Kaufmann, L. C. Hartmann and P. Haluska (2014). "Tumorgrafts as in vivo surrogates for women with ovarian cancer." *Clin Cancer Res* **20**(5): 1288-1297.

Winters, A. C. and K. M. Bernt (2017). "MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches." *Front Pediatr* **5**: 4.

Wunderlich, M., F. S. Chou, K. A. Link, B. Mizukawa, R. L. Perry, M. Carroll and J. C. Mulloy (2010). "AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3." *Leukemia* **24**(10): 1785-1788.

Xujun, W., N. R. Farnoud, M. Hadler, K. M. Knapp, E. M. Stein, M. S. Tallman, A. L. Donahue, D. Lipson, M. K. Nahas, P. J. Stephens, E. M. Sanford, K. Wang, G. M. Frampton, T. Brennan, G. A. Otto, R. Yelensky, V. A. Miller, R. L. Levine and S. A. Armstrong (2013). "Patient Derived Xenograft (PDX) Models Faithfully Recapitulate The Genetic Composition Of Primary AML." *Blood* **122**(21): 1328.

Yamauchi, T., T. Masuda, M. C. Canver, M. Seiler, Y. Semba, M. Shboul, M. Al-Raqad, M. Maeda, V. A. C. Schoonenberg, M. A. Cole, C. Macias-Trevino, Y. Ishikawa, Q. Yao, M. Nakano, F. Arai, S. H. Orkin, B. Reversade, S. Buonamici, L. Pinello, K. Akashi, D. E. Bauer and T. Maeda (2018). "Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-Specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS." *Cancer Cell* **33**(3): 386-400 e385.

Yang, J., A. Li, Y. Li, X. Guo and M. Wang (2018). "A novel approach for drug response prediction in cancer cell lines via network representation learning." *Bioinformatics*.

Zelent, A., M. Greaves and T. Enver (2004). "Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia." *Oncogene* **23**(24): 4275-4283.

Zhang, F., M. Wang, J. Xi, J. Yang and A. Li (2018). "A novel heterogeneous network-based method for drug response prediction in cancer cell lines." *Sci Rep* **8**(1): 3355.

Zhang, J., M. F. Walsh, G. Wu, M. N. Edmonson, T. A. Gruber, J. Easton, D. Hedges, X. Ma, X. Zhou, D. A. Yergeau, M. R. Wilkinson, B. Vadodaria, X. Chen, R. B. McGee, S. Hines-Dowell, R. Nuccio, E. Quinn, S. A. Shurtleff, M. Rusch, A. Patel, J. B. Beckfort, S. Wang, M. S. Weaver, L. Ding, E. R. Mardis, R. K. Wilson, A. Gajjar, D. W. Ellison, A. S. Pappo, C. H. Pui, K. E. Nichols and J. R. Downing (2015). "Germline Mutations in Predisposition Genes in Pediatric Cancer." *N Engl J Med* **373**(24): 2336-2346.

Zuber, J., J. Shi, E. Wang, A. R. Rappaport, H. Herrmann, E. A. Sison, D. Magoon, J. Qi, K. Blatt, M. Wunderlich, M. J. Taylor, C. Johns, A. Chicas, J. C. Mulloy, S. C. Kogan, P. Brown, P. Valent, J. E. Bradner, S. W. Lowe and C. R. Vakoc (2011). "RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia." *Nature* **478**(7370): 524-528.

# **CHAPITRE 5**

## **DISCUSSION, CONCLUSION ET PERSPECTIVES**

## 5 Discussion, conclusion et perspectives

Dans cette thèse, nous avons acquis de nouveaux éléments pour comprendre la mécanistique qui sous-tend la leucémogénèse médiée par des translocations chromosomiques de type *KMT2A*. Dans le chapitre 2, nous avons pu démontrer que la fusion de gènes induit des changements minimes dans les cellules leucémiques au niveau transcriptomique et épigénétique, notamment via le remodelage de la chromatine. Le chapitre 3 a mis en lumière les différences entre les modèles leucémiques porteurs des deux fusions de gènes *KMT2A-MLLT3* et *KMT2A-MLLT4*, et nous a permis d'émettre des hypothèses quant aux causes des écarts dans les probabilités de survie en clinique. De l'ensemble de cette thèse, nous pouvons en ressortir une série de points importants.

### 5.1 Un modèle avec des avantages et des inconvénients

Le premier point concerne le modèle en lui-même. Il se base sur l'intégration aléatoire d'une fusion oncogénique dans le génome de cellules souches CD34<sup>+</sup> et présente de nombreux avantages. Produire facilement des leucémies *humaines* permet de contrecarrer la rareté des échantillons en clinique. L'intégration de la fusion dans le génome de cellules provenant d'un donneur unique permet de s'abroger de l'hétérogénéité génétique retrouvée chez les patients. Par ailleurs, le modèle permet d'analyser l'impact *direct* d'une translocation chromosomique, aspect qu'il est impossible d'obtenir en clinique en raison de la présence de mutations secondaires. Malgré tout, les pistes d'amélioration sont nombreuses.

#### 5.1.1 Contrôler le site d'intégration de la fusion oncogénique

La première viserait à contrôler davantage les sites d'intégration de la fusion dans le génome des cellules, ce qui est impossible de faire dans le modèle actuel. Les données produites dans le chapitre 3 montrent une certaine hétérogénéité génétique entre les différents réplicats biologiques. Certes, la corrélation entre les échantillons du modèle est plus forte que celle observée entre deux échantillons de patients mais elle n'est pas parfaite et pourrait être améliorée. De nouvelles approches expérimentales utilisant des systèmes CRISPR-Cas9 permettraient de maîtriser le lieu d'intégration de la fusion, son niveau d'expression et réussir

ainsi à produire des échantillons leucémiques davantage reproductibles (Jeong, Jager et al. 2019, Sarrou, Richmond et al. 2020).

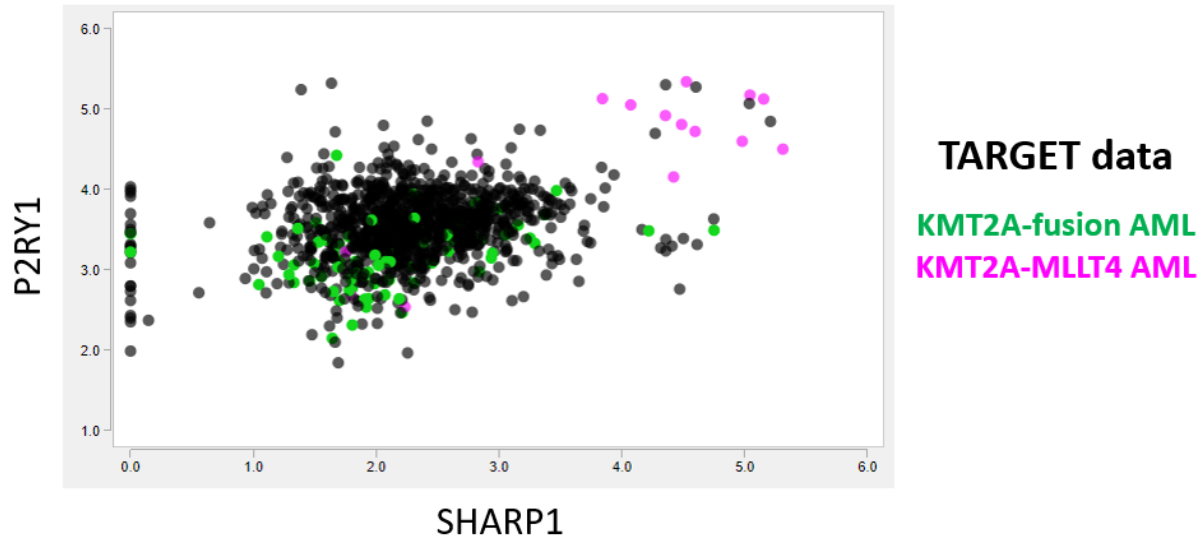
### 5.1.2 Se rapprocher encore davantage de la réalité clinique

Connaître les impacts directs de la présence d'une translocation chromosomique peut amener des pistes thérapeutiques intéressantes. Les données démontrent que le modèle est capable de récapituler ce qui est observé en clinique (chapitre 3).

Pourtant, certaines différences persistent et laissent à penser que le modèle ne suffit pas et reste éloigné de la réalité des patients pédiatriques chez qui il est rare d'observer uniquement une translocation chromosomique. La présence additionnelle de mutations des gènes *KRAS*, *NRAS* ou *KIT* (Bolouri, Farrar et al. 2018) complique l'analyse. C'est notamment la raison pour laquelle certains gènes ne sont pas exprimés dans le modèle mais fortement exprimés en clinique dans les lignées cellulaires ou dans les données de patients à notre disposition, et inversement. C'est le cas des gènes *SHARPI* et *P2RY1*, caractérisés comme potentiels biomarqueurs (Numata, Kwok et al. 2018) et fortement exprimés dans les leucémies pédiatriques spécifiquement médiées par la fusion *KM4* (**Figure 5.1**), mais non exprimés dans notre système modèle. Ce constat laisse à penser que *SHARPI* et *P2RY1* pourraient être impliqués davantage dans l'agressivité du cancer mais moins dans l'initiation et l'établissement de la leucémie causée par cette fusion en particulier.

Ces observations n'invalident pas la pertinence du modèle utilisé pour étudier spécifiquement les conséquences de la présence unique d'une translocation chromosomique. Il souligne l'importance de l'adapter pour se rapprocher davantage des profils mutationnels en clinique. Il pourrait être intéressant d'induire l'expression d'une fusion oncogénique et d'ajouter certaines mutations ponctuelles retrouvées particulièrement chez les patients pédiatriques. À notre connaissance, il n'existe actuellement aucun modèle de leucémies humaines de ce type. En établissant ce type de modèle, il serait alors possible : i) d'établir les profils transcriptomiques et épigénétiques dans d'autres contextes mutationnels et ainsi cartographier quelles combinaisons mutationnelles sont responsables de l'expression de certains biomarqueurs ; ii) de mesurer la sensibilité de potentiels composés thérapeutiques en fonction de la présence de

mutations secondaires en plus d'une fusion de gènes ; iii) de préciser encore davantage en quoi certains profils mutationnels prédisent les probabilités de survie chez les patients pédiatriques.



**Figure 5.1 – Expression des gènes *P2RY1* et *SHARP1* dans la cohorte de patients pédiatriques atteints d'une LMA (données TARGET).**

Chaque point correspond à un patient. Les couleurs vertes et roses correspondent respectivement à des données de patients avec une translocation de type *KMT2A* et avec une translocation *KMT2A-MLLT4*.

