Modifications of H3K4 methylation levels are associated with DNA hypermethylation in Acute Myeloid Leukemia

Stefania Scalea[1], Carmen Maresca[2] , Caterina Catalanotto[3], Rachele Marino[3], Carlo Cogoni[3], Anna Reale[1], Michele Zampieri[1] and Giuseppe Zardo[1]

1 Department of Experimental Medicine University of Rome "Sapienza" Italy: stefania.scalea@uniroma1.it;anna.reale@uniroma1.it; michele.zampieri@uniroma1.it; giuseppe.zardo@uniroma1.it

2 Oncogenomic and Epigenetic Unit, Regina Elena National Cancer Institute, Rome, Italy: carmen.maresca@ifo.gov.it

3 Department of Molecular Medicine University of Rome "Sapienza" Italy caterina.catalanotto@uniroma1.it; rachele.marino@uniroma1.it; carlo.cogoni@uniroma1.it

Correspondence: giuseppe.zardo@uniroma1.it

Running Title: H3K4me3 modifications and DNA methylation in AML

Keywords: DNA methylation; histone modifications; Acute Myeloid Leukemia

Abbreviations: H3K4me3, Histone 3 Lysine 4 trimethylated; H3K27me3, Histone 3 Lysine 27 trimethylated; H3K4me0, Histone 3 Lysine 4 unmethylated; hESC, Human Embryonic Stem Cells; hHSC/MPP, human Hematopoietic Stem or Progenitor cells; AML, Acute Myeloid Leukemia; CpGis, CpG islands; RLGS, Restriction Landmark Genome Scanning; TSS, Transcriptional Start Site; IRX2, Iroquois homeobox 2 gene; CABYR, Calcium-binding tyrosine phosphorylation-regulated gene; NKX6-1, Homeobox Protein NK-6 Homolog A gene; FOXE1, Forkhead Box E1 gene; NRG1, Neuregulin 1 gene; RA, retinoic acid; CSF1, Colony Stimulating Factor 1; CSF2, Colony Stimulating Factor 2; CSF3, Colony Stimulating Factor 3.

Abstract

The "instructive model" of aberrant DNA methylation in human tumors is based on the observation that CpG islands prone to hypermethylation in cancers are embedded in chromatin enriched in H3K27me3 in human embryonic stem cells (hESC). Recent studies also link methylation of CpG islands to the methylation status of H3K4, where H3K4me3 is inversely correlated with DNA methylation. To provide insight into these conflicting findings, we generated DNA methylation profiles for acute myeloid leukemia samples from patients and leukemic cell lines and integrated them with publicly available ChIp-seq data, containing H3K4me3 and H3K27me3 CpG island occupation in hESC, or hematopoietic stem or progenitor cells (hHSC/MPP). Hypermethylated CpG islands in AML samples displayed H3K27me3 enrichments in hESC and hHSC/MPP; however, ChIp analysis of specific hypermethylated CpG islands revealed a significant reduction in H3K4me3 signal with a concomitant increase in H3K4me0 levels as opposed to a nonsignificant increase in H3K27me3 marks. The integration of AML DNA methylation profiles with the ChIp-seq data in hESC and hHSC/MPP also led to the identification of Iroquois homeobox 2 (IRX2) as a previously unknown factor promoting differentiation of leukemic cells. Our results indicate that in contrast to the "instructive model," H3K4me3 levels are strongly associated with DNA methylation patterns in AML and have a role in the regulation of critical genes, such as the putative tumor suppressor *IRX2*.

Introduction

Epigenetic modifications sustain onset and progression of various human pathologies including carcinogenesis. Modification of established DNA methylation patterns is a common feature in all types of tumors, including leukemia, and is considered an event that intrinsically, or in association with genetic lesions, feeds carcinogenesis.

In tumors, DNA hypomethylation mainly affects tandem centromeric satellite α, juxtacentromeric (centromere-adjacent) satellite 2, interspersed Alu and long interspersed elements (LINE)-1 repeats (1). These events promote genomic instability (2-6) and reactivation of transposons (7-8). DNA hypermethylation is also observed at sequences with regulatory functions in gene expression, such as promoter CpG islands (CpGis) and CpG shores other than at enhancers and insulators (9-13). Aberrant DNA hypermethylation of the above mentioned regulatory genomic regions is generally

correlated with the repression of tumor suppressor (14-16), metastasis (17), and DNA repair genes (18-19) leading to the convincement that DNA hypermethylation is directly responsible for the observed gene silencing (20). However, recent experimental data obtained studying the methylome of normal tissues and derived cancer types suggested that aberrant DNA hypermethylation represents a secondary event following gene inactivation, as the majority of aberrantly hypermethylated genes in cancer are already repressed in tissue of origin (21).

Aberrant DNA hypermethylation of gene regulatory regions may be the consequence of the modified activity of DNA methyltransferases (DNMTs) (22, 23), aberrant recruitment of mutated transcription factors (24), mutations in demethylating enzymes, such as ten-eleven-translocation enzymes (TETs) and their associated co-factor pathways (isocitrate dehydrogenase; IDH1/2) (25), and changes in chromatin architecture depending on histone posttranslational modifications.

Specific, global and local, histone modifications are often associated with distinctive DNA methylation patterns (26). Several studies investigating the molecular basis of DNA hypermethylation propose an “instructive mechanism” of aberrant DNA methylation in tumors that relies on the histone modifications characterizing chromatin in embryonic and adult stem cells. Accordingly, CpGis prone to hypermethylation in tumors are embedded in chromatin enriched in H3K27me3-only or H3K27me3 in association with an H3K4me3 mark at the same *locus* (bivalent domain) in human embryonic and adult stem cells (27-31). This instructive model is supported by the observation that EZH1/2, a component of the polycomb repressive complex 2 (PRC2), is responsible for H3K27 methylation and may recruit *de novo* DNA methyltransferases (DNMTs) (32-33). However, in human as well as mouse embryonic stem cells (hESC; mESC), H3K27me3 is mainly located in bivalent domains coinciding with unmethylated CpGis (27, 29, 34-36). This observation might indicate that H3K27me3 and DNA methylation are mutually exclusive. However, several independent studies have observed a causal association between PRC2 recruitment, H3K27me3 and DNA hypermethylation during carcinogenesis (29, 31, 37-40). A possible explanation for this apparent contradiction arises from the observation that TET1 and TET2 have been found to be associated with the PRC2 complex at CpGis in mESC (41) and in cell lines overexpressing TETs (42).

