



Ferrata Storti Foundation

## 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</sup> Anne Bergeron,<sup>2,3</sup> Josée Hebert,<sup>4,5</sup> Soheil Meshinchi,<sup>6</sup> Sonia Cellot,<sup>7</sup> Frédéric Barabé<sup>2,3,8</sup> and Brian T. Wilhelm<sup>1,5</sup>

<sup>1</sup>Laboratory for High Throughput Biology, Institute for Research in Immunology and Cancer, Montréal, Québec, Canada; <sup>2</sup>Centre de Recherche en Infectiologie du CHUL, Centre de Recherche du CHU de Québec Université Laval, Québec City, Québec, Canada; <sup>3</sup>CHU de Québec Université Laval Hôpital Enfant-Jésus, Québec City, Québec, Canada; <sup>4</sup>Division of Hematology-Oncology and Leukemia Cell Bank of Quebec, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada; <sup>5</sup>Department of Medicine, Université de Montréal, Montréal, Québec, Canada; <sup>6</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, PA, USA; <sup>7</sup>Department of Pediatrics, Division of Hematology, Ste-Justine Hospital, Montréal, Québec, Canada and <sup>8</sup>Department of Medicine, Université Laval, Quebec City, Québec, Canada

Haematologica 2022  
Volume 107(1):86-99

### 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 *S100A8/9*, 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 subtle changes in the epigenome while co-opting the normal transcriptional machinery to drive leukemogenesis.

### Correspondence:

BRIAN T. WILHELM  
brian.wilhelm@umontreal.ca

Received: September 7, 2020.

Accepted: December 21, 2020.

Pre-published: December 30, 2020.

<https://doi.org/10.3324/haematol.2020.271619>

©2022 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>.

Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>,

sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



### Introduction

Leukemias are a class of blood disorders characterized by the uncontrolled proliferation of hematopoietic stem and progenitor cells (HSPC) that have acquired a block in the normal process of differentiation.<sup>1-3</sup> 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.<sup>4</sup> 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*).<sup>5</sup> 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<sup>4</sup> 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 the mixed lineage leukemia [*MLL*] gene). *KMT2A* is rearranged in approximately 10% of all leukemias<sup>6</sup> but the frequency in infant acute lymphocytic leukemia (ALL, >70%)<sup>7</sup> and AML (>35%)<sup>7</sup> is much higher. The *KMT2A* gene encodes a large (500 kDa) and complex 38-exon protein<sup>8,9</sup> that can be cleaved by caspase-1 into two separate fragments (*KMT2A-C* and *KMT2A-N*).<sup>10</sup> 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<sup>11</sup> 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.<sup>6</sup> One of the most common fusion partners is the *MLL2* gene (also known as *AF9*), which is found in 30% and 13% of *KMT2A*-rearranged AML and ALL,<sup>12</sup> respectively; this fusion is associated with an intermediate risk. The *KMT2A-MLL2* (KM3) fusion protein is part of the DOT1L complex (DOTCOM) and leads to aberrant expression of specific target genes marked by H3K79 methylation.<sup>13</sup> 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.<sup>14</sup> 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.<sup>15</sup> This adaptation of a prior human model<sup>16</sup> uses a retrovirus to deliver the human KM3 fusion gene which we have shown is sufficient to generate a human leukemia.<sup>15</sup> 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,<sup>15</sup> 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.<sup>17</sup> 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/S100A9* expression levels, whose relative ratios have been demonstrated to be critical for blocking differentiation of leukemic cells.<sup>18</sup> Collectively, these results suggest that the leukemic transformation of normal HSPC by KM3 involves a very subtle epigenetic shift that also implicates co-option of normal transcriptional networks.

## Methods

### 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. The characteristics of the patients' AML samples are shown in *Online Supplementary Table S1*. The model leukemias used in this study were generated as previously described.<sup>15</sup>

### 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 Quebec and Hotel Dieu de Levis. CD34<sup>+</sup> cord blood cells were positively selected as previously described<sup>15</sup> 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 min 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 min and resuspended in PBS + citrate 0.6% + bovine serum albumin 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 Posseld 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 cell 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.<sup>15</sup> The shRNA sequences used are shown in Table 1.

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

Cells were cross-linked using formaldehyde 1% (for 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<sup>™</sup>) 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 h 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 pH 8.0, 1 mM EDTA) and incubated overnight at 65°C with shaking (1,200 rpm) to reverse-crosslink. An RNaseA was added (for 1 h 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 (75 bp paired-end) according to standard Illumina protocols. For all RNA-sequencing experiments, RNA was purified from cells (1x10<sup>6</sup>) 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.*<sup>19</sup> and was performed using 50,000 cells. Samples were sequenced (75 bp paired-end) according to Illumina protocols. The primers used are shown in Table 2.

### DNA methylation capture sequencing

Genomic DNA from cell 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

**Table 1. List of primers used for shRNA-mediated knockdown of ADCY9 expression.**

Target	Sequence (5'→3')
Scramble	TGCTGTTGACAGTGAGCGCCCGCCTGAAGTCTCTGATTAATAGTGAAGCCACAGATGTATTAATCAGAGACTTCAGGCGGTTGCCTACTGCCTCGGA
ADCY9 – shRNA1	TGCTGTTGACAGTGAGCGCGACTGTCAAAACCTTTGATAATAGTGAAGCCACAGATGTATTATCAAAGTTTTGACAGCTTCCTACTGCCTCGGA
ADCY9 – shRNA2	TGCTGTTGACAGTGAGCGCGGTTATTTGACTTTTAGATAGTGAAGCCACAGATGTATCTAAAAGTCAAATAATACCCATGCCTACTGCCTCGGA

**Table 2. List of primers used to prepare assays for transposase-accessible chromatin (ATAC)-sequencing libraries.**

Primer name	Primer sequences (5'→3')
Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGCGGA	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGCTCTCGTGGGCTCGGAGATGT

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 were 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.