## 5.2 Une mécanique biologique qui soulève encore des questions

Le deuxième aspect qui ressort de notre étude relève du fonctionnement et des mécanismes oncogéniques des protéines de fusion.

Contrairement à ce que nous nous attendions à observer, les translocations chromosomiques profiteraient des programmes transcriptionnels établis dans les cellules sanguines considérées comme normales pour induire des changements minimes dans la structure et l'accessibilité de la chromatine ainsi que dans l'expression de certains gènes (chapitre 2). Cela rejoint une étude récente qui démontre que la fusion de gènes *KMT2A-MLLT3* préserve l'expression génique des cellules dans lesquelles elle est exprimée (Chen, Burkhardt et al. 2019). Par ailleurs, les protéines partenaires *MLLT3* et *MLLT4* sont respectivement présentes dans le noyau ou le

cytoplasme. Pourtant, les deux protéines de fusion fonctionnent sensiblement de la même manière dans le noyau en recrutant les complexes AEP et DOT1L pour induire l'expression de gènes cibles (chapitre 1). Ce point commun semblerait expliquer pourquoi il y a très peu de différences transcriptomiques et épigénétiques entre ces deux types de leucémies (chapitre 3) mais ne permettrait pas de justifier la variation dans la probabilité de survie des patients. Pris ensemble, toutes ces données laissent à penser qu'il est davantage probable que la différence dans le pronostic clinique s'expliquerait par un autre niveau de régulation, notamment protéique. C'est d'ailleurs ce que les résultats du chapitre 3 suggèrent, avec une modification dans les signatures de gènes codant pour les sous-unités ribosomales et intervenant dans les mécanismes de traduction.

Par ailleurs, la perte de la portion C-terminale de la protéine KMT2A entraîne la disparition du domaine TAD sur lequel peuvent se fixer des histones acétyltransférases (chapitre 1). Il pourrait être intéressant de cartographier l'enrichissement d'autres marques d'histones dans les cellules leucémiques porteuses de translocations de type *KMT2A*, notamment des marques activatrices au niveau des régions *enhancers* (H3K27ac), promotrices (H3K9ac) ou encore H4K16ac. Les différences seront probablement minimes mais peuvent apporter de nouvelles informations pour rechercher de potentiels biomarqueurs.

Une donnée manquante à notre analyse est la détermination des sites de fixation des protéines de fusion pour déterminer les gènes sur lesquels elles agissent directement. La question des cibles directes ou indirectes des fusions de type KMT2A serait donc très pertinente mais se heurte à un aspect technique. Il reste difficile de réaliser des approches par ChIP-seq en raison de la grande taille des protéines de fusion. À notre connaissance, il n'existe que des études indirectes qui ont tenté de cartographier les cibles directes de KMT2A-MLLT3. Certaines consistaient à réaliser une expérience de ChIP-seq sur les protéines nucléaires KMT2A et MLLT3 séparément et de considérer les cibles communes comme étant des cibles de la protéine de fusion (Prange, Mandoli et al. 2017). Bien que ces informations sur les cibles directes de la fusion donnent un aperçu au niveau de la mécanistique, elles sont incomplètes et ne seraient pas informatives dans le cas d'une protéine cytoplasmique comme MLLT4.

Il serait donc plus pertinent de modifier le plasmide codant pour la fusion dans notre modèle et d'y ajouter une étiquette identifiable par un anticorps. En revanche, la taille importante de la

protéine de fusion, son repliement et sa conformation tridimensionnelle risquent d'entraver l'accessibilité de l'anticorps. Par ailleurs, ajouter une étiquette peut également impacter artificiellement l'activité de la protéine de fusion.

Enfin, alors que la fusion *KM3* est largement étudiée dans la littérature, les mécanismes décrivant le fonctionnement de la protéine *KM4* restent encore mal compris. Pourtant, les quelques études publiées démontrent un rôle essentiel de la voie de signalisation RAS notamment via les domaines RA de la protéine *MLLT4* (Kuriyama, Harada et al. 1996). Elle agit comme régulateur du niveau de RAS-GTP dans les cellules (Manara, Baron et al. 2014). Une étude récente a montré que des mutations touchant la voie de signalisation RAS (*KRAS* et *NRAS*) concernent davantage les patients avec une translocation de type *KM4* comparativement aux patients porteurs d'autres types de translocations (comme *KM3*) (Bill, Mrozek et al. 2020). En clinique, le fait que le gène *RAS* soit muté ou non a un impact sur l'efficacité et la réponse aux molécules thérapeutiques (Lavalley, Baccelli et al. 2015). RAS semble jouer un rôle central dans la leucémogénèse médiée par la fusion de type *KMT2A* et peut-être plus spécifiquement dans celle causée par la fusion *KM4*. Pour vérifier cette hypothèse, il pourrait être intéressant d'exploiter les avantages de notre modèle d'étude. En combinant la présence d'une translocation chromosomique avec une mutation touchant la voie de signalisation RAS, il serait possible de mesurer si la leucémie est plus ou moins agressive via une analyse de la survie des souris. Par ailleurs, la détection de mutations de *RAS* chez les patients pourrait permettre de mieux prédire le pronostic en clinique et servir éventuellement de caractéristique supplémentaire dans la classification ENL décrite au chapitre 1.

### **5.3 Rechercher des biomarqueurs pour une médecine plus personnalisée**

Le troisième aspect issu de nos travaux repose sur l'identification de potentiels biomarqueurs, certains d'entre eux ayant démontré une spécificité et un rôle essentiel dans la leucémogénèse. Le modèle présentait l'avantage principal de mesurer l'impact direct de la présence d'une translocation chromosomique dans les cellules. Il s'est avéré plus difficile que prévu de trouver des biomarqueurs différentiellement exprimés et régulés de manière épigénétique.



Alors que des expériences fonctionnelles sont à effectuer pour valider la fonction des gènes *LPL* et *PROM1* dans les leucémies causées par une fusion *KM4* (chapitre 3), *ADCY9* et *CCR1* semblent être nouvellement identifiés dans les LMA médiées par la fusion *KM3* (chapitre 2). Il pourrait être intéressant d'élargir l'étude en déterminant les gènes essentiels à la viabilité des cellules leucémiques porteuses d'une translocation chromosomique de type *KMT2A* grâce à un criblage génomique par CRISPR-Cas9. Cette approche requiert une quantité importante de cellules pour réaliser l'expérience, un nombre difficilement atteignable avec notre modèle en raison du faible taux d'infection des cellules primaires par les rétrovirus ainsi que le temps limité de culture *in vitro*. En revanche, le criblage pourrait être établi dans des lignées cellulaires leucémiques porteuses (*Thp1*) ou non (*Kg1a*) de la fusion *KM3* et être comparé et validé avec des données existantes (notamment via Broad Depmap (Behan, Iorio et al. 2019)). Les gènes candidats peuvent ensuite être validés dans le modèle, à condition de vérifier leur profil d'expression à travers les différentes étapes de la leucémogénèse. Comme décrit dans le chapitre 4, chaque modèle a ses avantages et ses inconvénients : les lignées cellulaires restent très éloignées du modèle et de la clinique.

L'objectif de caractériser des biomarqueurs plus ou moins spécifiques des différentes fusions de gènes est *in fine* de : i) ajuster au mieux le pronostic clinique des patients en fonction de l'expression de ces derniers ; ii) pouvoir proposer des approches thérapeutiques plus efficaces pour se rapprocher d'une médecine plus personnalisée. Il reste aujourd'hui difficile de traiter les LMA causées par des réarrangements de type *KMT2A* (Steinhilber and Marschalek 2018). Des inhibiteurs ciblant les protéines interagissant avec la fusion oncogénique ont été développés et les premières études cliniques ont été lancées. Le manque de spécificité pour atteindre uniquement la protéine de fusion sans toucher la protéine sauvage *KMT2A* complique la découverte de traitements efficaces tout en limitant les potentiels effets secondaires. Il y a donc un besoin important de développer des traitements et des stratégies pour soit : i) inhiber spécifiquement la protéine de fusion, ou du moins réduire les effets secondaires sur le fonctionnement des cellules non-leucémiques ; ii) cibler les mécanismes qui résultent de la présence de la protéine de fusion ; iii) personnaliser le traitement selon le contexte mutationnel du patient.

## 5.4 Conclusion générale

Pour conclure, nos études ont permis de contribuer à la compréhension du développement leucémique causé directement par une translocation chromosomique de type *KMT2A*. Nos résultats démontrent que la transformation leucémique implique des modifications épigénétiques minimales et requiert également la coopération des réseaux transcriptionnels utilisés dans les cellules sanguines normales.

En exploitant les nombreux avantages d'un modèle de leucémies humaines, nous avons eu accès aux mécanismes à chacun des stades du développement du cancer. Le modèle nous a permis de définir des gènes régulés de manière épigénétique et différenciellement exprimés. Certains gènes pourraient jouer un rôle dans le processus de leucémogénèse médié par des fusions faisant intervenir le gène *KMT2A*. D'autres pourraient servir davantage à préciser le pronostic vital des patients en clinique. Le modèle gagnerait à être optimisé et amélioré pour se rapprocher encore davantage de la réalité en clinique et ainsi mieux refléter les différentes combinaisons mutationnelles. La translocation chromosomique suffit pour déclencher un développement cancéreux, mais il reste encore à déterminer comment certaines mutations additionnelles peuvent aggraver le pronostic des patients et modifier la sensibilité à des molécules thérapeutiques.

Enfin, l'accès à nos données génomiques et épigénomiques pourrait servir de base pour des recherches futures. L'utilisation récente de molécules ciblant les mécanismes épigénétiques est très prometteuse (Steinhilber and Marschalek 2018). Dans cette thèse, nous nous sommes concentrés plus particulièrement sur la méthylation de l'ADN, la chromatine et les marques d'histones. Il reste encore de nombreux autres mécanismes qui ne sont pas encore totalement compris, comme notamment les petits ARN interférents. Touchant particulièrement les patients pédiatriques, il y a un besoin urgent de cartographier les mécanismes d'action des fusions de gènes dans la biologie du cancer. Une meilleure caractérisation de la pathophysiologie de la leucémie pourrait ainsi permettre d'explorer des avenues thérapeutiques plus ciblées.