TET enzymes catalyze hydroxylation of 5-methylcytosine and the active demethylation process, thus maintaining the unmethylated state of CpGis.

Recent studies also indicate that methylation of CpGis is related to the methylation status of H3K4; the levels of methylated H3K4 (H3K4me3) tend to be inversely correlated with DNA methylation (43-45). This mutually exclusive nature of the association of H3K4me3 with CpGi methylation might be related to its role in regulating methyltransferase activity. The ATRX-DNMT3-DNMT3L domain (ADD) in the *de novo* DNA methyltransferase (DNMT) DNMT3a, for example, recognizes the unmethylated form of H3K4 (H3K4me0), which stimulates methyltransferase activity (46-47). Through structural and biochemical analyses, the ADD domain of DNMT3a has been shown to also interact with its catalytic domain (CD), but in the presence of H3K4me3, which results in the loss of the ability of DNMT3a to bind and methylate DNA (47). Thus, H3K4me0 and H3K4me3 have opposite effects on DNMT3a activity (48).

However, there is no evidence that an H3K4me3 protective effect against aberrant methylation exists or is lost in tumor models. Here, we generated DNA methylation profiles in acute myeloid leukemia (AML) samples from patients and several AML cell lines. These methylation profiles were integrated with publicly available chromatin immunoprecipitation-sequencing (ChIp-seq) data for H3K4me3 and H3K27me3 promoter CpGis occupation in hESC or hematopoietic stem and/or progenitor cells (hHSC/MPP). We observed that, in most cases, hypermethylated CpGis in AML display H3K27me3 occupancy in hESC and hHSC/MPP. ChIp analyses of specific hypermethylated CpGs in AML samples also showed a significant reduction in H3K4me3 signal with a concomitant increase in H3K4me0 levels as opposed to a nonsignificant increase in the H3K27me3 mark particularly in AML patient samples. The integration of AML DNA methylation profiles with ChIp-seq data in hESC and hHSC/MPP also led to the identification of Iroquois homeobox 2 (IRX2) as a previously unknown factor promoting differentiation in leukemia. Our data suggest that H3K4me3 levels underlie critical gene regulating DNA methylation patterns in AML.

Results

DNA methylation analysis of AML samples by RLGS.

Using restriction landmark genome scanning (RLGS), we analyzed the methylation status of ~ 3000 CpGis in 19 AML samples from 19 patients and 7 AML cell lines, classified according to FAB criteria. The distribution of AML samples and cell lines was as follows: for patient samples, 5 M1, 5 M2, 2 M3, 3 M4, and 4 M5; for cell lines, HL-60, M2; Kasumi-1, M2; NB-4, M3; ML-2, M4; AML-193, M5; MONOMAC-1, M5; HEL, M6. RLGS was performed using the methylation sensitive enzyme Not1 as previously described (49-50). Aberrant methylation of target CpGis was identified comparing the methylation profiles of every single AML sample with a *normal master profile.* The normal *master profile* was generated considering CpGis that were present and concomitantly unmethylated in 1 peripheral blood sample, in 2 CD34+ hHSC/MPP samples and in 2 CD34- mature cell samples co-purified from umbilical cord blood from 2 healthy donors (to diminish the contribution of common restriction-site polymorphism to apparent spot loss). All CpG analyzed nucleotides were included in canonical CpGis that did not include or were part of repetitive sequences. This allowed the analyses of 976 CpGis for patient AMLs and 1188 CpGs for AML cell lines.

284 and 534 CpGis sites were found to be hypermethylated at least once in AML patient samples (Figure 1) and AML cell lines, respectively (data not shown). Although the analysis was performed on a limited number of samples, the percentage of DNA methylation events did not depend on the stage of differentiation block (FAB classification). However, as expected, the percentage of DNA methylation events in AML cell lines was significantly higher than in AML patient samples (data not shown).

Integration of AML DNA methylation data with histone posttranslational modifications detected in hESC and hHSC/MPP.

To explore the possibility of an instructive model of aberrant DNA methylation in AML, we integrated the methylation data of every methylated CpGi with the corresponding histone marks in hESC. We referred to 3 different publicly available databases, containing annotated genome wide H3K4me3 and H3K27me3 enrichments in hESC (51-53). We considered only those hypermethylated CpGis which were reported to be

associated with the same histone mark in all 3 datasets in hESC. This restricted the analyses to 131 loci for AML patient samples (Figure 2) and 223 for AML cell lines (data not shown).

Our analysis revealed that 28,7% of aberrantly methylated CpGis in AML patient samples were marked by a bivalent domain, 30,3% by H3K27me3 and 40,1% by H3K4me3 in hESC (Figure 3 panel A, columns A, B, and C, respectively). We grouped data according to the frequency of hypermethylation, which we defined as the number of AML samples presenting the same methylated CpG island. In this analysis, frequently methylated CpGis were mainly marked by a domain containing H3K27me (only or as a bivalent domain) when compared to hypermethylated CpGis marked by H3K4me3 in hESC (n = 9+7 *vs* n = 4; Figure 3 panel A column B).

Similarly, in AML cell lines, 34,1% of aberrantly methylated CpGis were marked by H3K4me3, 22.9% by H3K27me3 and 43% by a bivalent domain (Figure 3 panel B columns A, B, and C, respectively) in hESC. When data were grouped according to the frequency of hypermethylation, frequently methylated CpGis were mainly marked in hESC by a domain containing H3K27me (only or as a bivalent domain) when compared to hypermethylated CpGis marked by H3K4me3 (Figure 3 panel B columns B and C). Importantly, AML patient samples and also AML cell lines contained a discrete number of hypermethylated sites (n = 11) marked by H3K4me3-only in hESC (Figure3 panel B column C). This finding might be due to a switch from H3K4me3 to H3K27me3 (only or as bivalent domain) at these CpGis during differentiative restriction of hESC to hHSC/MPP, making them sensitive to aberrant hypermethylation in AML.

Histone modifications distinguish the 5' promoter region of *IRX2* in AML samples relative to CD34+ hHSC/MPP and mature CD34- cells.