### Bioinformatics analysis of ChIP-seq and ATAC-seq

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-seq 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)<sup>20</sup> 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 of -1,000 to +3,000 bp for H3K79me2, and -2,000 to +2,000 bp for H3K4me3. MACS2 (version 2.2.4)<sup>21</sup> 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<sup>22</sup> 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`<sup>23</sup> was used to visualize global repartition of mapped reads across the genome. Previously published data for lineage-specific ATAC-seq in a range of normal blood cells<sup>24</sup> were 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 hypersensitive sites and transcription factor clusters generated from the ENCODE consortium V3 were analyzed by custom Perl Scripts. Public ATAC-seq 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-seq data (accessions EGAD00001002710 and EGAD00001002709) were downloaded, remapped and analyzed with MACS2 as described above. Gene set enrichment analysis was performed using software (version 3.0) from the Broad Institute<sup>25</sup> and a pre-ranked list based on significantly differentially expressed genes between CD34<sup>+</sup> and CD34+KM3 cells using the R package DESeq2.<sup>26</sup>

### Mapping methyl-seq 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 previous-

ly described Perl scripts.<sup>27</sup> 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 differentially 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.<sup>28</sup> To define the DMC, we specified two parameters: *q*-values below 0.01 and a threshold of percentage methylation difference ≤40%. All functional annotations for genomic regions were defined based on GRCh37/hg19 assembly downloaded from the 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 transcription start site for each RefSeq transcript.

## Results

### Transcriptional consequences of KM3 expression

To characterize the short-term transcriptional changes induced by the KM3 fusion, we analyzed the RNA-seq data from the initial unmanipulated cord blood cells and these cells after their transduction with the KM3 fusion. As previously described,<sup>16</sup> 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 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., *MN1*<sup>30</sup>). This observation is supported by expression data from patients with *KMT2A*-translocated AML in the TARGET study,<sup>4</sup> which showed an equivalent loss of *MN1* and *GPR56*<sup>4</sup> (Figure 1B) and a similar loss of expression in studies of *GPR56* seen in adult AML.<sup>31</sup> 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 (Online Supplementary Figure S1).

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,<sup>32</sup> MAGICTRICKS,<sup>33</sup> and ShinyGO<sup>34</sup>) produced differing lists of potential transcription factors (Figure 1C), but factors in the CEBP and FOS families, along with NFE2, were highlighted consistently, at differing ranks. Analysis of RNA-seq data confirmed that the expression of all members of the CEBP and FOS families increased during *in vitro* culture (Figure

1D), along with transient changes in ETS, MAF, and JUN families (*Online Supplementary Figure S2*). We also observed an increase in the expression of *S100A9* and *S100A8* (Figure 1E), which have been described as direct targets of CEBPβ.<sup>35</sup> It has been recently demonstrated that the ratio of these calcium-binding proteins is critical for blocking AML differentiation,<sup>18</sup> 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* increased 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 CEBPβ, have also previously been implicated in the process of normal myeloid differentiation,<sup>36</sup> which suggests that transcription factors normally present are co-opted to activate genes that are part of a leukemic tran-