## 5.5 Références

Behan, F. M., F. Iorio, G. Picco, E. Goncalves, C. M. Beaver, G. Migliardi, R. Santos, Y. Rao, F. Sassi, M. Pinnelli, R. Ansari, S. Harper, D. A. Jackson, R. McRae, R. Pooley, P. Wilkinson, D. van der Meer, D. Dow, C. Buser-Doepner, A. Bertotti, L. Trusolino, E. A. Stronach, J. Saez-Rodriguez, K. Yusa and M. J. Garnett (2019). "Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens." *Nature* **568**(7753): 511-516.

Bill, M., K. Mrozek, J. Kohlschmidt, A. K. Eisfeld, C. J. Walker, D. Nicolet, D. Papaioannou, J. S. Blachly, S. Orwick, A. J. Carroll, J. E. Kolitz, B. L. Powell, R. M. Stone, A. de la Chapelle, J. C. Byrd and C. D. Bloomfield (2020). "Mutational landscape and clinical outcome of patients with de novo acute myeloid leukemia and rearrangements involving 11q23/KMT2A." *Proc Natl Acad Sci U S A* **117**(42): 26340-26346.

Bolouri, H., J. E. Farrar, T. Triche, Jr., R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvin, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard and S. Meshinchi (2018). "The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions." *Nat Med* **24**(1): 103-112.

Chen, X., D. B. Burkhardt, A. A. Hartman, X. Hu, A. E. Eastman, C. Sun, X. Wang, M. Zhong, S. Krishnaswamy and S. Guo (2019). "MLL-AF9 initiates transformation from fast-proliferating myeloid progenitors." *Nat Commun* **10**(1): 5767.

Jeong, J., A. Jager, P. Domizi, M. Pavel-Dinu, L. Gojenola, M. Iwasaki, M. C. Wei, F. Pan, J. L. Zehnder, M. H. Porteus, K. L. Davis and M. L. Cleary (2019). "High-efficiency CRISPR induction of t(9;11) chromosomal translocations and acute leukemias in human blood stem cells." *Blood Adv* **3**(19): 2825-2835.

Kuriyama, M., N. Harada, S. Kuroda, T. Yamamoto, M. Nakafuku, A. Iwamatsu, D. Yamamoto, R. Prasad, C. Croce, E. Canaani and K. Kaibuchi (1996). "Identification of AF-6 and canoe as putative targets for Ras." *J Biol Chem* **271**(2): 607-610.

Lavallee, V. P., I. Baccelli, J. Krosch, B. Wilhelm, F. Barabe, P. Gendron, G. Boucher, S. Lemieux, A. Marinier, S. Meloche, J. Hebert and G. Sauvageau (2015). "The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias." *Nat Genet* **47**(9): 1030-1037.

Manara, E., E. Baron, C. Tregnago, S. Aveic, V. Bisio, S. Bresolin, R. Masetti, F. Locatelli, G. Basso and M. Pigazzi (2014). "MLL-AF6 fusion oncogene sequesters AF6 into the nucleus to trigger RAS activation in myeloid leukemia." *Blood* **124**(2): 263-272.

Numata, A., H. S. Kwok, A. Kawasaki, J. Li, Q. L. Zhou, J. Kerry, T. Benoukraf, D. Bararia, F. Li, E. Ballabio, M. Tapia, A. J. Deshpande, R. S. Welner, R. Delwel, H. Yang, T. A. Milne,

R. Taneja and D. G. Tenen (2018). "The basic helix-loop-helix transcription factor SHARP1 is an oncogenic driver in MLL-AF6 acute myelogenous leukemia." Nat Commun **9**(1): 1622.

Prange, K. H. M., A. Mandoli, T. Kuznetsova, S. Y. Wang, A. M. Sotoca, A. E. Marneth, B. A. van der Reijden, H. G. Stunnenberg and J. H. A. Martens (2017). "MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia." Oncogene **36**(23): 3346-3356.

Sarrou, E., L. Richmond, R. J. Carmody, B. Gibson and K. Keeshan (2020). "CRISPR Gene Editing of Murine Blood Stem and Progenitor Cells Induces MLL-AF9 Chromosomal Translocation and MLL-AF9 Leukaemogenesis." Int J Mol Sci **21**(12).

Steinhilber, D. and R. Marschalek (2018). "How to effectively treat acute leukemia patients bearing MLL-rearrangements ?" Biochem Pharmacol **147**: 183-190.

# **ANNEXE I**

## **ANALYSE TRANSCRIPTOMIQUE ET GÉNOMIQUE DU CANCER : OUTILS BIOINFORMATIQUES ET ENJEUX ACTUELS**

# **AI Analyse transcriptomique et génomique du cancer : outils bioinformatiques et enjeux actuels**

## **AI.1 Préface**

Les avancées dans les technologies de séquençage nouvelle génération ont eu un impact majeur dans le domaine de la génomique du cancer. Avec les quantités importantes de données générées à partir d'échantillons de tumeurs, la communauté scientifique a dû améliorer et développer des outils pour analyser les données brutes. Bien qu'une grande partie de cet effort ait été concentrée sur l'optimisation d'algorithmes pour obtenir des résultats plus rapides et plus précis, l'accessibilité des données finales reste un enjeu important. De grandes quantités de données peuvent être générées mais ne sont pas facilement accessibles aux chercheurs n'ayant pas les ressources ni l'expérience pour les analyser. Le but de cette revue de littérature est de présenter les outils bioinformatiques disponibles pour explorer des données de *RNA-seq* mais également de souligner le besoin d'en développer de nouveaux pour faciliter l'accessibilité à ce type de données.

## **AI.2 Contributions**

J'ai contribué à l'écriture et la révision du manuscrit conjointement avec Brian T Wilhelm.

## **AI.3 Référence de la publication**

Thomas Milan, Brian T Wilhelm. "Mining Cancer Transcriptomes: Bioinformatic Tools and the Remaining Challenges." *Molecular Diagnosis & Therapy* (2017). PMID: 28229366.

## **AI.4 Article**

# **Mining Cancer Transcriptomes: Bioinformatic Tools and the Remaining Challenges.**

**Thomas Milan<sup>1</sup>, Brian T Wilhelm<sup>1,2</sup>**

<sup>1</sup>Laboratory for High Throughput Biology, Institute for Research in Immunology and Cancer

<sup>2</sup>Department of Medicine, Université de Montréal, Montréal, QC, Canada

\*Correspondence to:

Brian T Wilhelm

Institute for Research in Immunology and Cancer

Université de Montréal

P.O. Box 6128, Station Centre-Ville

Montreal, Québec, Canada H3C 3J7

Phone : +1 514 343.6111 ext 0923

Fax : +1 514 343.7780

E-mail : [brian.wilhelm@umontreal.ca](mailto:brian.wilhelm@umontreal.ca)

## **AI.4.1 Abstract**

The development of next-generation sequencing technologies has had a profound impact on the field of cancer genomics. With the enormous quantities of data being generated from tumor samples, researchers have had to rapidly adapt tools or develop new ones to analyse the raw data to maximize its value. While much of this effort has been focused on improving specific algorithms to get faster and more precise results, the accessibility of the final data for the research community remains a significant problem. Large amounts of data exist but are not easily available to researchers who lack the resources and experience to download and reanalyze them. In this article, we focus on RNA-seq analysis in the context of cancer genomics and discuss the bioinformatic tools that are available to explore these data. We also highlight the importance of developing new and more intuitive tools to provide easier access to public data and discuss the related issues of data sharing and patient privacy.

### **Key points**

Advances in DNA sequencing technologies have enabled unprecedented growth of the data from cancer genomics projects, which has in turn fueled the development of diverse bioinformatic tools to analyze these data.

The enormous investments into cancer genomics have provided a detailed initial view of the genetics of cancer; however, these data likely hold many other important insights that will only be uncovered through the development of new intuitive tools for data exploration.

In addition to new tools, the maximal value of the investments into cancer genomics will only be fully realized when all of the underlying genomic and clinical data from patients are freely available to be reanalyzed.



## AI.4.2 Introduction

The last decade has seen a dramatic change in the study of cancer biology, driven by rapid and continuous technological advances in DNA sequencing. The application of these advances has impacted on virtually every aspect of biomedical research and resulted in enormous amounts of data being generated for virtually every experimental model or system and context. In particular, these advances have fundamentally altered the paradigm of cancer genomics (Mardis and Wilson 2009), making it feasible for the first time to envision personalized treatment based on tumour genome sequencing in a clinical setting in a timeframe of weeks.

Pioneering large-scale studies of cancer patient cohorts such as The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research 2008, Cancer Genome Atlas Research 2011, Cancer Genome Atlas 2012, Cancer Genome Atlas Research, Kandoth et al. 2013, Cancer Genome Atlas Research, Ley et al. 2013) and affiliated projects have leveraged this technology to generate encyclopedic descriptions of many common cancer types. These datasets have arguably had the same impact on cancer biology as the first published human genome had for genetics, opening a detailed window into every aspect of “broken” genomes. While radical improvements in both the cost and the speed of sequencing technologies have made it possible to generate the raw data required for these projects, it has simultaneously highlighted two critically important bottlenecks in the entire process. The first is our remarkably incomplete understanding of cancer biology. In the majority of large-scale studies of cancer patients, the frequency of recurrently mutated genes follows an asymptotic curve, creating a “long tail” of mutated genes (Garraway and Lander 2013). Understanding why patients at the far end of the distribution, in some cases with no recurrently mutated genes, develop the same cancer as those at the other end remains a critical test for the field.

The second bottleneck, which is the subject of this review, is the development of bioinformatic tools to maximize the value of the raw data being generated. In fact, the essential elements of the process of next-generation sequencing bioinformatic analysis are relatively straightforward and have not changed significantly since the initial application of the technology. The sequencing technologies that currently dominate the market place produce billions of short-sequence reads that must be mapped to a reference genome before they can be

biologically interpreted. After this step, the sequence data can be analyzed to calculate gene expression values, detect mutations, RNA editing, translocations and alternative splicing.

To keep pace with the ever-increasing raw output of sequencing machines, bioinformaticians have historically focused efforts on developing faster mapping algorithms, since virtually every other aspect of analysis depends on this first step. Numerous innovative approaches have enabled the development of algorithms that represent a reasonable compromise between speed and sensitivity. At the same time, many of these bioinformatic tools were designed primarily to address the efficiency of the specific algorithm implemented rather than make them easier to use. As a result, the use of these tools is largely restricted to those with the technical competence to compile and install them.