In order to investigate the possibility that the hypermethylation of DNA sequences marked by the presence of H3K4me3 in hESC might be due to an enrichment in H3K27me3 during differentiative restriction of hESC to hHSC/MPP, we selected 5 target genes marked in hESC by H3K4me3-only or in association with H3K27me3 (as a bivalent domain) among CpGis with the highest frequency of hypermethylation in AML patient samples and cell lines. The following 5 genes were examined: *IRX2, NKX6-1,*

FOXE1, CABYR and *NRG1*. RLGS provides DNA methylation analysis for only one CpG dinucleotide within its restriction sequence. To validate hypermethylation of selected targets, corresponding CpG islands were analyzed using bisulfite sequencing in the 19 AML patient samples. We analyzed the following CpGis:

1. CpG #689 (UCSC Genome Browser GRCh38/hg38) corresponding to *IRX2* gene marked by a bivalent domain in hESC;
2. CpG #242 (UCSC Genome Browser GRCh38/hg38) corresponding to *NKX6-1* gene marked by a bivalent domain in hESC;
3. CpG #253 (UCSC Genome Browser GRCh38/hg38) corresponding to *FOXE1* gene marked by a bivalent domain in hESC;
4. CpG #58 (UCSC Genome Browser GRCh38/hg38) corresponding to *CABYR* gene marked by H3K4me3 only in hESC
5. CpG #169 (UCSC Genome Browser GRCh38/hg38) corresponding to *NRG1* gene marked by H3K4me3 only in hESC.

Bisulfite sequencing revealed that all target CpGis were characterized by heavy and frequent hypermethylation in AML patient samples (Figure 4).

Using chromatin immunoprecipitation (ChIp), we verified H3K4me3 and H3K27me3 enrichments in the 5' promoter regions of our selected target genes in CD34+ hHSC/MPP and in the corresponding CD34- mature cells (containing 70-80% mature granulocytes) purified from cord blood. Bivalent domains are located at gene promoters centered around the canonical TSS over a region within 1-1.5 kb upstream or downstream the TSS itself. In our ChIp experiments, for each target, we amplified a DNA sequence (≈ 250bp long) within a region of -1kb from the TSS (Figure 5). We also analyzed a region at least +2kb away from the TSS (intragenic) where bivalent domains and H3K4me3 and H3K27me3 enrichments are not expected (Figure 5). The 5' promoter regions of *FOXE1* and *NKX6-1* in hHSC/MPP and CD34- mature cells were clearly marked by a bivalent domain as reported for hESC (51-53), and for *CABYR* and *NRG1* by H3K4me3 modification also as in hESC (51-53). The only exception was represented by the 5' promoter region of *IRX2* called as a bivalent domain in hESC (51-53) but marked by H3K4me3-only in hHSC/MPP and $CD34^{-}$ mature cells (Figure 5).

Our results indicated that no substantial change in the histone modification pattern that

could fit in the instructive model of DNA hypermethylation occurred at the promoter regions of these target genes during differentiative restriction from hESC to hHSC/MPP and CD34- mature cells. Moreover, the 5' *IRX2* promoter region revealed a change in histone modifications, from a bivalent domain in hESC to H3K4me3-only in hHSC/MPP and $CD34^{-}$ mature cells. These results contradicted our explanation for the association of the relatively high number of hypermethylated CpGis, in AML samples, with a switch toward H3K27me3 modification during differentiative restriction from hESC to CD34+ hHSC/MPP and to mature CD34- cells.

H3K4me3, H3K27me3 and H3K4me0 enrichments at selected targets in AML samples relative to normal hematopoietic cell types.

Several studies have identified an association between PRC2 activity, H3K27me3 enrichment and DNA hypermethylation in tumors (29, 31, 36-40). To investigate the increased DNA methylation in our selected target genes (*IRX2, NKX6- 1, FOXE1, CABYR* and *NRG1*), we analyzed H3K4me3 and H3K27me3 enrichment in their promoter regions using ChIp. AML patient samples and cell lines were examined and compared to the respective enrichments in CD34+ hHSC/MPP cells and mature CD34- cells. The results from ChIp demonstrated that H3K27me3 levels were increased in target gene promoter regions both in AML patient samples and cell lines when compared to the H3K27me3 levels in CD34+ hHSC/MPP cells and CD34- mature cells (Figure 6). However, nonparametric Wilcoxon Rank Sum Test analysis revealed that H3K27me3 enrichments were not significant in patient samples, except in the case of *IRX2*. In contrast, H3K27me3 levels were significantly increased in all target gene promoters, except for *NKX6-1*, in AML cell lines. Surprisingly, the opposite tendency was observed for H3K4me3 enrichment. H3K4me3 levels in the target genes were significantly decreased in all AML patient samples when compared to H3K4me3 enrichment in CD34+ hHSC/MPP cells and CD34- mature cells. However, such changes were only observed in 2 genes, *CABYR* and *IRX2*, in AML cell lines. Thus, the results indicated that in AML patient samples, H3K4me3 levels on hypermethylated promoter regions of the target genes were significantly decreased whereas H3K27me levels did not change significantly.

These results corroborate several studies indicating that H3K4me3 levels are a critical regulator of DNMTase activity (43-47). H3K4me3 has been shown to bind the ADD domain of *de novo* DNMTs causing the inhibition of their activity. Thus, H3K4me3 would protect underlying DNA sequences from methylation. In contrast, H3K4me0, which is the unmethylated form of H3K4, would not bind to the ADD domain, thus allowing *de novo* DNMT activity and consequently hypermethylation of the underlying DNA sequences. Using ChIp, we therefore investigated H3K4me0 and H3K4me4 levels at the 5' promoter regions of our target genes in AML patient samples and compared them with the enrichment observed in control CD34+ hHSC/MPP and CD34- mature cells. ChIp results revealed that H3K4me0 levels were significantly higher in the AML patient samples than in control CD34+ hHSC/MPP and mature CD34- cells. Moreover, H3K4me0 were also significantly higher when compared to H3K4me3 enrichment within the AML patient cohort (Figure 7). In contrast, changes in H3K4me3 levels were not statistically significant, relative to H3K4me0 enrichment in control CD34+ hHSC/MPP and mature CD34- cells.

In conclusion, our data indicate that CpGi hypermethylation observed for our target genes in AML patient samples might be associated with an increase in H3K4me0 levels, and consequently a decrease in H3K4me3, rather than a significant change in H3K27me3 levels.

Aberrant expression of critical histone methylases and demethylases in AML.

Our data indicated that H3K27me3, H3K4me3 and H3K4me0 enrichments, on specific hypermethylated target genes were significantly modified in AML differentiation block when compared to their levels in normal CD34+ HSC/MPP and mature CD34- cells. These changes might correlate with aberrant function of H3K27 and H3K4 histone methylases and demethylases. We therefore quantified the expression of epigenetic *writers* and *erasers* critical in controlling H3K4me3, H3K27me3 and h3K4me0 levels in normal mature CD34- cells and in AML patient samples and cell lines.