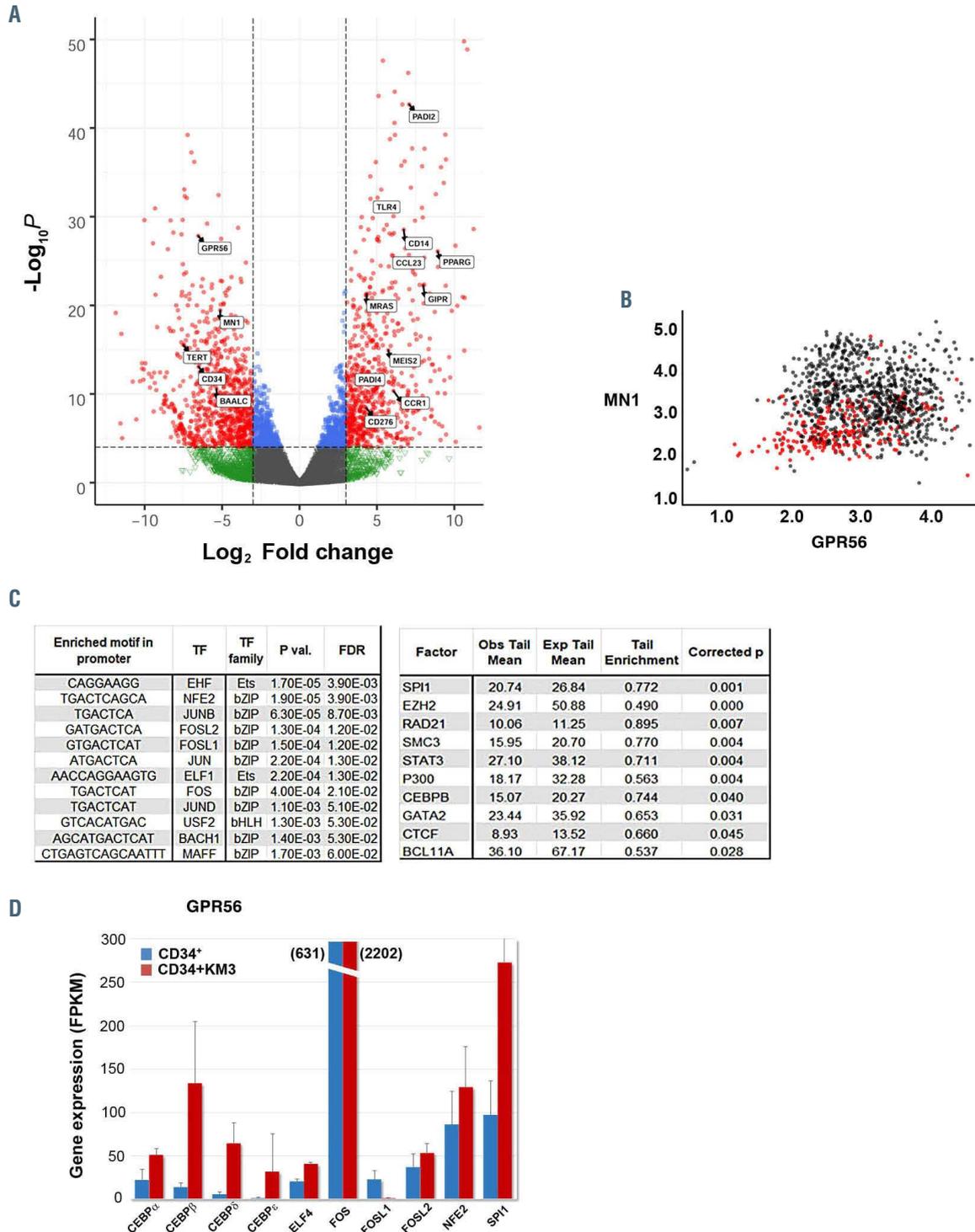
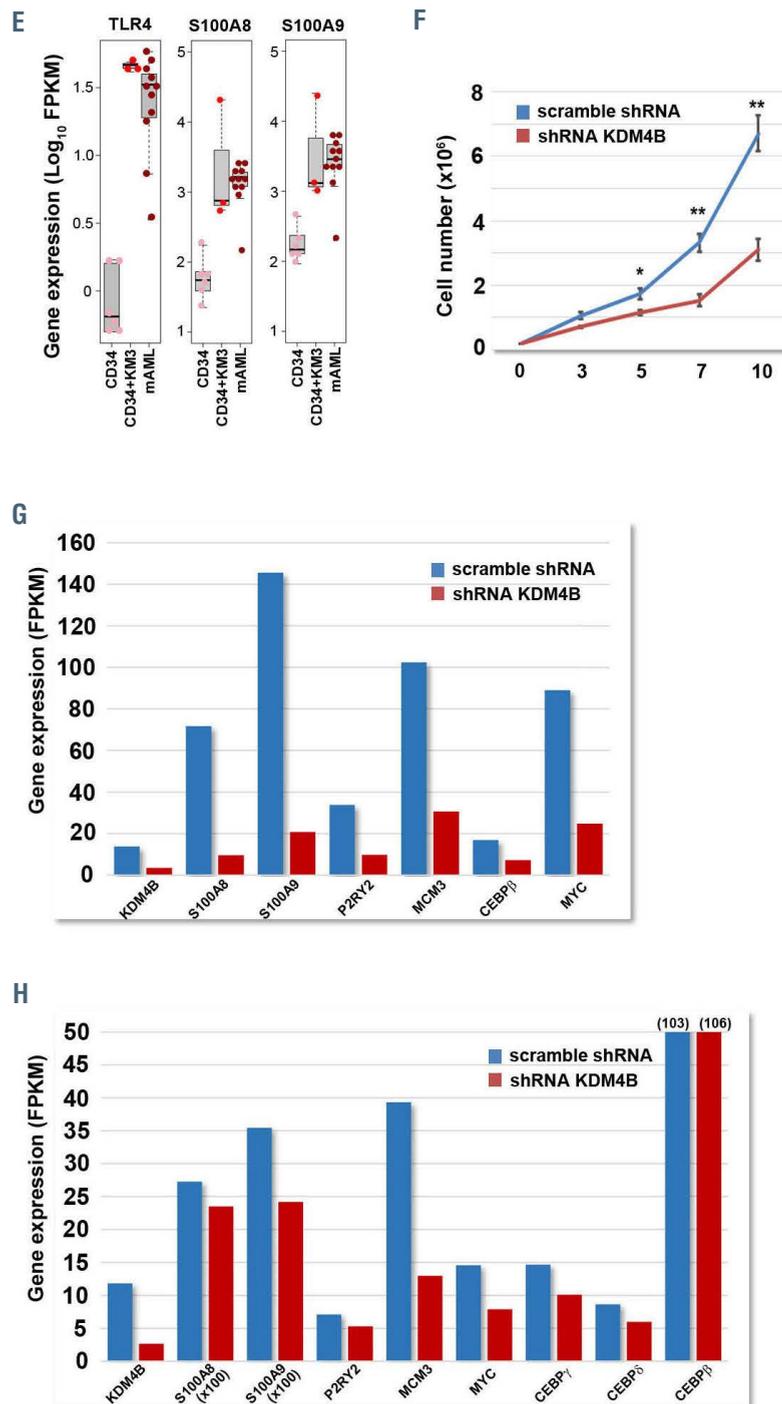


Figure 1. Continued on following page.



**Figure 1.** Gene expression changes in the *KMT2A-MLL3* leukemia model. (A) Volcano plot of differentially expressed genes (red dots are genes with a  $P$  value  $<10^{-5}$ , and a log<sub>2</sub> fold change  $>3$ ) from RNA-sequencing data of CD34<sup>+</sup> 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 *MN1* and *GPR56* expression 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<sup>34</sup> or MAGICTRICKS.<sup>33</sup> (D) Gene expression levels of transcription factor families highlighted in (C) in either CD34<sup>+</sup> 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*, *CEBPβ* and other cell-cycle related genes. (H) shRNA knockdown of *KDM4B* in CD34<sup>+</sup> cells transduced with the KM3 fusion show gene expression decreases very similar to those in THP-1 cells. FPKM: fragments per kilobase of exon per million mapped reads

scriptional 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 (Online Supplementary Table S2). This analysis highlighted a single gene with this profile, *KDM4B*, which has been implicated in murine AML, albeit with functional redundancy.<sup>37</sup> The shRNA-mediated

knockdown of *KDM4B* in THP-1 cells showed a significant impact on growth (Figure 1F) and RNA-sequencing from these cells revealed a profound downregulation of *S100A8* and *S100A9*, *CEBPβ*, *P2RY2* (a KM3-AML biomarker<sup>17</sup>), and other cell cycle regulators (Figure 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. The data showed very similar gene expression changes in these genes with the exception of *CEBPβ*, although other CEBP family members were downregulated (Figure 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.