### AI.4.3 RNA-seq in Cancer Genomics

Since its first application to cancer cell lines, transcriptome sequencing (Bainbridge, Warren et al. 2006) (RNA-seq) has become one of the principal tools used to investigate cancer samples. Although similar approaches for examining gene expression patterns in cancers using microarrays have already been used extensively, RNA-seq provides a more precise and less biased view of these data (Wilhelm, Briau et al. 2011, Zhao, Fung-Leung et al. 2014). Because total RNA is collected without regard to genome annotation, it is possible to simultaneously measure gene expression and splicing differences, detect mutations in expressed genes, and measure allelic biases in a single experiment. As a result, RNA-seq enables a relatively inexpensive yet comprehensive view of the genetic changes in tumors as well as discovery of novel transcribed elements.

The technical aspects of RNA-seq are comprehensively reviewed elsewhere (Griffith, Walker et al. 2015, Conesa, Madrigal et al. 2016); however, as a very brief overview, once RNA is extracted from tumor cells, the ribosomal RNA (rRNA) (representing a large majority of the material) is typically removed. This can be performed either through selective enrichment of polyadenylated transcripts or selective depletion of rRNA using complementary biotinylated probes. While data from both approaches are largely similar, potential biases are introduced into the results because of differences in the efficiency of material capture. The RNA is then fragmented followed by end repair and ligation of sequencing adaptors and reverse transcription. Sequencing libraries are then quantified and sequenced, typically to a depth of ~150M reads for patient samples where comprehensive examination will be performed (e.g., splicing, sub-clonal analysis) or 30M reads to simply look at expression levels similar to the levels proposed by ENCODE (Landt, Marinov et al. 2012).

Once the raw data is generated, the detailed analysis can be launched, which for cancer samples typically focuses on gene expression (typically reported as reads per kilobase of transcript per million (RPKM) mapped reads (Mortazavi, Williams et al. 2008)) and mutation detection. Although the steps required for the initial analysis (e.g., read-mapping, variant detection, etc.) do not vary much in most typical analyses, the underlying software used has been continually evolving to improve the performance and accuracy of the downstream analysis. Indeed, despite the inherent power of RNA-seq, many of the more subtle aspects of the

bioinformatic analysis can be highly susceptible to biases introduced through either software or data analysis. Therefore, while we discuss some of the various types of software tools available, it is important to note that no tool or collection of tools will allow an automated analysis or interpretation of cancer RNA-seq data without human intervention.

In recent years, as RNA-seq has been applied to large cancer cohorts, there has been a growing appreciation that a typical analysis produces only a summary of the bulk population of tumor cells, representing the mean of gene expression values of millions of cells. While bulk cultures can be still highly informative for genetic variations present in small numbers of cells (Ding, Ley et al. 2012), extracting expression data from such mixed populations is not feasible, leading to a loss of information from subclonal populations of tumor cells that are often responsible for disease relapse (Kreso and Dick 2014). To address this problem and gain additional biological insight into tumor heterogeneity, methods to sequence the transcriptome of individual cells (single-cell RNA-seq) have been developed (Tang, Barbacioru et al. 2009, Islam, Kjallquist et al. 2011, Ramskold, Luo et al. 2012). Since its initial development, single-cell RNA-seq has become a powerful tool in cancer research and has been used to characterize transcriptional heterogeneity in human squamous cell carcinoma of the bladder (Zhang, Zhang et al. 2016), patient-derived melanoma (Gerber, Willscher et al. 2016), and lung adenocarcinoma (Min, Kim et al. 2015). Such studies have also been added by similar approaches using normal tissues where the normal cellular processes of differentiation have been characterized (Farlik, Halbritter et al. 2016). In addition to these publications, additional initiatives, such as the recently announced “Human Cell Atlas” project (2016), have been launched to apply this approach on a much larger scale to build a complete, comprehensive, and integrative map describing each human cell. This ambitious initiative would potentially have an important impact on the understanding and treatment of diseases such as cancer.

#### **AI.4.4 Currently Available Bioinformatic Platforms for RNA-seq**

The tools available to analyze RNA-seq data can be broadly grouped into three groups. The first and most traditional type of tools are standalone task-dedicated programs typically optimized to perform very specific computations. These often represent individual steps in a bioinformatic pipeline or analysis schema, where individual tools are selected for each task and then often linked together in a largely automated fashion to allow a standardized analysis to be performed. A selection of standalone tools for RNA-seq is shown in **Table AI.1**. A complete listing is beyond the scope of this article, but excellent reviews are available that compare the strengths and weaknesses of the various tools by analysis task. It is important to note that the vast majority of bioinformatic tools are “biologically agnostic” in that they treat data from cancer and normal samples identically. It remains the responsibility of the user to determine the significance of the biological differences uncovered by the various bioinformatic tools.

Computational task	Examples and references
Read mapping	
<p>The initial step in most standard analyses, alignment of RNA-seq reads to a reference genome is often complicated by the gaps created by aligned cDNA* reads to genomic DNA containing introns. Output is a typically binary sequence alignment (BAM) file.</p>	STAR (Dobin, Davis et al. 2013)
	HISAT (Kim, Langmead et al. 2015)
	TopHat2 (Kim, Pertea et al. 2013)
Splicing analysis	
<p>The identification of alternative splicing events can be performed either through measurement of expression levels of different transcript isoforms or through statistical analysis of reads that indicate the inclusion/exclusion of specific exons.</p>	Cufflinks/Cuffdiff (Trapnell, Hendrickson et al. 2013)
	MISO (Katz, Wang et al. 2010)
	MATS (Shen, Park et al. 2012)
	DEXSeq (Anders, Reyes et al. 2012)
Allele-specific expression	
<p>Although RNA-seq data are required for the identification of biased expression of alleles of a gene, this analysis also requires exact knowledge of the genotype (typically through exome sequencing).</p>	ASEReadCounter (GATK) (McKenna, Hanna et al. 2010)
	MAMBA (Rivas, Pirinen et al. 2013)
	AlleleSeq (Rozowsky, Abyzov et al. 2011)
Variant / mutation detection	
<p>Variants in RNA-seq data are typically detected using the same software as for DNA sequence; however, because of issues in aligning spliced reads to a reference genome,</p>	SAMtools (Li 2011)
	Unified Genotyper (GATK) (McKenna, Hanna et al. 2010)
	Freebayes (arXiv:1207.3907)

systematic mapping errors can be reported as variants.	VarScan2 (Koboldt, Zhang et al. 2012)
Differential gene expression	
One of the most common types of analyses, most methods for the detection of differential expression rely on statistical models of read counts to calculate significant changes.	DESeq2 (Love, Huber et al. 2014)
	CuffDiff (Trapnell, Hendrickson et al. 2013)
	EdgeR (Robinson, McCarthy et al. 2010)
	baySeq (Hardcastle and Kelly 2010)
Fusion detection	
For RNA-seq data for cancer samples, fusion genes can be identified through analysis of paired-end sequencing data to find reads that align to different chromosomes.	SOAPfusion (Wu, Zhang et al. 2013)
	TOPHat-fusion (Kim and Salzberg 2011)
	ChimeraScan (Iyer, Chinnaiyan et al. 2011)

\**cDNA = complementary DNA*

**Table AI.1 – Standalone analysis tools for RNA-seq data.**

Given the number of programs available, many researchers are often uncertain about how to decide which tool to use. For many common tasks, there is still a surprising lack of consolidation in terms of software, and this is complicated by the fact that some test conditions used in publication of new analysis tools (e.g., the use of “synthetic” data) can create an apparent software-specific advantage. In addition, in many cases, the biological questions being asked are too varied to enable a common general solution: every software tool has its unique profile of sensitivity, and the output will also be influenced by the quality of initial material and data. Despite these potential complications, most publications rely on a handful of well-tested tools, any of which will be suitable for most users.

The second group of bioinformatic resources can be described as analysis suites, since they typically represent a pre-assembled collection of tools, often including the most commonly used software listed in **Table AI.1**. These analysis suites are either downloaded to run locally or, more usefully, can be accessed through web-based graphical user interfaces. The significant advantage of accessing these resources through a web-based service is that a wide range of tools are often available without the difficulties for end users of compiling and installing these tools locally in a Unix environment. In addition, the web servers hosting the tools can access significantly more computational power than is available through desktop workstations. With “drag and drop” functionality or menu-based selection, users are typically able to analyze their data in a much more autonomous fashion using these suites, although the final interpretation of the data remains a product of the tools used (e.g., it is still possible to select a non-optimal tool for RNA-seq analysis). Also included in this list of tools are a number of programs specifically designed for use with single-cell RNA-seq (scRNA-seq) data. While some features of scRNA-seq analysis are essentially no different than those for bulk cultures (e.g., read mapping, generating expression counts), certain other aspects have been refined to account for the unique data source, including quality control (McCarthy, Campbell et al. 2017) and identification of major cell types present in the bulk population (Guo, Wang et al. 2015). Other scRNA-seq tools allow the integration of data (e.g., to explore and combine genetic, epigenetic, and transcriptomic heterogeneity (Hou, Guo et al. 2016)) or analysis steps (Diaz, Liu et al. 2016) (**Table AI.2**).



Name	Description		
Galaxy	An extensive collection of NGS data analysis tools (including for RNA-seq) accessed through a website menu that allows comprehensive analysis pipelines to be easily built, saved and shared	URL	<a href="https://usegalaxy.org">https://usegalaxy.org</a>
		Usage	Online or local install
		Reference	(Afgan, Baker et al. 2016)
		Commercial?	N
NCBI-Genomics	Downloadable tool suite that allows analysis and visualization of NGS data	URL	<a href="http://www.ncbi.nlm.nih.gov/tools/gbench">http://www.ncbi.nlm.nih.gov/tools/gbench</a>
		Usage	Local install
		Reference	NA
		Commercial?	N
GATK	Comprehensive analysis pipeline built on best practices standards that focuses on mapping and variant calling	URL	<a href="https://software.broadinstitute.org/gatk">https://software.broadinstitute.org/gatk</a>
		Usage	Online or local install
		Reference	(McKenna, Hanna et al. 2010)
		Commercial?	N
GenomeSpace	Cloud-based analysis framework designed to function as a "bridge" to allow seamless transfer of data between external tools (Galaxy, Cytoscape, IGV, UCSC, etc.) with automatic format conversions	URL	<a href="http://www.genomespace.org">http://www.genomespace.org</a>
		Usage	Online
		Reference	(Qu, Garamszegi et al. 2016)
		Commercial?	N