Our data indicated that Jumonji/ARID domain-containing protein 1A (JARID1A) and B (JARID1B), involved in demethylation of H3K4me3 and me2, and lysine-specific histone demethylase 1 (LSD1), involved in H3K4me1 demethylation, were significantly over-

expressed. Enhancer of zeste homolog 2 (EZH2), which underlies H3K27 trimethylation, was also significantly over-expressed, while Jumonji domain-containing protein 3 (JMJD3), which demethylates H3K27me3, was also significantly down-regulated (Figure 8). However, these two last enzyme activities, which facilitate H3K27me3 enrichment, were possibly balanced by a significant increase in expression of *JARID2* in AML patient samples. JARID2 lacks demethylase activity but inhibits trimethylation of H3K27 (54). This result might also explain the apparent discrepancy observed for the significant increase in H3K27me3 levels in AML cell lines where *JARID2* was found to be down-regulated. Therefore, these data support the alteration of H3K27me3, H3K4me3 and H3K4me0 enrichments detected in ChIp analysis of our selected target genes in AML samples.

The occupancy of JARID1B and LSD1 at the 5' promoter region of selected target genes increases significantly in AML cell lines.

In order to give a functional role to the over-expression and down-regulation of the critical writers and erasers affecting the levels of H3K27 and H3K4 methylation in AML samples, ChIp experiments were carried out to determine the presence of EZH2, JMJD3, JARID1B and LSD1 proteins at the 5' promoter region (proximal) and at a control region (distal) of our target genes. These experiments were performed on mature CD34- cells and on the AML cell lines HL60, ML-2 and AML193. Compared to the mature CD34- cells, ChIP (Figure 9) and expression data (Figure 8) from AML cell lines suggest the following key issues: a) the overexpression of the H3K27 methylase EZH2 did not result in any significant increase in its occupancy at the selected target genes, with the exception of IRX2, where EZH2 binding signal was increased; b) the downregulation of the H3K27me3 demethylase JMJD3 correlated to a reduced signal of its occupancy on the CABYR, FOXE1 and NRG1 genes, but not on IRX2 and NKX6-1 genes. c) the overexpression of the H3K4me2/3 demethylase JARID1b and of the H3K4me1 demethylase LSD1 was associated with an increase in their occupancy at the promoter region of all gene targets, with the exception of NRG1 gene. Overall, these data suggest that in AML cell lines the increase of H3K27me3 on specific gene promoters may depends on a reduced expression and recruitment of JMJD3 (as for CABYR, FOXE1 and

NRG1 genes) rather than on an increased expression and binding of EZH2 (as for IRX2 gene). To note, comparable levels of H3K27me3 matched comparable binding signals for both EZH2 and JMJD3 at the NKX6-1 gene promoter between CD34- cells and AML cell lines. On the other hand, the increased expression and binding of JARID1b and LSD1 support and explain the reduction of H3K4me3 and the increase of H3K4me0 signals observed on the promoters of the CABYR, FOXE1, IRX2, NKX6-1 genes, whereas no significant accumulations of JARID1B and LSD1 were observed on NRG1 promoter region.

Ectopic expression of *IRX2* promotes differentiation in AML cells.

CpGis hypermethylation is often associated with transcriptional inhibition of the corresponding gene. qRT-PCR demonstrated that many of the hypermethylated targets (*IRX2*, *NKX6-1*, *FOXE1*, *CABYR* and *NRG1*) in AML patient samples were silenced in most leukemic samples (Figure 10). Loss of expression was particularly prevalent among samples for *IRX2*. Interestingly, IRX2 has been suggested to function in lineage-specific epithelial cell differentiation in the breast tissue, and to act as a tumor suppressor or oncogene depending on the cellular and tissue context (55-62). Therefore, we used infection with lentiviral expression constructs to restore *IRX2* in HL60 and AML-193 cell lines to determine the gene's impact on differentiation block in AML. We also infected AML cells with a normal karyotype obtained from a new patient. All cells were collected after 48 hours, and *IRX2* cDNA expression was quantified and compared with the expression in control cells, parental cells or cells infected with the empty vector. *IRX2* expression was successfully restored in cell lines, as observed in qRT-PCR (Figure 11A).
Morphologic analyses of HL60- and AML-193-IRX2 cell lines compared to control cells, displayed signs of cell maturation such as chromatin condensation with nuclear segmentation, increased granulation and reduced cytosolic basophilia (Figure 11B). These characteristics were also visible in the patient sample AML-IRX2 (Figure 11B). In this case, the blast number was dramatically reduced: 45% blasts in AML- IRX2 *vs* 58% in AML-EV and 73% in AML-uninfected.
We further investigated ectopic expression of IRX2 in HL60 cell lines by analyzing the expression of key cytokines involved in the proliferation and differentiation of

hematopoietic precursor cells (Figure 11C), the percentage of CD11b positive cells (Figure 11D) and the effect of retinoic acid (RA) treatment (Figure 11E). IRX2 expression in HL60 cells (HL60-IRX2) compared with controls (HL60- uninfected and HL60-EVcells), had no effect on the expression of macrophage colony stimulating factor 1 (CSF1) but did promote *CSF2* and *CSF3* expression, which are implicated in the differentiation of granulo-macro and granulocytic cells respectively (Figure 11C). In addition, 4 days after infection, HL60-IRX2 cells exhibited a significant, even if not quantitatively relevant, increase in CD11b positive cells (Figure 11D), and an increased sensitivity to RA particularly in the first 12 hours of treatment (Figure 11E).

Discussion.

Several authors have used genome wide approaches, such as ChIP-seq, to study histone modifications associated with 5' promoter CpGis in hESCs and in tumor cells (27-31; 34-36). The main evidence produced by these studies is that the chromatin domains of promoter CpGis are characterized by a surprising variability in histone modifications. Nonetheless, two main different histone modifications seem to occur in promoter CpGis, H3K4me3 and H3K27me3, which are associated with active and inactive promoters, respectively. However, several promoter CpGis are characterized by the concomitant and partially overlapping presence of H3K4me3 and H3K27me3. These chromatin areas are defined as "bivalent domains" and associated promoters are considered to be maintained in a poised transcriptional *status*. During cell differentiation, the resolution of the bivalent domain in favor of H3K4me3 or H3K27me3 modification, would assure the complete activation or definitive repression of the corresponding gene, in such a way that specific cell differentiation programs could be activated. The study of the same chromatin domains in tumor cell models demonstrated that when CpGi promoters are enriched in H3K27me3-only or in association with H3K4me3, they are very often hypermethylated. This observation has yielded the "instructive model" of aberrant DNA methylation, in which H3K27me3 gene promoters would have a high probability for aberrant hypermethylation during tumor transformation. The integration of our DNA methylation data, obtained using RLGS of AML patient samples and cell lines, with ChIp-seq data of the same sites in hESC, reveals the presence of H3K27me3-only or in association with

H3K4me3 thus confirming the tendency of promoter CpGis to be hypermethylated in the presence of H3K27me3 modification in AML samples.