### 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,<sup>38,39</sup> 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,<sup>40</sup> we assessed DNA methylation levels genome-wide at 84 Mbp 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,<sup>15</sup> the methylation data between replicates were highly consistent (*Online Supplementary Figure S3*) and were also consistent with published data regarding global methylation levels (*Online Supplementary Figure S4*).<sup>41</sup>

Hierarchical clustering of DMC (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 2A). 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,<sup>15</sup> 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,<sup>41</sup> DMC observed in both model and patient KM3-AML were profoundly biased ( $\sim 89\%$ ) towards demethylation (Figure 2B). In opposition to this trend, model leukemias generated from the same donor cells but which developed a lymphoid phenotype (B-cell ALL) had primarily hyper-methylated 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 (*Online Supplementary Figures S5 and S6*).

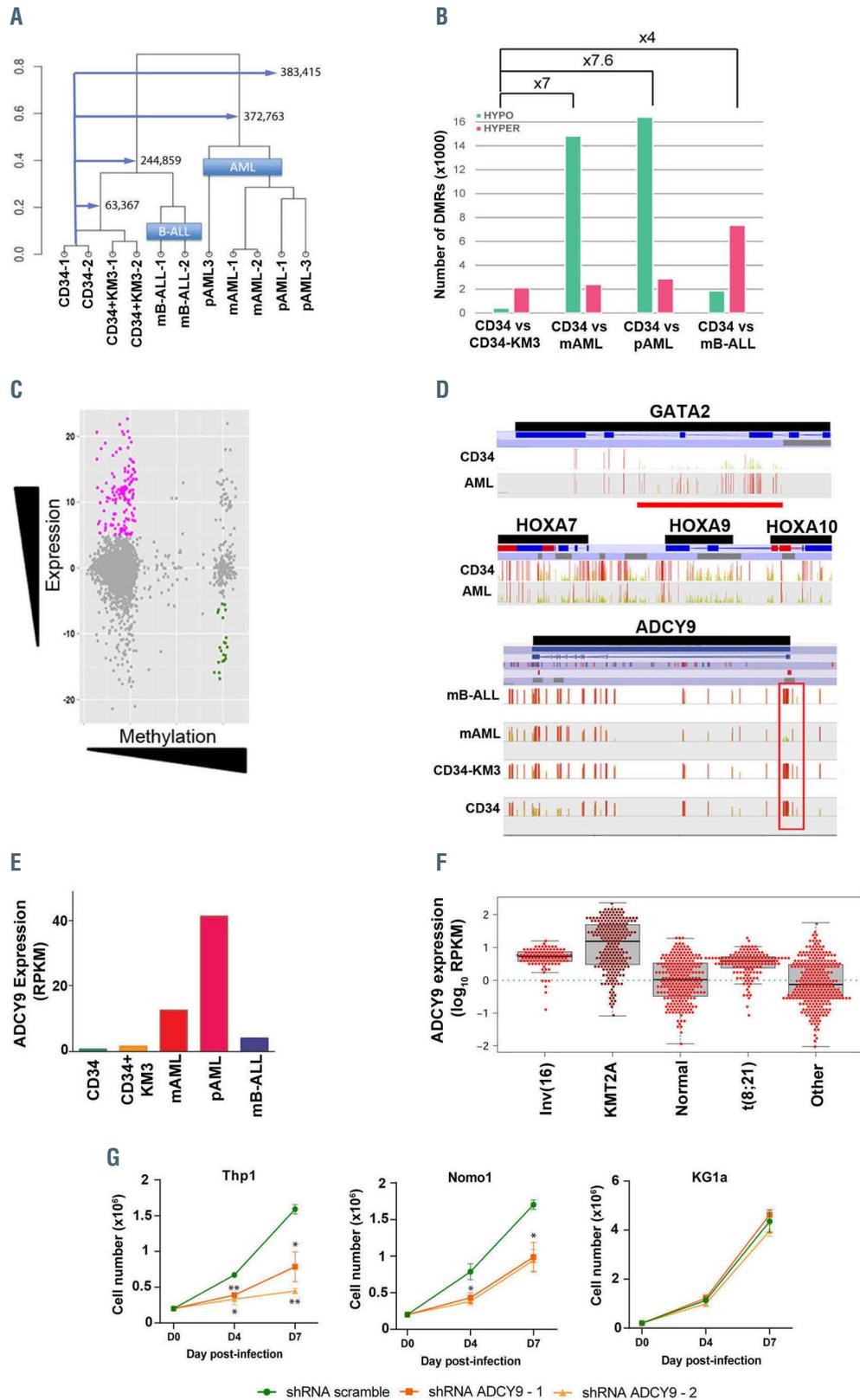
To clarify the importance of the localization of these changes and their potential impact, we next examined the correlation between DNA methylation changes and gene expression. Despite the general view regarding the impact of DNA methylation changes in the promoter regions and transcription levels, we observed only a weak global correlation between the two (Figure 2C). While we could find examples of expected changes (e.g., in previously described *KMT2A* target genes such as *GATA2*,

*HOXA7* and *HOXA9*<sup>42,43</sup>) (Figure 2D), 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<sup>44</sup> as well as previous bulk studies.<sup>45</sup> 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 2E) and the gene is only expressed in the *KMT2A*-rearranged model and patients' leukemias (Figure 2F). *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 was recently identified as a fusion partner of the *KMT2D* gene in a case of ALL.<sup>46</sup> We therefore performed shRNA knock-downs 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 2G), 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+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<sup>47</sup>) based on shared gene ontology terms (*Online Supplementary Figure S7*). 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. The 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.