easyRNASeq	R/Bioconductor package for post-mapping analysis/annotation for RNA-seq including DEG detection	<b>URL</b>	<a href="http://bioconductor.org/packages/release/bioc/html/easyRNASeq.html">http://bioconductor.org/packages/release/bioc/html/easyRNASeq.html</a>
		<b>Usage</b>	Local install
		<b>Reference</b>	(Delhomme, Padioleau et al. 2012)
		<b>Commercial?</b>	N
ViennaNGS	Collection of PERL-based programs that allow post-mapping analysis of RNA-seq data, including expression, splicing, and visualization through UCSC browser	<b>URL</b>	<a href="https://github.com/mtw/Bio-ViennaNGS">https://github.com/mtw/Bio-ViennaNGS</a>
		<b>Usage</b>	Local install
		<b>Reference</b>	(Wolfinger, Fallmann et al. 2015)
		<b>Commercial?</b>	N
GenePattern	Web-based framework that allows an extensive collection of analysis modules (including for RNA-seq) to be assembled (by drag and drop) into a shareable pipeline	<b>URL</b>	<a href="http://software.broadinstitute.org/cancer/software/genepattern">http://software.broadinstitute.org/cancer/software/genepattern</a>
		<b>Usage</b>	Online and local install
		<b>Reference</b>	(Reich, Liefeld et al. 2006)
		<b>Commercial?</b>	N
BaseSpace	Free-for-use, cloud-based collection of public software tools ("apps") that allow data analysis in addition to storage	<b>URL</b>	<a href="https://basespace.illumina.com/home/index">https://basespace.illumina.com/home/index</a>
		<b>Usage</b>	Online
		<b>Reference</b>	NA
		<b>Commercial?</b>	Y
Strand NGS	Commercial analysis suite that includes functionality for RNA-seq (DEG detection, splicing, etc)	<b>URL</b>	<a href="http://www.strand-ngs.com">http://www.strand-ngs.com</a>
		<b>Usage</b>	Local install
		<b>Reference</b>	NA
		<b>Commercial?</b>	Y

SINCERA	Pipeline based on R/Bioconductor packages and expression values from scRNA-seq, and used for the identification and distinction of major cell types and gene signatures	URL	<a href="https://research.cchmc.org/pbge/sincera.html">https://research.cchmc.org/pbge/sincera.html</a>
		Usage	Local install
		Reference	(Guo, Wang et al. 2015)
		Commercial?	N
Sincell	R package that specifically provides cell-state hierarchies in the process of cell differentiation thanks to statistical support	URL	<a href="http://bioconductor.org/packages/release/bioc/html/sincell.html">http://bioconductor.org/packages/release/bioc/html/sincell.html</a>
		Usage	Local install
		Reference	(Julia, Telenti et al. 2015)
		Commercial?	N
Scell	Integrated software tool with an intuitive graphical interface that includes a complete analysis of scRNA-seq experiments (mapping, gene-expression quantification, clustering, gene-expression gradients)	URL	<a href="https://github.com/diazlab/SCell">https://github.com/diazlab/SCell</a>
		Usage	Local install
		Reference	(Diaz, Liu et al. 2016)
		Commercial?	N
Scater	R/Bioconductor package for quality control and normalization of scRNA-seq data	URL	<a href="http://bioconductor.org/packages/scater">http://bioconductor.org/packages/scater</a>
		Usage	Local install
		Reference	(McCarthy, Campbell et al. 2017)
		Commercial?	N

**Table AI.2 – Tool suites for RNA-seq analysis.**

\**DEG* differentially expressed genes, *NGS* next-generation sequencing, *scRNA* single-cell RNA, *UCSC* University of California Santa Cruz

The last group of tools highlights an encouraging trend in the field of large-scale studies of investing significant resources to prioritise not only data generation but also the development of bioinformatic tools to provide access to data after the initial publication. Because of the focus on facilitating the reanalysis of published data, these tools blur the distinction between databases and analysis tools; however, custom data can also be uploaded in some cases. A brief summary of these tools is shown in **Table AI.3**.

Name	Types of tools available	Amount and origins of data	Visualization
ICGC data portal	Advanced search form allows extensive summaries to be generated for individual genes or patients within or across cancer projects  (International Cancer Genome, Hudson et al. 2010)	70 cancer genome projects (TCGA and others), 16,000 samples	Interactive, detailed visualization of donors, genes, and mutations by project and across cancer tissue
GDC-Data portal	Restricted to data-transfer tools with links to cBioPortal and BROAD firehose	39 cancer genome projects (TCGA and TARGET), 14,500 samples	Interactive, detailed visualization of samples, statistics and subgroups by project and cancer tissue
cBioPortal	Highly interactive web plots of published cancer cohorts and extensive gene-centric tools for expression, mutations, survival, etc. (Gao, Aksoy et al. 2013)	147 cancer genomics datasets (TCGA and others) 31,400 samples	Visualization of genomic alterations in a set of genes, samples, and pathways.
UCSC Cancer Genomics Browser	Tools to display, investigate, and analyse cancer genomics data and its associated clinical information  (Zhu, Sanborn et al. 2009)	473 datasets, of 25 cancers from TCGA from the CCLE and Stand Up To Cancer.	Several datasets can be viewed at the same time to compare different data and cancer types.

canEvolve	Gene, microRNA, and protein expression profiles. Copy number alterations, protein-protein interaction information  (Samur, Yan et al. 2013)	90 cancer genomics studies from more than 10,000 patients, including 15 TCGA datasets containing 4,800 patients profiles	Various types, including heatmaps, scatterplots or network structure
CGWB	Computational platform to integrate clinical tumor mutation profiles with the reference human genome  (Zhang, Finney et al. 2007)	User data	Genome browser, heatmap viewer, plots
NONCODE	Non-coding RNA from 16 species such as human and mouse. Basic information (location, strand, exon number, length and sequence) and advanced information (expression profile, disease relation, relative literature)  (Bu, Yu et al. 2012)	Regularly updated (in 2016, 487,164 long non-coding RNA). From literature (PubMed) and many other databases (Ensembl, RefSeq, lncRNAdb, GENCODE)	Various, including tables and sequence/expression plots with tabular gene information
GDSC	Drug sensitivity data, integrated with large genomic datasets from the COSMIC  (Yang, Soares et al. 2013)	Drug-response data in cancer cell lines, describing response to 265 anticancer drugs across 1074 cancer cell lines	Various, including volcano plots for genetic anomalies, data tables, and gene expression heatmaps

GTEX	eQTL, gene expression, isoform usage in ~50 normal tissues.  (Mele, Ferreira et al. 2015)	RNA-seq data and genotype data for 12,000 samples of normal tissue	Boxplots for expression and gene cartoons for isoform usage across tissues
------	---	--	--

**Table AI.3 – Data repositories with integrated analysis tools.**

*CCLE* Cancer Cell Line Encyclopedia, *CGWB* Cancer Genome WorkBench, *COSMIC* Catalogue of Somatic Mutations in Cancer database, *eQTL* expression Quantitative Trait Loci, *GDC* Genomic Data Commons, *GDSC* Genomics of Drug Sensitivity in Cancer, *ICGC* International Cancer Genome Consortium, *TCGA* The Cancer Genome Atlas, *GTEX* gene-tissue expression, *PMID* PubMed unique identifier, *UCSC* University of California Santa Cruz

## **AI.4.5 Challenges / Current Gaps**

As noted, the investment of millions of dollars in cancer genomics has provided an invaluable reference framework for a wide range of cancers while highlighting our inability to precisely define the genetic drivers of every patient. In turn, this gap in understanding highlights the importance of the development of new and intuitive tools to enable researchers to explore their data in an unbiased fashion and make new connections that can fill in the gaps in our current knowledge. While many technical challenges remain, other aspects relate to larger issues about the way in which information is shared.

### **AI.4.5.1 Data Visualization / Exploration**

One of the most daunting challenges for researchers working with a large patient dataset is to try and uncover the common elements responsible for the disease. Given the incredible genetic heterogeneity of cancer, along with the normal variations in human genomes, this process often involves iterative examinations of the data to uncover connections. As a result, the time involved in arriving at an experimentally testable hypothesis becomes the limiting factor. For example, imagine a researcher with data for gene expression and somatic mutations in 500 patient samples of breast cancer who wants to understand how the two might be correlated. The researcher may have an idea of several well-known genes to start looking at, but what about potentially novel genes? It is possible that the expression of a particular gene may be highly correlated with an infrequent mutation, highlighting a new biological subgroup within the patients. If the researcher has access to a tool that allows this type of data exploration, checking for such a correlation might take several minutes, leading to a clear “yes” or “no”. On the other hand, if the researcher performing this analysis manually (e.g., using formulas in spreadsheets) over several hours is far less likely to test this possibility and might instead focus on the well-known genes. Therefore, the ability to perform data exploration to rapidly test (and/or eliminate) several potential hypotheses before focusing the analysis on a specific aspect is extremely valuable. While there are many ways to potentially address this bottleneck, solutions that create an informative visualization of the data can also leverage the highly evolved ability of human brains for pattern recognition (Mattson 2014). Despite the obvious importance of simple but powerful tools to integrate and visualize data, there are still remarkable few that allow users to simultaneously take full advantage of the years of investment in important “static”



resources such as KEGG (Kanehisa, Sato et al. 2016), dbSNP (Sherry, Ward et al. 1999), and GO (Ashburner, Ball et al. 2000) when analyzing their data.

#### **AI.4.5.2 Data Sharing and Standards**

A basic principle of science is that experimental results, or in the context of cancer genomics, the analysis and interpretation of the data, should be reproducible. For this to be possible in cancer genomics, the underlying raw data and the methodology used for the initial analysis must be publically stored. This creates a variety of challenges for those submitting the data, the storage repositories, and users who wish to reuse these data. With respect to annotation of cancer genomics RNA-seq data, the bulk of these have been generated in the context of large-scale projects with sufficient resources to ensure the sample data and procedures used have been adequately documented. As a result, small projects for which information may be incomplete are likely to have a relatively smaller impact.

The rate at which data are being generated remains a significant challenge for public databases, and the amount of data being generated each year is estimated to reach the level of exabytes in the near future (Stephens, Lee et al. 2015). Although hardware costs for storage of this information continue to decrease, additional efforts will likely be required to help researchers find relevant information for their project (mirroring what many internet search engines did as the number of web pages grew exponentially). Lastly, for end users wishing to reuse published data, downloading raw data from large cohorts and manipulating these files remains one of the biggest challenges. Although data compression methods have been developed that can improve the portability of data (Benoit, Lemaitre et al. 2015, Zhang, Li et al. 2015, Saha and Rajasekaran 2016), some methods come with the cost of some data loss during compression. An alternative solution to this problem is to focus on cloud-based storage with built-in tools suites to obviate the problem of data transmission. The current *ad hoc* process of archiving raw data, although preferable in many ways, will very likely be unsustainable within several years.