However, our data also show that CpGis labeled by H3K4me3-only in hESC, are not immune to aberrant methylation in AML.. This event was particularly prominent in AML patient samples. This result might be explained by the instructive model of DNA hypermethylation, if during restrictive differentiation from hESC to HSC/MPP, these promoter CpGis could have acquired the H3K27me3 modification. Our ChIp experiments showed that this was not the case. Our selected hypermethylated targets maintained the same histone modifications observed in hESC in HSC/MPP progenitor cells (CD34+ cells), except for *IRX2*. For *IRX2*, the three datasets used in our analysis were in agreement for a bivalent domain in hESC that appeared to be resolved into mono-label H3K4me3 in hematopoietic progenitors (CD34+ cells). Despite this result, the *IRX2* CpG island appeared highly methylated in AML samples as indicated by our RLGS and bisulfite data. ChIp experiments, analyzing the presence of H3K27me3, H3K4me3 and H3K4me0 on selected hypermethylated promoter CpGis in control HSC/MPP CD34+ and mature CD34- cells, demonstrated that despite nonsignificant enrichment in H3K27me3, we detected a significant decrease in H3K4me3 levels (especially in AML patient samples) and a significant increase in H3K4me0 enrichment in AML samples. We also showed that alterations in the expression of the principal *writers* and *erasers* of H3K4me3 and H3K27me3 modifications are well correlated with their occupancy at the promoter region of our target genes in AML cell lines (Figure 9). These data also agreed with the changes of H3K27me3, H3K4me3 and H3K4me0 enrichments observed at the same regions (Figure 6 and 7). For instance: a non significant occupancy of EZH2 (Figure 9, see CABYR, FOXE1 and NRG1) corresponded to a significant occupancy of JMJD3, thus explaining the significant increase in H3K27me3 levels observed in AML cell lines (Figure 6); on the contrary, EZH2 accumulation was significant at IRX2 promoter region and this was correlated with a non-significant diminution of JMJD3, thus the relative occupancies of these two chromatin effectors might explain the increased levels of H3K27me3 levels at IRX2 promoter region. This suggests that gene promoter H3K27me3 levels in AML cells may depend on the deregulated expression of EZH2 and JMJD3 and on a gene-specific control of their recruitment on chromatin. As regard

NKX6-1 both the accumulation of EZH2 and JMJD3 were not significant if compared to CD34- cells and this seemed in agreement with the levels of H3K27me3 observed in AML cell lines (Figure 6), but, more importantly, the over-expression of H3K4me2/3 demethylase JARID1b and H3K4me1 demethylase LSD1 (Figure 8) resulted in a significant increased occupancy of the promoter region of all targets except for NRG1 (Figure 9), still this was in agreement with the non significant diminution of H3K4me3 diminution in AML cell lines. Overall our data support the results of several studies indicating H3K4me3 and H3K4me0 levels as a critical regulator of DNMTase activity (43-47).

Among our selected target genes, we chose to further examine *IRX2* in functional studies. Although characterized by a bivalent domain in hESC, the *IRX2* promoter acquired an H3K4me3 mark in hHSC/MPP, indicating that its expression and function may contribute to cell lineage specification. In AML samples, the *IRX2* promoter was heavily hypermethylated, and exhibited high levels of H3K4me0 and low levels of H3K4me4. Above all, *IRX2* was not expressed in all AML patient samples. Finally, ectopic *IRX2* expression in 2 AML cell lines and 1 patient sample demonstrated a role for the gene in overcoming the block in hHSC differentiation in the disease. Expression of specific granulo-macro growth factors was concomitantly increased, and HL60 cells exhibited increased sensitivity to RA treatment. In conclusion, our data suggest that there is a strong link between H3K4me3 levels and DNA methylation, and therefore, bring into question the relevance of the “instructive model” through which DNA hypermethylation is guided by H3K27me3, at least in AML. In addition, we unmasked a role for the loss of *IRX2* in the onset of AML.

Materials and Methods

Reagents.

Reagents included all-trans-retinoic acid Ara-C (cytosine-1-β-D- arabinofuranoside), and RNase-A (Sigma-Aldrich; St. Louis, MO, USA).

Human AML samples, AML cell lines and immunophenotyping.

Normal CD34+ HSC/HPC and mononuclear CD34− cell fractions were purified from the cord blood of 3 healthy donors using immunomagnetic column separation (Miltenyi Biotec Inc.; Auburn, CA; and STEMCELL Technologies; Cambridge, MA, USA) as previously described (63-64). Cord blood was provided by the UOS Regional Bank of Cord Blood (Dr. Maria Screnci, Rome, Italy). The purity of column-selected CD34+ cells (> 95%) was assessed by flow cytometric (FACS) analysis. Cells were labeled with anti-CD34-APC human (Miltenyi Biotech Inc.) and sorted with the FACSAria III (Becton Dickinson, BD Biosciences; Franklin Lakes, NJ, USA). Acute myeloid leukemia (AML) samples (n = 19) were provided by Division of Hematology Department of Cellular Biotechnologies and Hematology (University of Rome “Sapienza”) and were obtained from the peripheral blood of newly diagnosed leukemia patients showing more than 60% leukemic infiltration. Cytogenetic analyses were performed as previously reported (63-64). Cases were classified according to the French-American-British (FAB) classification scheme (65). Peripheral blood lymphocytes were isolated from the peripheral blood of a single healthy donor. HL-60 (ECACC-Sigma-Aldrich cat#98070106), NB-4 and HEL cell lines were cultured in RPMI 1640 supplemented with 10% FBS and 1% penicillin-streptomycin. Kasumi-1 and ML-2 cell lines were maintained in RPMI 1640 supplemented with 20% FBS and 1% penicillin-streptomycin. The AML-193 cell line was cultured in Iscove's MDM supplemented with 10% FBS, 1% penicillin- streptomycin and 20% conditioned medium from the cell line 5637 (DSMZ ACC 35). The MONO-MAC-1 cell line was cultured in RPMI supplemented with 10% FBS, 1% penicillin-streptomycin, 2 mM L-glutamine, 1× non-essential amino acids and 1 mM sodium pyruvate. Immunophenotyping of the HL60 cell line was performed using direct immunofluorescence staining of cells with APC-conjugated mouse anti-human CD11b (clone ICRF44) and CD14 (clone M5E2; Becton Dickinson Pharmingen). A minimum of 50,000 events was recorded for each sample in a FACSAria III (Becton Dickinson, BD Biosciences). NB-4, HEL, Kasumi-1, ML-2, AML-193 and Mono- Mac-1 were kindly prvided by Prof. Pier Giuseppe Pelicci (Director of Molecular Mechanisms of Cancer and Aging Unit of European Institute of Oncology-IEO Milan.)