### 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<sup>11</sup>) 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 (*Online Supplementary Figure S7*) and expected spatial distribution, and the majority of peaks seen in the model leukemias were common to both the initial and transduced cells (Figure 3A). 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<sup>20</sup> we identified a set of genes ( $n=331$ ) (*Online Supplementary Table S2*) with coordinated epige-



**Figure 2. DNA methylation changes in model leukemias.** (A) Hierarchical clustering of differentially methylated cytosines (DMC) identified by methyl-sequencing 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).

netic and expression changes (Figure 3B). 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. *CCR1* is expressed in different types of blood cells, can bind a number of different ligands

and is induced during normal monocyte development.<sup>48</sup> 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.<sup>15</sup> RNA-sequencing data from our models and genetically

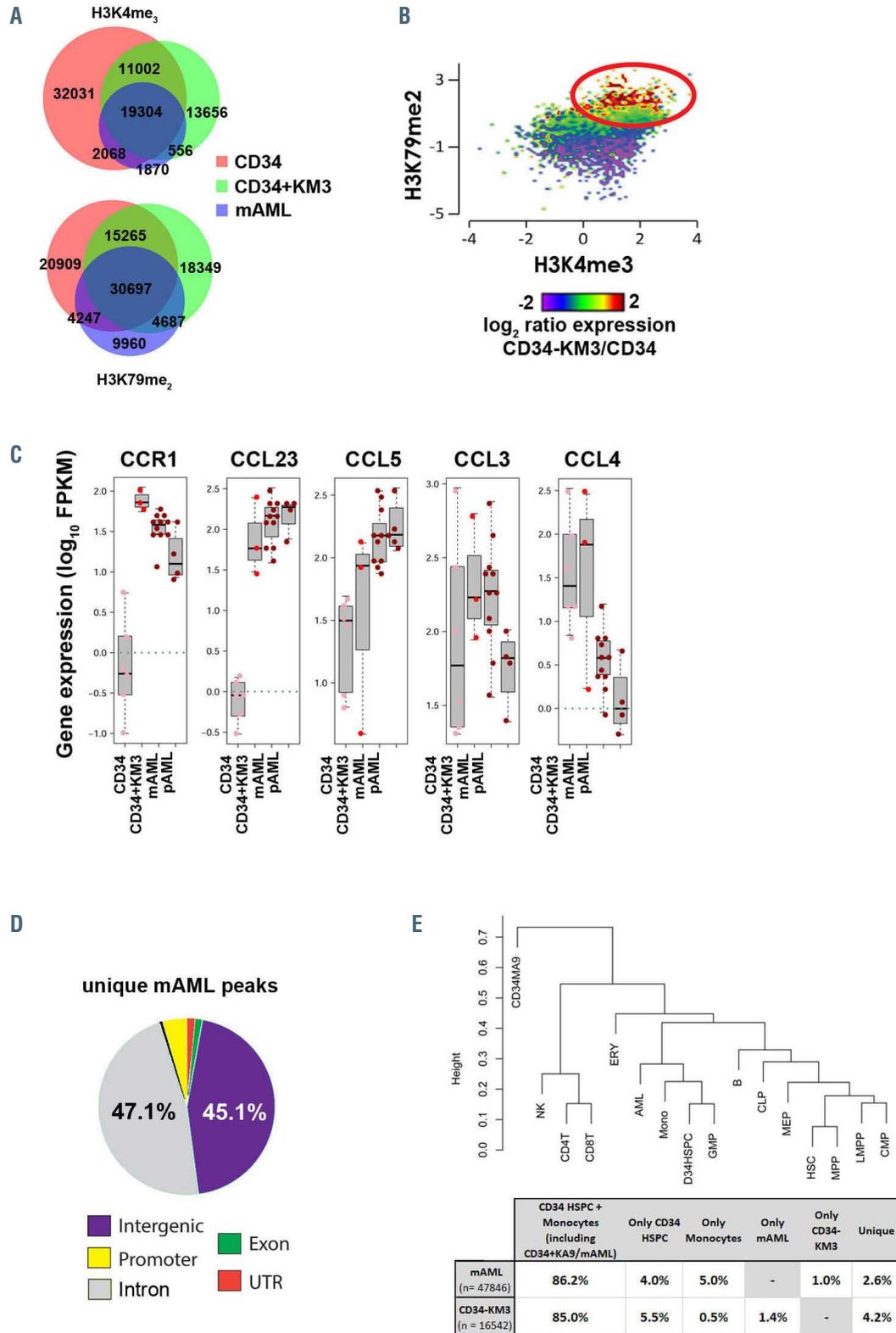


Figure 3. Legend on following page.

F

Factor	Sites	Rank in mAML		CD34	CD34+KM3	mAML
		peaks	Rank in total			
CTCF	122	1	1	19.6	18.6	17.6
FOS	111	2	2	631.2	2202.3	2135.4
JUND	80	3	5	575.0	127.1	642.1
EP300	74	4	3	28.4	22.7	28.8
SP1	69	5	12	97.4	273.2	278.7
IKZF1	61	6	11	57.8	55.1	60.7
SP1	56	7	6	18.8	30.7	26.2
IKZF2	51	8	55	11.4	18.2	0.3
ATF2	50	9	26	16.2	11.8	11.9
RUNX3	49	10	24	48.1	24.3	13.8
FOXA1	44	11	7	0.0	0.0	0.1
POLR2A	44	12	9	36.8	48.1	59.1
CEBPB	41	13	8	14.2	133.5	68.7
HNF4A	40	14	10	0.0	0.0	0.3
DPF2	39	15	22	31.5	28.6	18.9
MTA2	38	16	51	75.1	42.9	68.9
RXRA	38	17	16	12.0	38.3	27.1
RELB	37	18	88	26.1	12.3	17.9
NFIC	35	19	39	9.3	10.1	11.5
TCF12	34	20	25	22.7	19.4	18.7
BATF	32	21	72	17.2	17.2	11.3
FOXA2	32	22	20	0.0	0.0	0.0
STAT3	30	23	36	54.0	53.7	27.7
TBX21	29	24	97	2.1	0.5	0.0
EED	28	25	83	10.9	14.7	8.7
EZH2	28	26	102	20.1	9.3	17.2
FOSL2	28	27	45	37.2	53.3	223.5
GATA3	27	28	23	4.7	0.1	1.4
JUNB	27	29	80	365.9	618.0	417.8
TRIM28	27	30	43	484.5	144.9	293.9