### **AI.4.5.3 Privacy Issues and Access to Patient Clinical Data**

As valuable as the raw RNA-seq data from cancer patient genomes are, the true value is often realized when correlating all of the genetic anomalies with the clinical parameters of patient cohorts and with treatment outcomes. Since the start of cancer genome sequencing, accessing both the raw sequence data along with such clinical data has been difficult in many cases, limiting the utility of the data. This can be partly explained through data-release policies designed to allow researchers who have generated the data sufficient time to publish their findings. In addition, privacy concerns with respect to patient data are often cited as the reason behind the long and bureaucratic process typically required to access patient data. While privacy issues do have a reasonable basis, legal protections are already in place in many jurisdictions that would prevent the inappropriate use of such information (McGuire and Majumder 2009, Otlowski, Taylor et al. 2012, Joly, Ngueng Feze et al. 2013). Moreover, while most researchers view only the final data from anonymous banked samples, these samples have all been collected under informed consent, typically with the explicit understanding that patients are sharing the genetic information about their disease in the hopes of helping others. In stark contrast to these altruistic motivations, some recent publications have suggested that the sharing of the scientific data derived from clinical studies should actually be discouraged. Part of the justification presented is to prevent “data parasites” from gaining benefit from the effort to generate the data (Taichman, Backus et al. 2016) (including validating the original study’s conclusions). Such dangerous and irrational arguments not only run contrary to spirit of the participating patients but also undermine the scientific premise of the reproducibility of experimental results and their interpretation.

### **AI.4.5.4 Potential Solutions**

It is clear that cancer genomics RNA-seq data can improve not only our understanding of cancer but also ultimately the diagnosis and treatment of it. Therefore, it is critical to identify ways to address the problems described. For the types of bioinformatic tools available, the challenge of creating simple tools for data exploration rests squarely on the shoulders of research groups developing novel analysis tools. That is, at least as much thought should go into the design of the user interface as into any technical aspect related to the algorithm of the software. Achieving this goal will also require a shift from software currently being developed to study

cancer genomes through simple static displays and small incremental improvements in speed or efficiency. Such improvements hold little value for improving patient treatments when, for example, we cannot understand the biological importance for the majority of single nucleotide variations that current software tools would accurately identify. Addressing this need will call for new and creative ways to both display data and allow users to explore it in multiple dimensions. In principle, large genomics centers or groups of centers are probably best placed to take the lead in the development of such tools.

The problems of data storage, annotation, and sharing are collective problems for the entire research community. As such, they require solutions that are both precise enough to deal with cancer genomics data and flexible enough to also apply to most other data types. There is a potential opportunity now to support the move towards closer integration of cloud-based storage and the data analysis tools for cancer genomics. As noted, there are already a number of examples of sites (e.g., cBioportal (Gao, Aksoy et al. 2013), GenomeSpace (Qu, Garamszegi et al. 2016)) where data and analysis tools co-exist; however, there is still duplication of effort in some cases (e.g., accessing cBioPortal (Gao, Aksoy et al. 2013)/Galaxy (Afgan, Baker et al. 2016) through GenomeSpace (Qu, Garamszegi et al. 2016)). Given the growth and popularity of such sites (which speaks to the importance of the needs that they meet), their continued evolution may eventually lead towards a natural consolidation into a single central resource. One group with an enormous vested interest in this process includes the governmental (and private) funding agencies who have collectively invested hundreds of millions of dollars into cancer genomics projects. While some funders have recognized the importance of not only generating the data but also accurately curating it and ensuring it can easily be reused, the overall efforts in this regard are on a significantly smaller scale. If even a small percentage of the money invested into cancer genomics projects could be shifted from sequencing to a sustained effort to curate the data and develop tools to make reanalysis trivial, these agencies could easily leverage their investments for sustainable long-term research gains, to the benefit of all.

For the last issue of patient privacy and access to clinical data, there is perhaps no simple solution to address all concerns. Even if the methodology for obtaining “informed consent” could somehow be improved (Jefford and Moore 2008) (e.g., whereby patients could explicitly choose to waive their medical privacy), the impact of such efforts could still be obstructed by

researchers who harbour proprietary rights over patient data generated with public funding. In our modern and technologically interconnected society, we are reminded on a weekly basis that the privacy of electronic personal data (medical or otherwise), and security over that privacy, simply do not exist. In such an environment, the extensive and often onerous bureaucratic regulations that have been established with the intent of protecting patient privacy make little sense. It is certain that any significant move towards making clinical data of donors freely available would likely bring new ethical and societal challenges. Nevertheless, it is difficult to argue that this would be worse than the current situation, where the potential returns on the investments into cancer genomics are systematically impeded.

#### **AI.4.6 Conclusion**

To maximize the value of the research funding invested in generating cancer genomics data, and to accelerate our understanding of cancer, any researcher working in the field should be able to confidently explore their data (or that of others) to generate testable hypotheses regardless of their level of expertise in bioinformatics. Although many bottlenecks still exist, one of the most pressing ultimately remains our ability to interpret the masses of data currently being generated.

#### **AI.4.7 Additional information**

**Conflict of Interest:** Thomas Milan and Brian T. Wilhelm have no conflicts of interest.

**Funding:** This work was supported by Grants to BTW from the Fonds de Recherche du Québec en Santé (32900), and the Terry Fox Research Institute (TFRI-NI-1042).

## AI.4.8 References

(2016). "Human Cell Atlas." from <https://www.humancellatlas.org/>.

Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko and J. Goecks (2016). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update." *Nucleic Acids Res* **44**(W1): W3-W10.

Anders, S., A. Reyes and W. Huber (2012). "Detecting differential usage of exons from RNA-seq data." *Genome Res* **22**(10): 2008-2017.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nat Genet* **25**(1): 25-29.

Bainbridge, M. N., R. L. Warren, M. Hirst, T. Romanuik, T. Zeng, A. Go, A. Delaney, M. Griffith, M. Hickenbotham, V. Magrini, E. R. Mardis, M. D. Sadar, A. S. Siddiqui, M. A. Marra and S. J. Jones (2006). "Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach." *BMC Genomics* **7**: 246.

Benoit, G., C. Lemaître, D. Lavenier, E. Drezen, T. Dayris, R. Uricaru and G. Rizk (2015). "Reference-free compression of high throughput sequencing data with a probabilistic de Bruijn graph." *BMC Bioinformatics* **16**: 288.

Bu, D., K. Yu, S. Sun, C. Xie, G. Skogerbo, R. Miao, H. Xiao, Q. Liao, H. Luo, G. Zhao, H. Zhao, Z. Liu, C. Liu, R. Chen and Y. Zhao (2012). "NONCODE v3.0: integrative annotation of long noncoding RNAs." *Nucleic Acids Res* **40**(Database issue): D210-215.

Cancer Genome Atlas, N. (2012). "Comprehensive molecular characterization of human colon and rectal cancer." *Nature* **487**(7407): 330-337.

Cancer Genome Atlas Research, N. (2008). "Comprehensive genomic characterization defines human glioblastoma genes and core pathways." *Nature* **455**(7216): 1061-1068.

Cancer Genome Atlas Research, N. (2011). "Integrated genomic analyses of ovarian carcinoma." *Nature* **474**(7353): 609-615.

Cancer Genome Atlas Research, N., C. Kandoth, N. Schultz, A. D. Cherniack, R. Akbani, Y. Liu, H. Shen, A. G. Robertson, I. Pashtan, R. Shen, C. C. Benz, C. Yau, P. W. Laird, L. Ding, W. Zhang, G. B. Mills, R. Kucherlapati, E. R. Mardis and D. A. Levine (2013). "Integrated genomic characterization of endometrial carcinoma." *Nature* **497**(7447): 67-73.

Cancer Genome Atlas Research, N., T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E.

Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing and G. Eley (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." *N Engl J Med* **368**(22): 2059-2074.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczesniak, D. J. Gaffney, L. L. Elo, X. Zhang and A. Mortazavi (2016). "A survey of best practices for RNA-seq data analysis." *Genome Biol* **17**: 13.

Delhomme, N., I. Padioleau, E. E. Furlong and L. M. Steinmetz (2012). "easyRNASeq: a bioconductor package for processing RNA-Seq data." *Bioinformatics* **28**(19): 2532-2533.

Diaz, A., S. J. Liu, C. Sandoval, A. Pollen, T. J. Nowakowski, D. A. Lim and A. Kriegstein (2016). "SCell: integrated analysis of single-cell RNA-seq data." *Bioinformatics* **32**(14): 2219-2220.

Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson and J. F. DiPersio (2012). "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing." *Nature* **481**(7382): 506-510.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* **29**(1): 15-21.

Farlik, M., F. Halbritter, F. Muller, F. A. Choudry, P. Ebert, J. Klughammer, S. Farrow, A. Santoro, V. Ciaurro, A. Mathur, R. Uppal, H. G. Stunnenberg, W. H. Ouwehand, E. Laurenti, T. Lengauer, M. Frontini and C. Bock (2016). "DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation." *Cell Stem Cell* **19**(6): 808-822.

Gao, J., B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander and N. Schultz (2013). "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal." *Sci Signal* **6**(269): p11.

Garraway, L. A. and E. S. Lander (2013). "Lessons from the cancer genome." *Cell* **153**(1): 17-37.

Gerber, T., E. Willscher, H. Loeffler-Wirth, L. Hopp, D. Schadendorf, M. Scharl, U. Anderegg, G. Camp, B. Treutlein, H. Binder and M. Kunz (2016). "Mapping heterogeneity in patient-derived melanoma cultures by single-cell RNA-seq." *Oncotarget*.

Griffith, M., J. R. Walker, N. C. Spies, B. J. Ainscough and O. L. Griffith (2015). "Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud." *PLoS Comput Biol* **11**(8): e1004393.

Guo, M., H. Wang, S. S. Potter, J. A. Whitsett and Y. Xu (2015). "SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis." *PLoS Comput Biol* **11**(11): e1004575.

Hardcastle, T. J. and K. A. Kelly (2010). "baySeq: empirical Bayesian methods for identifying differential expression in sequence count data." *BMC Bioinformatics* **11**: 422.

Hou, Y., H. Guo, C. Cao, X. Li, B. Hu, P. Zhu, X. Wu, L. Wen, F. Tang, Y. Huang and J. Peng (2016). "Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas." *Cell Res* **26**(3): 304-319.