Ethical Statement.

Human AML samples were obtained from patients after informed consent and the study was conformed to the standards set by the Declaration of Helsinki.

Restriction Landmark Genome Scanning (RLGS).

Briefly, RLGS was performed according to published protocols using the enzyme combination of NotI-EcoRV-HinfI (49-50). Nonspecific sheared ends AML genomic DNA (1 to 3 μg) were blocked in a 10-μL reaction by the addition of nucleotide analogues (αS-dGTP , αS-dCTP , ddA TP , and ddTTP) and DNA polymerase I (2 U; 37°C, 20 minutes), followed by enzyme inactivation (65°C, 30 minutes). We adjusted the buffer and digested DNA (37°C, 2 hours) with NotI (20 U; Promega, Madison, WI), which is sensitive to methylation. Sequenase (version 2.0; U.S.B., Cleveland, OH, USA) was used to fill in the NotI ends with [α-32P]dGTP and [α-32P]dCTP (Amersham GE Healthcare Europe, Freiburg, Germany) for 30 minutes at 37°C. We digested the labeled DNA (37°C, 1 hour) with EcoRV (20 U; Promega) and separated a portion by electrophoresis through a 60-cm–long, 0.8% agarose tube gel (first-dimension separation). The agarose gel was next equilibrated in HinfI digestion buffer, and the DNA was digested in the gel with HinfI (700 U; Promega) at 37°C for 2 hours. The agarose gel was placed horizontally (rotated 90° relative to the first direction of electrophoresis) across the top of a nondenaturing 5% polyacrylamide gel, stabilized with molten agarose, and subjected to electrophoresis.

Bisulfite sequencing

Bisulfite sequencing assay was performed as described (66) on bisulfite treated genomic DNA (5 μg) using the following primer pairs:

IRX2

F5’-TGATGATGGTTAGTATGATTTTTT-3’

R5’-CACAAACTCTTATTAATTCCAAAAC-3’

NKX6-1

F5’-GTTGGGGTTGAAGTAGAGTT-3’

R5’-TCTAAACACCCACAACCCAAA-3’

FOXE-1

F5'-GTTTAGTTTAGAGTTGGGGTT-3'

R5'-ATTAAATCCTCTCTCCATTCC-3'

CABYR

F5'-GTTTAGATTTAGTTAGAAAAGGGGAA-3'

R5'-AACTCCCTAAAACATAATAATCTCC-3'

NRG-1

F: 5'-GTTTAGAGTTTAGGGTAAGGGATA-3'

R: 5'-CTAACACTAAACTAAATTTCATCTACTTC-3'

Chromatin Immunoprecipitation (ChIp)

AML cell lines were immunoprecipitated using the following antibodies for histone modifications: mouse monoclonal anti-human H3K4me3 (Abcam #ab8580; Cambridge, MA, USA), mouse monoclonal anti-human H3K27me3 (Abcam #ab6002; Cambridge MA, USA), mouse monoclonal anti- human H3Kme0 (Merck-Millipore #17-675; Burlington, MA, USA), mouse monoclonal anti-human H3K4me3 (Merck Millipore #17-678); rabbit polyclonal anti-human H3 (Abcam #ab1791; Cambridge MA,USA), and for histone modifiers: rabbit monoclonal anti-human LSD1 (Cell Signaling Technology #2148; Danvers, MA, USA), polyclonal rabbit anti-human JARID1B (Cell Signaling Technology #3273; Danvers, MA, USA), polyclonal rabbit anti-human EZH2 (Abcam #ab186006; Cambridge, MA, USA) polyclonal rabbit anti-human JMJD3 (Abcam #ab85392; Cambridge MA, USA). We used mouse anti-human IgG1, mouse anti-human IgG2a/b antibodies and rabbit anti-human (Merck Millipore) as immunoprecipitation controls. 5' promoter genomic regions, within 1kb from the putative TSS, of *FOXE1*, *IRX2*, *NKX6- 1, NRG1,* and *CABYR* genes were amplified from immunoprecipitated DNAs using the following primers:

***IRX2*:**

F5'-CTCCTCTTTCAGA TTTGTGTTGT-3'

R5'-TCACCTGAGCTACCTGTAATG3';

***NKX6-1*:**

F5'CAGCCTTTGACTCTCCTCTC-3',

R5’-GTGACGTCCCTCAAAGTTCT-3’;

CABYR:

F5’-TCCCAGGTTCAAGCAA TTCT-3’

R5’-A TCACGAGGTCAGGAGTTCA-3’;

FOXE-1:

F5’-GGTACAGTCGAGGGACCTGG-3’

R5’-CCCCAA TCGTGGCTGCCACA-3’;

NRG-1:

F5’-ACTCGGCAGAACTGAGTGAC-3’

R5’-GTGCCTGTTAGAAATCTTCAGC -3’

For each target we also amplified an intragenic region where bivalent domains are not expected to be found. In this case primer pairs used were:

IRX2:

F5’-AGGAATCGCTCGCTCAAAT-3’

R5’-GTCTCGGGGCGCTTTCTT-3’;

NKX6-1:

F5’-GAAGCTGTCTAACTGGCCTGA-3’,

R5’-AAGGAGGGTGGGGACTACAAAG-3’;

CABYR:

F5’-ATTCCTGCAGGTACAAAGCA-3’

R5’-ACAGGGTTTCACCATCTTGG-3’;

FOXE-1:

F5’-GCTGGTTTTCCCTGTCTCTG-3’

R5’-CCACCATTGCTGCCAAATAC-3’;

NRG-1:

F5’-TGCCAGTATTTTGGGTGAGT-3’

R5’-CATTGTATGTACCTTAGTGG-3’

All primers pairs were designed with Primer express Version 3.0 software (Applied Biosystems; Foster City, CA, USA):

RT-PCR quantitation was performed in triplicate using the SYBR Green dye detection method. Values obtained for each immunoprecipitated sample were quantified *versus* the

respective input and calculated following the $2^{-\Delta CT}$ method.