G

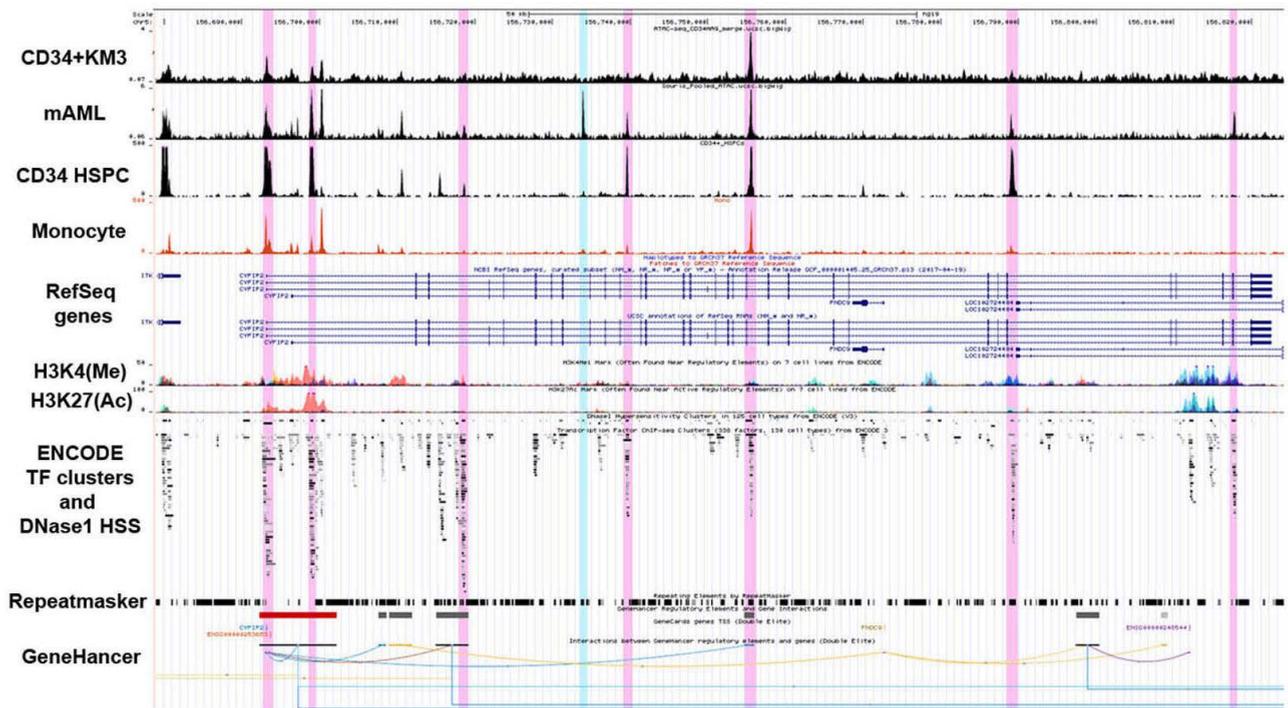


Figure 3. Epigenetic analysis changes in *KMT2A-MLL3* acute myeloid leukemia. (A) Venn diagram of peaks determined by MACS2 in different stages of the model system (CD34<sup>+</sup>, CD34+KM3, KM3 AML model) for H3K4Me3 and H3K79Me2 (n=3 for chromatin immunoprecipitation [ChIP]-sequencing of each mark). (B) A scatterplot of ChIP-seq enrichments for each histone mark in an area ± 2,000 bp (H3K4me3) or -1,000 to +3000 bp (H3K79me2) around the transcription start site for each gene is shown, with individual genes colored by the log<sub>2</sub> ratio (CD34<sup>+</sup>/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-seq peaks unique to model leukemias (mAML) show that the majority are either in intronic or intergenic regions. (E) Hierarchical clustering of ATAC-seq peaks in published data from different normal blood cell types compared to leukemic cells. NK: natural killer cells; CD4T: CD4<sup>+</sup> T lymphocytes; CD8T: CD8<sup>+</sup> T lymphocytes; ERY: erythrocytes; AML: acute myeloid leukemia (model); Mono: monocytes; CD34<sup>+</sup> HSPC: CD34<sup>+</sup> 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 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-seq peaks in CD34+KM3, mAML, CD34<sup>+</sup> 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.

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 3C). External RNA-sequencing data from the BloodSpot database (*Online Supplementary Figure S9A*) or TARGET cohort<sup>4</sup> also confirm the *KMT2A*-specific expression of *CCL23* (*Online Supplementary Figure S9B*) and there is a small but significant difference in survival of pediatric AML patients with high levels of *CCL23* (*Online Supplementary Figure S9C*). Although previous studies have highlighted the importance of *CCR1* in different cancers (e.g., breast<sup>49</sup> and prostate<sup>50</sup> cancer) 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) of the inhibitor.<sup>51</sup> 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.

#### **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 3D). As a point of comparison for our ATAC-sequencing data, we downloaded previously published and normalized ATAC-sequencing data generated from normal blood cells<sup>24</sup> along with raw ATAC-sequencing data from normal CD34<sup>+</sup> cells and monocytes from the BLUEPRINT project.<sup>52</sup> Using peaks identified via 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 (*Online Supplementary Figure S10*), analysis of potential transcription factor binding motifs once again identified factors, such as PU.1, that are critical for normal myeloid cell differentiation (*Online Supplementary Figure S11*). 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 3E). The clear separation of CD34+KM3 cells, despite the small absolute number of unique peaks (and more peaks common to CD34<sup>+</sup> HSPC than to monocytes), suggests that the chromatin of these cells is subtly affected by their culture *in vitro*.

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 3F). 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<sup>53</sup>), 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 3G). 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.