International Cancer Genome, C., T. J. Hudson, W. Anderson, A. Artez, A. D. Barker, C. Bell, R. R. Bernabe, M. K. Bhan, F. Calvo, I. Eerola, D. S. Gerhard, A. Guttmacher, M. Guyer, F. M. Hemsley, J. L. Jennings, D. Kerr, P. Klatt, P. Kolar, J. Kusada, D. P. Lane, F. Laplace, L. Youyong, G. Nettekoven, B. Ozenberger, J. Peterson, T. S. Rao, J. Remacle, A. J. Schafer, T. Shibata, M. R. Stratton, J. G. Vockley, K. Watanabe, H. Yang, M. M. Yuen, B. M. Knoppers, M. Bobrow, A. Cambon-Thomsen, L. G. Dressler, S. O. Dyke, Y. Joly, K. Kato, K. L. Kennedy, P. Nicolas, M. J. Parker, E. Rial-Sebbag, C. M. Romeo-Casabona, K. M. Shaw, S. Wallace, G. L. Wiesner, N. Zeps, P. Lichter, A. V. Biankin, C. Chabannon, L. Chin, B. Clement, E. de Alava, F. Degos, M. L. Ferguson, P. Geary, D. N. Hayes, T. J. Hudson, A. L. Johns, A. Kasprzyk, H. Nakagawa, R. Penny, M. A. Piris, R. Sarin, A. Scarpa, T. Shibata, M. van de Vijver, P. A. Futreal, H. Aburatani, M. Bayes, D. D. Botwell, P. J. Campbell, X. Estivill, D. S. Gerhard, S. M. Grimmond, I. Gut, M. Hirst, C. Lopez-Otin, P. Majumder, M. Marra, J. D. McPherson, H. Nakagawa, Z. Ning, X. S. Puente, Y. Ruan, T. Shibata, M. R. Stratton, H. G. Stunnenberg, H. Swerdlow, V. E. Velculescu, R. K. Wilson, H. H. Xue, L. Yang, P. T. Spellman, G. D. Bader, P. C. Boutros, P. J. Campbell, P. Flicek, G. Getz, R. Guigo, G. Guo, D. Haussler, S. Heath, T. J. Hubbard, T. Jiang, S. M. Jones, Q. Li, N. Lopez-Bigas, R. Luo, L. Muthuswamy, B. F. Ouellette, J. V. Pearson, X. S. Puente, V. Quesada, B. J. Raphael, C. Sander, T. Shibata, T. P. Speed, L. D. Stein, J. M. Stuart, J. W. Teague, Y. Totoki, T. Tsunoda, A. Valencia, D. A. Wheeler, H. Wu, S. Zhao, G. Zhou, L. D. Stein, R. Guigo, T. J. Hubbard, Y. Joly, S. M. Jones, A. Kasprzyk, M. Lathrop, N. Lopez-Bigas, B. F. Ouellette, P. T. Spellman, J. W. Teague, G. Thomas, A. Valencia, T. Yoshida, K. L. Kennedy, M. Axton, S. O. Dyke, P. A. Futreal, D. S. Gerhard, C. Gunter, M. Guyer, T. J. Hudson, J. D. McPherson, L. J. Miller, B. Ozenberger, K. M. Shaw, A. Kasprzyk, L. D. Stein, J. Zhang, S. A. Haider, J. Wang, C. K. Yung, A. Cros, Y. Liang, S. Gnaneshan, J. Guberman, J. Hsu, M. Bobrow, D. R. Chalmers, K. W. Hasel, Y. Joly, T. S. Kaan, K. L. Kennedy, B. M. Knoppers, W. W. Lowrance, T. Masui, P. Nicolas, E. Rial-Sebbag, L. L. Rodriguez, C. Vergely, T. Yoshida, S. M. Grimmond, A. V. Biankin, D. D. Bowtell, N. Cloonan, A. deFazio, J. R. Eshleman, D. Etemadmoghadam, B. B. Gardiner, J. G. Kench, A. Scarpa, R. L. Sutherland, M. A. Tempero, N. J. Waddell, P. J. Wilson, J. D. McPherson, S. Gallinger, M. S. Tsao, P. A. Shaw, G. M. Petersen, D. Mukhopadhyay, L. Chin, R. A. DePinho, S. Thayer, L. Muthuswamy, K. Shazand, T. Beck, M. Sam, L. Timms, V. Ballin, Y. Lu, J. Ji, X. Zhang, F. Chen, X. Hu, G. Zhou, Q. Yang, G. Tian, L. Zhang, X. Xing, X. Li, Z. Zhu, Y. Yu, J. Yu, H. Yang, M. Lathrop, J. Tost, P. Brennan, I. Holcatova, D. Zaridze, A. Brazma, L. Egevard, E. Prokhortchouk, R. E. Banks, M. Uhlen, A. Cambon-Thomsen, J. Viksna, F. Ponten, K. Skryabin, M. R. Stratton, P. A. Futreal, E. Birney, A. Borg, A. L.



Borresen-Dale, C. Caldas, J. A. Foekens, S. Martin, J. S. Reis-Filho, A. L. Richardson, C. Sotiriou, H. G. Stunnenberg, G. Thoms, M. van de Vijver, L. van't Veer, F. Calvo, D. Birnbaum, H. Blanche, P. Boucher, S. Boyault, C. Chabannon, I. Gut, J. D. Masson-Jacquemier, M. Lathrop, I. Pauporte, X. Pivot, A. Vincent-Salomon, E. Tabone, C. Theillet, G. Thomas, J. Tost, I. Treilleux, F. Calvo, P. Bioulac-Sage, B. Clement, T. Decaens, F. Degos, D. Franco, I. Gut, M. Gut, S. Heath, M. Lathrop, D. Samuel, G. Thomas, J. Zucman-Rossi, P. Lichter, R. Eils, B. Brors, J. O. Korbel, A. Korshunov, P. Landgraf, H. Lehrach, S. Pfister, B. Radlwimmer, G. Reifemberger, M. D. Taylor, C. von Kalle, P. P. Majumder, R. Sarin, T. S. Rao, M. K. Bhan, A. Scarpa, P. Pederzoli, R. A. Lawlor, M. Delledonne, A. Bardelli, A. V. Biankin, S. M. Grimmond, T. Gress, D. Klimstra, G. Zamboni, T. Shibata, Y. Nakamura, H. Nakagawa, J. Kusada, T. Tsunoda, S. Miyano, H. Aburatani, K. Kato, A. Fujimoto, T. Yoshida, E. Campo, C. Lopez-Otin, X. Estivill, R. Guigo, S. de Sanjose, M. A. Piris, E. Montserrat, M. Gonzalez-Diaz, X. S. Puente, P. Jares, A. Valencia, H. Himmelbauer, V. Quesada, S. Bea, M. R. Stratton, P. A. Futreal, P. J. Campbell, A. Vincent-Salomon, A. L. Richardson, J. S. Reis-Filho, M. van de Vijver, G. Thomas, J. D. Masson-Jacquemier, S. Aparicio, A. Borg, A. L. Borresen-Dale, C. Caldas, J. A. Foekens, H. G. Stunnenberg, L. van't Veer, D. F. Easton, P. T. Spellman, S. Martin, A. D. Barker, L. Chin, F. S. Collins, C. C. Compton, M. L. Ferguson, D. S. Gerhard, G. Getz, C. Gunter, A. Guttmacher, M. Guyer, D. N. Hayes, E. S. Lander, B. Ozenberger, R. Penny, J. Peterson, C. Sander, K. M. Shaw, T. P. Speed, P. T. Spellman, J. G. Vockley, D. A. Wheeler, R. K. Wilson, T. J. Hudson, L. Chin, B. M. Knoppers, E. S. Lander, P. Lichter, L. D. Stein, M. R. Stratton, W. Anderson, A. D. Barker, C. Bell, M. Bobrow, W. Burke, F. S. Collins, C. C. Compton, R. A. DePinho, D. F. Easton, P. A. Futreal, D. S. Gerhard, A. R. Green, M. Guyer, S. R. Hamilton, T. J. Hubbard, O. P. Kallioniemi, K. L. Kennedy, T. J. Ley, E. T. Liu, Y. Lu, P. Majumder, M. Marra, B. Ozenberger, J. Peterson, A. J. Schafer, P. T. Spellman, H. G. Stunnenberg, B. J. Wainwright, R. K. Wilson and H. Yang (2010). "International network of cancer genome projects." *Nature* **464**(7291): 993-998.

Islam, S., U. Kjallquist, A. Moliner, P. Zajac, J. B. Fan, P. Lonnerberg and S. Linnarsson (2011). "Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq." *Genome Res* **21**(7): 1160-1167.

Iyer, M. K., A. M. Chinnaiyan and C. A. Maher (2011). "ChimeraScan: a tool for identifying chimeric transcription in sequencing data." *Bioinformatics* **27**(20): 2903-2904.

Jefford, M. and R. Moore (2008). "Improvement of informed consent and the quality of consent documents." *Lancet Oncol* **9**(5): 485-493.

Joly, Y., I. Ngueng Feze and J. Simard (2013). "Genetic discrimination and life insurance: a systematic review of the evidence." *BMC Med* **11**: 25.

Julia, M., A. Telenti and A. Rausell (2015). "Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq." *Bioinformatics* **31**(20): 3380-3382.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe (2016). "KEGG as a reference resource for gene and protein annotation." *Nucleic Acids Res* **44**(D1): D457-462.

Katz, Y., E. T. Wang, E. M. Airoidi and C. B. Burge (2010). "Analysis and design of RNA sequencing experiments for identifying isoform regulation." *Nat Methods* **7**(12): 1009-1015.

- Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." Nat Methods **12**(4): 357-360.
- Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol **14**(4): R36.
- Kim, D. and S. L. Salzberg (2011). "TopHat-Fusion: an algorithm for discovery of novel fusion transcripts." Genome Biol **12**(8): R72.
- Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding and R. K. Wilson (2012). "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing." Genome Res **22**(3): 568-576.
- Kreso, A. and J. E. Dick (2014). "Evolution of the cancer stem cell model." Cell Stem Cell **14**(3): 275-291.
- Landt, S. G., G. K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B. E. Bernstein, P. Bickel, J. B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K. I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A. J. Hartemink, M. M. Hoffman, V. R. Iyer, Y. L. Jung, S. Karmakar, M. Kellis, P. V. Kharchenko, Q. Li, T. Liu, X. S. Liu, L. Ma, A. Milosavljevic, R. M. Myers, P. J. Park, M. J. Pazin, M. D. Perry, D. Raha, T. E. Reddy, J. Rozowsky, N. Shores, A. Sidow, M. Slattery, J. A. Stamatoyannopoulos, M. Y. Tolstorukov, K. P. White, S. Xi, P. J. Farnham, J. D. Lieb, B. J. Wold and M. Snyder (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia." Genome Res **22**(9): 1813-1831.
- Li, H. (2011). "Improving SNP discovery by base alignment quality." Bioinformatics **27**(8): 1157-1158.
- Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.
- Mardis, E. R. and R. K. Wilson (2009). "Cancer genome sequencing: a review." Hum Mol Genet **18**(R2): R163-168.
- Mattson, M. P. (2014). "Superior pattern processing is the essence of the evolved human brain." Front Neurosci **8**: 265.
- McCarthy, D. J., K. R. Campbell, A. T. Lun and Q. F. Wills (2017). "Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R." Bioinformatics.
- McGuire, A. L. and M. A. Majumder (2009). "Two cheers for GINA?" Genome Med **1**(1): 6.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.
- Mele, M., P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segre, S. Djebali, A. Niarchou, G. T. Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T.