RNA extraction and analysis.

Total RNA was extracted from cells using the TRIzol RNA isolation system (Invitrogen/ThermoFisher Scientific; Waltham, MA, USA). cDNA was made from the extracted total RNA (1 μg) with the High Capacity RNA-to-cDNA Kit (Applied Biosystems). *IRX2*, *FOXE1*, *NKX6-1*, *NRG1*, *CABYR*, glyceraldehyde phosphate dehydrogenase (GAPDH), *JARID1A*, *JARID1B*, *LSD1*, *JMJD3*, *JARID2*, and *EZH2*

IRX2

F5’-TACCAGAAGCAAGCACGAGA-3’

R5’-TACTTGTCCTTGCACTCCGA-3’

FOXE1

F5’-GTTCACACGTTCCCCGCAA T-3’ R5’-TGGA TCCTGGTCTCTGGTGT-3’

CABYR

F5’-TTGACAGCACTGCAAAGGCA-3’

R5’-GATCACCATCAGCAATAGGG-3’

NRG1

F5’-TACTGTCACCCAGACTCCTA-3’

R5’-TCTACAGGTGACATACGAGC-3’

NKX6-1

F5’-CTGCCACGCTTTAGCAGCCTGA-3’

R5’-TTGAGGCGCTCTGTCTCCGAG-3’

JARID1A

F5’-TCCTCGTGCCTATCACTCT-3’

R5’-CCTTAGGCGTCGGTAATGAT-3’

JARID1b

F5’-A TTGCCTCAAAGGAA TTTGG-3’

R5’-CA TCACTGGCA TGTTGTTCA-3’

LSD1

F5’-GGTAACAGGTCTTGGAGGGA-3’

R5’-GGCTTCATAAAGTGGGCATT-3’

JMJD3

F5’-GGAGTACCCTGCATGGAGAT-3’

R5’-GGGTATGGATATGGGTGAGG-3’

JARID2

F5’-CA TCCCAAGTGTCCTCCACT-3’

R5’-GAGTAGCTGGAGGGGGTAG-3’

EZH2

F5’-TTACTGCTGGCACCGTCTGATGTG-3’

R5’-TGTCTGCTTCATCCTGAGAAATAATCTCC-3’

CSF1

F5’-CCTTGACAAGGACTGGAATA-3’

R5’-GGTACAGGCAGTTGCAATCA-3’

CSF2

F5’-CTGAACCTGAGTAGAGACAC-3’

R5’-GGCAGTGCTGCTTGTAGT-3’

CSF3

F5’-ACAGTGCACTCTGGACAGTG-3’

R5’-ACAGCTTGTAGGTGGCAC-3’

obtained for each sample were quantified *versus GAPDH* cDNA levels and calculated following the $2^{-\Delta CT}$ method.

Plasmid constructs and lentiviral infection.

The pIRES2-DsRed2-IRX2 plasmid was generated by cloning the *IRX2* coding sequence (BamH1-SalI) into the pIRES2-DsRed2 vector. The coding sequence was amplified by an MCF7 cell line cDNA preparation. We used the following primers:

F5’-cgggatcCACCATGTCCTACCCGCAG-3’

R5’-acgcgtcgacTGCTCGGCCCTTCTATAGGTA-3’

The control plasmid was the empty vector pIRES2-DsRed2. Plasmids were verified by sequencing. Lentivirus was produced in HEK293T cells transfected with the pIRES2-DsRed2-IRX2 and pIRES2-DsRed2 constructs using a third-generation lentiviral system. HL-60, AML-193 and AML patient samples were infected with lentiviral particles, and

red fluorescent protein-positive cells were sorted.

References

1. Ehrlich M. (2009) DNA hypomethylation in cancer cells. *Epigenomics* 1(2), 239–259.

2. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. (2003) Induction of tumors in mice by genomic hypomethylation. *Science* 300, 489-92.

3. Eden A, Gaudet F, Waghmare A, Jaenisch R (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300, 455.

4. Irizarry RA, Ladd-Acosta C., Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyan S, Feinberg AP. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186.

5. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, Kotliar D, Zhang W, Ghandi M, Garraway L, Fernandes SM, Livak KJ, Gabriel S, Gnirke A, Lander ES, Brown JR, Neuberg D, Kharchenko PV, Hacohen N, Getz G, Meissner A, Wu CJ (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 26, 813–825.

6. Mudbhary R, Hoshida Y, Chernyavskaya Y, Jacob V, Villanueva A, Fiel M, Chen X, Kojima K, Thung S, Bronson RT, Lachenmayer A, Revill K, Alsinet C, Sachidanandam R, Desai A, SenBanerjee S, Ukomadu C, Llovet JM, Sadler KC (2014) UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell* 25, 196–209.

7. Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014) Somatic retrotransposition in human cancer revealed by whole- genome and exome sequencing. *Genome Res*. 24, 1053–1063.

8. Tubio, J. M., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., et al.. (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343.

9. Pfeifer GP (2018) Defining Driver DNA Methylation Changes in Human Cancer. *Int J*

Mol Sci., 12;19(4).
10.Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA. (2014) Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res.* 9, 1421-32.
11. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyan S, Feinberg AP. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet*. 41, 178–186.
12. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd- Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41, 1350–1353.
13. Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342.
14. Greger V, Debus N, Lohmann D, Höpping W, Passarge E, Horsthemke B. (1994) Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma. *Hum Genet.* 94, 491-6.
15. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* 91, 9700-4
16. Stirzaker C, Millar DS, Paul CL, Warnecke PM, Harrison J, Vincent PC, Frommer M, Clark SJ.Extensive (1997) DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.* 57, 2229-37.
17. Graff JR, Gabrielson E, Fujii H, Baylin SB, Herman JG. (2000) Methylation patterns of the E-cadherin 5' CpG island are unstable and reflect the dynamic, heterogeneous loss of E-cadherin expression during metastatic progression. *J Biol Chem.* 275, 2727-32.
18. Costello JF, Futscher BW, Kroes RA, Pieper RO. (1994) Methylation-related chromatin structure is associated with exclusion of transcription factors from and