## **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<sup>13,54,55</sup> and DNA methylation seen in the disease.<sup>56</sup> 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<sup>57</sup> showing that an aberrant hypomethylation profile is found in KM3 leukemias. Interestingly, relative to the initial CD34<sup>+</sup> cells, CD34+KM3 and B-cell ALL have a hypermethylation phenotype, whereas the model leukemia 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.<sup>44</sup> Our data (e.g., *Online Supplementary Figure S7*) 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.<sup>58</sup> Forced expression of DNMT3B delayed KM3 leukemia development<sup>59</sup> whereas the deletion of *DNMT1* prevented KM3 AML progression.<sup>60</sup> 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 blood (of all lineages<sup>61</sup>) (Online Supplementary Figure S12), and is capable of inducing cAMP accumulation,<sup>62</sup> 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<sup>63,64</sup> 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 were 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<sup>65</sup> (Lin<sup>-</sup>FSC<sup>hi</sup>SSC<sup>int</sup>CD34<sup>-</sup>CD15<sup>int</sup>CD49d<sup>hi</sup>CD33<sup>hi</sup>CD11b<sup>-</sup>CD16<sup>-</sup>) and granulocyte-monocyte progenitors<sup>65</sup> (Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD123<sup>+</sup>) 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 G $\alpha$  pathway resulting in the inhibition of cAMP production.<sup>66</sup> In this context, it is interesting to note that *ADCY9* is specifically upregulated in *KMT2A*-AML (Figure 2F), potentially as counterbalance to the activity of *CCR1* and G $\alpha$  and to maintain cAMP levels.

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.<sup>67</sup> 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 that 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 observed, largely in introns and intergenic regions (Figure 3D), also raises the question of the potential biological role as novel enhancers. Similar distributions in ATAC-sequencing data have been noted<sup>68</sup> and intronic enhancers have been shown to regulate the expression of genes in a tissue-specific manner,<sup>69</sup> 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 genetic 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.

#### Disclosures

No conflicts of interest to disclose.

#### Contributions

BTW and TM designed the study, analyzed data and drafted the manuscript. TM and KL performed ChIP-sequencing and ATAC-sequencing, analyzed the data and generated figures. MC performed methyl-sequencing experiments, analyzed data, generated figures and drafted the manuscript. SS-T-H assisted with the ATAC-sequencing analysis and generation of figures. KL, ER, EB, and AB helped with molecular biology work, cloning and cell cultures. EB, AB and FB generated the model on mice from CD34<sup>+</sup> cells. SM provided expression and clinical data for patients' samples and JH and SC collected, characterized, and banked the local patients' samples used in this study. All authors discussed the results and participated in the revision of the manuscript.

#### Acknowledgments

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 their expert technical assistance with experimental work.

#### Funding

This work was supported by grants from the Cole Foundation (to 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 Québec - Santé, Faculté de Médecine, Université de Montréal, Leteo 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; to 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, supported by the Cancer

Research Network of the Fonds de Recherche du Québec en Santé.

#### 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.

## References

- Andersson AK, Ma J, Wang J, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet.* 2015;47(4):330-337.
- Saultz JN, Garzon R. Acute myeloid leukemia: a concise review. *J Clin Med.* 2016;5(3):33.
- De Kouchkovsky I, Abdul-Hay M. Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood Cancer J.* 2016;6(7):e441.
- Bolouri H, Farrar JE, Triche T Jr, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med.* 2018;24(1):103-112.
- Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-2074.
- Winters AC, Bernt KM. MLL-Rearranged leukemias-an update on science and clinical approaches. *Front Pediatr.* 2017;5:4.
- Sam TN, Kersey JH, Linabery AM, et al. 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.* 2012;58(6):836-839.
- Gu Y, Nakamura T, Alder H, et al. 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.* 1992;71(4):701-708.
- Tkachuk DC, Kohler S, Cleary ML. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell.* 1992;71(4):691-700.
- Hsieh JJ, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol.* 2003;23(1):186-194.
- Slany RK. The molecular biology of mixed lineage leukemia. *Haematologica.* 2009;94(7):984-993.
- Meyer C, Burmeister T, Groger D, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia.* 2018;32(2):273-284.
- Bernt KM, Zhu N, Sinha AU, et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell.* 2011;20(1):66-78.
- Milne TA. Mouse models of MLL leukemia: recapitulating the human disease. *Blood.* 2017;129(16):2217-2223.
- Barabe F, Gil L, Celton M, et al. Modeling human MLL-AF9 translocated acute myeloid leukemia from single donors reveals RET as a potential therapeutic target. *Leukemia.* 2017;31(5):1166-1176.
- Barabe F, Kennedy JA, Hope KJ, Dick JE. Modeling the initiation and progression of human acute leukemia in mice. *Science.* 2007;316(5824):600-604.
- Lagace K, Barabe F, Hebert J, Cellot S, Wilhelm BT. Identification of novel biomarkers for MLL-translocated acute myeloid leukemia. *Exp Hematol.* 2017;56:58-63.
- Laouedj M, Tardif MR, Gil L, et al. S100A9 induces differentiation of acute myeloid leukemia cells through TLR4. *Blood.* 2017;129(14):1980-1990.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol.* 2015;109:21.
- Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol.* 2016;23(4):349-357.
- Thomas R, Thomas S, Holloway AK, Pollard KS. Features that define the best ChIP-seq peak calling algorithms. *Brief Bioinform.* 2017;18(3):441-450.
- Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics.* 2017;18(1):287.
- Shen L, Shao N, Liu X, Nestler E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics.* 2014;15:284.
- Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016;48(10):1193-1203.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-15550.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571-1572.
- Wang HQ, Tuominen LK, Tsai CJ. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics.* 2011;27(2):225-231.
- Pennacchio LA, Ahituv N, Moses AM, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature.* 2006;444(7118):499-502.
- Sharma A, Jyotsana N, Gabdoulline R, et al. MN1 is indispensable for MLL-rearranged acute myeloid leukemia. *Haematologica.* 2020;105(5):1294-1305.
- Pabst C, Bergeron A, Lavalley VP, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood.* 2016;127(16):2018-2027.
- Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576-589.
- Roopra A. MAGICTRICKS: a tool for predicting transcription factors and cofactors that drive gene lists. *bioRxiv.* 2019:492744.
- Ge SX, Jung D. ShinyGO: a graphical enrichment tool for ani-mals and plants. *bioRxiv.* 2018:315150.
- Endoh Y, Chung YM, Clark IA, Geczy CL, Hsu K. IL-10-dependent S100A8 gene induction in monocytes/macrophages by double-stranded RNA. *J Immunol.* 2009;182(4):2258-2268.
- Tsukada J, Yoshida Y, Kominato Y, Auron PE. The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. *Cytokine.* 2011;54(1):6-19.
- Agger K, Miyagi S, Pedersen MT, et al. Jmjd2/Kdm4 demethylases are required for expression of Il3ra and survival of acute myeloid leukemia cells. *Genes Dev.* 2016;30(11):1278-1288.
- Sandoval JE, Huang YH, Muise A, Goodell MA, Reich NO. 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.* 2019;294(13):4898-4910.
- Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* 2010;363(25):2424-2433.
- Lee EJ, Luo J, Wilson JM, Shi H. Analyzing the cancer methylome through targeted bisulfite sequencing. *Cancer Lett.* 2013;340(2):171-178.
- Akalin A, Garrett-Bakelman FE, Kormaksson M, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* 2012;8(6):e1002781.
- Celton M, Forest A, Gosse G, et al. Epigenetic regulation of GATA2 and its impact on normal karyotype acute myeloid leukemia. *Leukemia.* 2014;28(8):1617-1626.
- Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun.* 2015;6:8489.
- Farlik M, Halbritter F, Müller F, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell.* 2016;19(6):808-822.
- Hodges E, Molaro A, Dos Santos CO, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell.* 2011;44(1):17-28.
- Mankuzhy NP, Walling E, Anderson B, Mody R. Cryptic ETV6-ABL1 fusion and MLL2 truncation revealed by integrative clinical sequencing in multiply relapsed acute lymphoblastic leukemia. *J Pediatr Hematol Oncol.* 2019;41(8):653-656.
- Pomaznoy M, Ha B, Peters B. GOnet: a tool