- Dermitzakis, K. G. Ardlie and R. Guigo (2015). "Human genomics. The human transcriptome across tissues and individuals." Science **348**(6235): 660-665.
- Min, J. W., W. J. Kim, J. A. Han, Y. J. Jung, K. T. Kim, W. Y. Park, H. O. Lee and S. S. Choi (2015). "Identification of Distinct Tumor Subpopulations in Lung Adenocarcinoma via Single-Cell RNA-seq." PLoS One **10**(8): e0135817.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.
- Otlowski, M., S. Taylor and Y. Bombard (2012). "Genetic discrimination: international perspectives." Annu Rev Genomics Hum Genet **13**: 433-454.
- Qu, K., S. Garamszegi, F. Wu, H. Thorvaldsdottir, T. Liefeld, M. Ocana, D. Borges-Rivera, N. Pochet, J. T. Robinson, B. Demchak, T. Hull, G. Ben-Artzi, D. Blankenberg, G. P. Barber, B. T. Lee, R. M. Kuhn, A. Nekrutenko, E. Segal, T. Ideker, M. Reich, A. Regev, H. Y. Chang and J. P. Mesirov (2016). "Integrative genomic analysis by interoperation of bioinformatics tools in GenomeSpace." Nat Methods **13**(3): 245-247.
- Ramskold, D., S. Luo, Y. C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtkova, J. F. Loring, L. C. Laurent, G. P. Schroth and R. Sandberg (2012). "Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells." Nat Biotechnol **30**(8): 777-782.
- Reich, M., T. Liefeld, J. Gould, J. Lerner, P. Tamayo and J. P. Mesirov (2006). "GenePattern 2.0." Nat Genet **38**(5): 500-501.
- Rivas, M. A., M. Pirinen, M. J. Neville, K. J. Gaulton, L. Moutsianas, T. D. C. Go, C. M. Lindgren, F. Karpe, M. I. McCarthy and P. Donnelly (2013). "Assessing association between protein truncating variants and quantitative traits." Bioinformatics **29**(19): 2419-2426.
- Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.
- Rozowsky, J., A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, N. Bhardwaj, M. Rubin, M. Snyder and M. Gerstein (2011). "AlleleSeq: analysis of allele-specific expression and binding in a network framework." Mol Syst Biol **7**: 522.
- Saha, S. and S. Rajasekaran (2016). "NRGC: a novel referential genome compression algorithm." Bioinformatics.
- Samur, M. K., Z. Yan, X. Wang, Q. Cao, N. C. Munshi, C. Li and P. K. Shah (2013). "canEvolve: a web portal for integrative oncogenomics." PLoS One **8**(2): e56228.
- Shen, S., J. W. Park, J. Huang, K. A. Dittmar, Z. X. Lu, Q. Zhou, R. P. Carstens and Y. Xing (2012). "MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data." Nucleic Acids Res **40**(8): e61.
- Sherry, S. T., M. Ward and K. Sirotkin (1999). "dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation." Genome Res **9**(8): 677-679.

- Stephens, Z. D., S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha and G. E. Robinson (2015). "Big Data: Astronomical or Genomical?" PLoS Biol **13**(7): e1002195.
- Taichman, D. B., J. Backus, C. Baethge, H. Bauchner, P. W. de Leeuw, J. M. Drazen, J. Fletcher, F. A. Frizelle, T. Groves, A. Haileamlak, A. James, C. Laine, L. Peiperl, A. Pinborg, P. Sahni and S. Wu (2016). "Sharing Clinical Trial Data--A Proposal from the International Committee of Medical Journal Editors." N Engl J Med **374**(4): 384-386.
- Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao and M. A. Surani (2009). "mRNA-Seq whole-transcriptome analysis of a single cell." Nat Methods **6**(5): 377-382.
- Trapnell, C., D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn and L. Pachter (2013). "Differential analysis of gene regulation at transcript resolution with RNA-seq." Nat Biotechnol **31**(1): 46-53.
- Wilhelm, B. T., M. Briau, P. Austin, A. Faubert, G. Boucher, P. Chagnon, K. Hope, S. Girard, N. Mayotte, J. R. Landry, J. Hebert and G. Sauvageau (2011). "RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity." Blood **117**(2): e27-38.
- Wolfinger, M. T., J. Fallmann, F. Eggenhofer and F. Amman (2015). "ViennaNGS: A toolbox for building efficient next- generation sequencing analysis pipelines." F1000Res **4**: 50.
- Wu, J., W. Zhang, S. Huang, Z. He, Y. Cheng, J. Wang, T. W. Lam, Z. Peng and S. M. Yiu (2013). "SOAPfusion: a robust and effective computational fusion discovery tool for RNA-seq reads." Bioinformatics **29**(23): 2971-2978.
- Yang, W., J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, S. Ramaswamy, P. A. Futreal, D. A. Haber, M. R. Stratton, C. Benes, U. McDermott and M. J. Garnett (2013). "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells." Nucleic Acids Res **41**(Database issue): D955-961.
- Zhang, J., R. P. Finney, W. Rowe, M. Edmonson, S. H. Yang, T. Dracheva, J. Jen, J. P. Struewing and K. H. Buetow (2007). "Systematic analysis of genetic alterations in tumors using Cancer Genome WorkBench (CGWB)." Genome Res **17**(7): 1111-1117.
- Zhang, X., M. Zhang, Y. Hou, L. Xu, W. Li, Z. Zou, C. Liu, A. Xu and S. Wu (2016). "Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder." Oncotarget **7**(40): 66069-66076.
- Zhang, Y., L. Li, J. Xiao, Y. Yang and Z. Zhu (2015). FQZip: Lossless Reference-Based Compression of Next Generation Sequencing Data in FASTQ Format. Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems - Volume 2. H. Handa, H. Ishibuchi, Y.-S. Ong and K.-C. Tan. Cham, Springer International Publishing: 127-135.
- Zhao, S., W. P. Fung-Leung, A. Bittner, K. Ngo and X. Liu (2014). "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells." PLoS One **9**(1): e78644.

Zhu, J., J. Z. Sanborn, S. Benz, C. Szeto, F. Hsu, R. M. Kuhn, D. Karolchik, J. Archie, M. E. Lenburg, L. J. Esserman, W. J. Kent, D. Haussler and T. Wang (2009). "The UCSC Cancer Genomics Browser." Nat Methods **6**(4): 239-240.

# **ANNEXE II**

## **CURRICULUM VITAE**

## **All Curriculum vitae**

### **All.1 Publications**

**Milan T.**, Celton M., Lagacé K., Roques É., Safa-Tahar-Henni S., Bresson E., Bergeron A., Hebert J., Meshinchi S., Cellot S., Barabé F., Wilhelm B.T. Epigenetic changes in human model KMT2A leukemias highlight early events during leukemogenesis. *Hematologica*. 30 Déc 2020.

Cardin S.\*, Bilodeau M.\*, Roussy M., Aubert L., **Milan T.**, Jouan L., Rouette A., Laramée L., Gendron P., Duchaine J., Decaluwe H., Spinella JF., Mourad S., Couture F., Sinnott D., Haddad É., Landry JR., Ma J., Humphries K., Roux P., Hébert J., Gruber TA., Wilhelm BT and Cellot S. Human models of NUP98-KDM5A megakaryocytic leukemia in mice contribute to uncovering new biomarkers and therapeutic vulnerabilities. *Blood Advances*. 12 Nov 2019

**Milan T.**, Canaj H., Villeneuve C., Ghosh A., Barabé F., Cellot S., Wilhelm B.T. Gene Fusions in Pediatric Leukemia: moving towards more accurate models. *Experimental Hematology*. 30 Mai 2019

**Milan T.** & Wilhelm B.T. Mining Cancer Transcriptomes : Bioinformatic Tools and the Remaining Challenges. *Current Opinion for Molecular Diagnosis & Therapy*. 22 Fév 2017

### **All.2 Participations à des congrès scientifiques**

#### **All.2.1 Présentations orales**

**Mars 2021:** 9th IRIC scientific day (Université de Montréal)

**Novembre 2020:** XXVI Journée scientifique des programmes de biologie moléculaire (Université de Montréal)

**Juin 2019:** CSMB 62nd annual conference model systems in cancer research (Montréal, QC)

**Novembre 2017:** Annual Scientific Meeting 2017. The Terry Fox Research Institute (Vancouver, BC)

**Mai 2017:** 7th IRIC scientific day (Université de Montréal)

## **AII.2.2 Présentations par affiche**

### *International*

- Novembre 2019:** 5th Canadian Cancer Research Conference (Ottawa, ON)
- Août 2019:** 48th annual scientific meeting of the International Society for Experimental Hematology (ISEH) (Brisbane, AUSTRALIE)
- Novembre 2017:** 4th Canadian Cancer Research Conference (Vancouver, BC)

### *Régional*

- Octobre 2018:** 12th International Symposium of IRIC (Machine Learning Applied to Biological Systems) (Université de Montréal)
- Juin 2016:** Montreal International Symposium on Angiogenesis and Metastasis (Treating disseminated cancer by targeting the epigenome). Université McGill (Montréal, QC)
- Juin 2016:** 10th International Symposium of IRIC (The Cell Cycle : from mechanisms to therapy) (Université de Montréal)

### *Académique*

- Mai 2019:** XXV Journée scientifique des programmes de biologie moléculaire (Université de Montréal)
- Mai 2019:** Journée de la fondation Cole
- Mai 2018:** 8th IRIC scientific day (Université de Montréal)
- Mai 2018:** XXIV Journée scientifique des programmes de biologie moléculaire (Université de Montréal)
- Mai 2017:** XXIII Journée scientifique des programmes de biologie moléculaire (Université de Montréal)