- for interactive gene ontology analysis. *BMC Bioinformatics*. 2018;19(1):470.
48. Schaller MA, Kallal LE, Lukacs NW. A key role for CC chemokine receptor 1 in T-cell-mediated respiratory inflammation. *Am J Pathol*. 2008;172(2):386-394.
  49. Shin SY, Lee DH, Lee J, et al. C-C motif chemokine receptor 1 (CCR1) is a target of the EGF-AKT-mTOR-STAT3 signaling axis in breast cancer cells. *Oncotarget*. 2017;8(55):94591-94605.
  50. Kato T, Fujita Y, Nakane K, et al. 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*. 2013;64(1):251-257.
  51. Naya A, Sagara Y, Ohwaki K, et al. Design, synthesis, and discovery of a novel CCR1 antagonist. *J Med Chem*. 2001;44(9):1429-1435.
  52. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell*. 2016;167(7): 1897.
  53. Assi SA, Imperato MR, Coleman DJL, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*. 2019;51(1):151-162.
  54. Garcia-Cuellar MP, Buttner C, Bartenhagen C, Dugas M, Slany RK. Leukemogenic MLL-ENL fusions induce alternative chromatin states to drive a functionally dichotomous group of target genes. *Cell Rep*. 2016;15(2): 310-322.
  55. Guenther MG, Lawton LN, Rozovskaia T, et al. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev*. 2008;22(24):3403-3408.
  56. Erfurth FE, Popovic R, Grembecka J, et al. MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. *Proc Natl Acad Sci U S A*. 2008;105(21):7517-7522.
  57. Alvarez S, Suela J, Valencia A, et al. DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PLoS One*. 2010;5(8): e12197.
  58. Zheng Y, Zhang H, Wang Y, et al. Loss of Dnmt3b accelerates MLL-AF9 leukemia progression. *Leukemia*. 2016;30(12):2373-2384.
  59. Schulze I, Rohde C, Scheller-Wendorff M, et al. Increased DNA methylation of Dnmt3b targets impairs leukemogenesis. *Blood*. 2016;127(12):1575-1586.
  60. Trowbridge JJ, Sinha AU, Zhu N, et al. Haploinsufficiency of Dnmt1 impairs leukemia stem cell function through derepression of bivalent chromatin domains. *Genes Dev*. 2012;26(4):344-349.
  61. Hay SB, Ferchen K, Chetal K, Grimes HL, Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp Hematol*. 2018;68:51-61.
  62. Cumbay MG, Watts VJ. Novel regulatory properties of human type 9 adenylate cyclase. *J Pharmacol Exp Ther*. 2004;310(1):108-115.
  63. Perez DR, Smagley Y, Garcia M, et al. Cyclic AMP efflux inhibitors as potential therapeutic agents for leukemia. *Oncotarget*. 2016;7(23):33960-33982.
  64. Gausdal G, Wergeland A, Skavland J, et al. Cyclic AMP can promote APL progression and protect myeloid leukemia cells against anthracycline-induced apoptosis. *Cell Death Dis*. 2013;4(2):e516.
  65. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res*. 2019;47(D1):D881-D885.
  66. Gilliland CT, Salanga CL, Kawamura T, Trejo J, Handel TM. The chemokine receptor CCR1 is constitutively active, which leads to G protein-independent, beta-arrestin-mediated internalization. *J Biol Chem*. 2013;288(45):32194-32210.
  67. Chen X, Burkhardt DB, Hartman AA, et al. MLL-AF9 initiates transformation from fast-proliferating myeloid progenitors. *Nat Commun*. 2019;10(1):5767.
  68. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol*. 2020;21(1):22.
  69. Borsari B, Villegas-Mirón P, Laayouni H, et al. Intronic enhancers regulate the expression of genes involved in tissue-specific functions and homeostasis. *bioRxiv*. 2020 Aug 22. doi: 10.1101/2020.08.21.260836 [preprint, not peer-reviewed]