suppressed expression of the O-6-methylguanine DNA methyltransferase gene in human glioma cell lines. *Mol Cell Biol.* 14, 6515-21.
19. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A.* 95, 6870-5.
20. Weinberg R.A. (2007) The Biology of Cancer. New York: Garland Science,.
21. Sproul D., Meehan R.R. (2013) Genomic insights into cancer-associated aberrant CpG island hypermethylation. *Brief Funct. Genomics* 12, 174-90.
22. Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* 72, 141–196;
23. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Danenberg PV, Laird PW. (1999) CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression. *Cancer Res.* 59, 2302–2306
24. Di Croce L., Raker VA, Corsaro M, Fazi F, Fanelli M, Faretta M, Fuks F, Lo Coco F, Kouzarides T, Nervi C, Minucci S, Pelicci PG. (2002) Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* 295, 1079-82.
25.Wu, X, Zhang, Y. (2017) TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534.
26. Atlasi Y, Stunnenberg HG. (2017) The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet.* 18, 643-658.
27. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H. (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre- marks genes for de novo methylation in cancer. *Nat Genet.* 39, 232-6.
28. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW. (2007) Epigenetic stem cell signature in cancer. *Nat Genet.* 39,157–158.
29. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins

DN, Herman JG, Baylin SB (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet.* 39, 237–242.
30. Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbruger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D, Collisson E, Zhu J, Yegnasubramanian S, Matsui W, Baylin SB (2012) A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res.* 22, 837– 849.
31. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun HM, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM, Tollenaar RA, Laird PW. (2012) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* 22, 271–282.
32. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871-4.
33. Wu X, Gong Y, Yue J, Qiang B, Yuan J, Peng X. (2008) Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing *Nucleic Acids Res*. 36, 3590-9.
34. Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, Stunnenberg HG. (2012) Sequential ChIP- bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* 22, 1128-38.
35. Court F, Arnaud P. (2017) An annotated list of bivalent chromatin regions in human ES cells: a new tool for cancer epigenetic research. *Oncotarget.* 8, 4110-4124.
36. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553- 60.
37. Chinaranagari S, Sharma P, Chaudhary J (2014) EZH2 dependent H3K27me3 is involved in epigenetic silencing of ID4 in prostate cancer. *Oncotarget* 5, 7172-82.

38. Stern JL, Paucek RD, Huang FW, Ghandi M, Nwumeh R, Costello JC, Cech TR. (2017) Allele-Specific DNA Methylation and Its Interplay with Repressive Histone Marks at Promoter-Mutant TERT. *Genes Cell Rep.* 21, 3700-3707.

39. Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP- seq) directly informs methylation status of histone-modified DNA. *Genome Res.* 22, 1120-7.

40. Cai Y, Lin JR, Zhang Q, O'Brien K, Montagna C, Zhang ZD (2018) Epigenetic alterations to Polycomb targets precede malignant transition in a mouse model of breast cancer. *Sci Rep.* 8: 5535, doi: 10.1038/s41598-018- 24005-x.

41. Neri F, Incarnato D, Krepelova A, Rapelli S, Pagnani A, Zecchina R, Parlato C, Oliviero S (2013) Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol.* 14, R91, doi: 10.1186/gb-2013-14-8-r91.

42.Kong L, Tan L, Lv R, Shi Z, Xiong L, Wu F, Rabidou K, Smith M, He C, Zhang L, Qian Y, Ma D, Lan F, Shi Y, Shi YG (2016) A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. *Nucleic Acids Res* 44, 8682–8692.

43. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet.* 39, 457-66.

44. Okitsu CY , Hsieh CL (2007) DNA methylation dictates histone H3K4 methylation. *Mol Cell Biol.* 27, 2746-57.

45. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 15, 464(7291).

46. Steen K. T. Ooi, Qiu C, BernsteinE, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714–717.

47. Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA

methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep*. 10, 1235–1241.
48. Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, Tian C, Wang J, Cong Y, Xu Y (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640-4.
49. Costello JF Smiraglia DJ Plass C. (2002) Restriction landmark genome scanning *Methods* 27:144–149.
50. Zardo G, Tiirikainen M, Hong C, Misra A, Feuerstein BG, Volik S, Collins CC, Lamborn KR, Bollen A, Pinkel D, Albertson DG, Costello JF. (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat Genet.* 32, 453-8.
51. Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1, 299- 312.
52. Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA, Bourque G, Oh S, Ruan Y, Ng HH, Wei CL. (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells *Cell Stem Cell* 1, 286-98.
53. Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4, 80-93.
54. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290-302.
55. Adamowicz M, Radlwimmer B, Rieker RJ, Mertens D, Schwarzbach M, Schraml P, Benner A, Lichter P, Mechtersheimer G, Joos S (2006) Frequent amplifications and abundant expression of TRIO, NKD2, and IRX2 in soft tissue sarcomas. *Genes Chromosomes Cancer* 45, 829-38.
56. Liu T, Zhou W, Zhang F, Shi G, Teng H, Xiao J, Wang Y. (2014) Knockdown of IRX2 inhibits osteosarcoma cell proliferation and invasion by the AKT/MMP9 signaling pathway. *Mol Med Rep* 10, 169-74.
57. Kang H, Wilson CS, Harvey RC, Chen IM, Murphy MH, Atlas SR, Bedrick EJ,

Devidas M, Carroll AJ, Robinson BW, Stam RW, Valsecchi MG, Pieters R, Heerema NA, Hilden JM, Felix CA, Reaman GH, Camitta B, Winick N, Carroll WL, Dreyer ZE, Hunger SP, Willman CL (2012) Gene expression profiles predictive of outcome and age in infant acute lymphoblastic leukemia: a Children's Oncology Group study *Blood* 119, 1872-81.

58. Rauch TA, Wang Z, Wu X, Kernstine KH, Riggs AD, Pfeifer GP (2012) DNA methylation biomarkers for lung cancer *Tumour Biol.* 33, 287–96.

59. Sato T, Arai E, Kohno T, Takahashi Y, Miyata S, Tsuta K, Watanabe S, Soejima K, Betsuyaku T, Kanai Y. (2014) Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease. *Int J Cancer* 135, 319-34.

60. Kamalakaran S, Varadan V, Giercksky Russnes HE, Levy D, Kendall J, Janevski A, Riggs M, Banerjee N, Synnestvedt M, Schlichting E, Kåresen R, Shama Prasada K, Rotti H, Rao R, Rao L, Eric Tang MH, Satyamoorthy K, Lucito R, Wigler M, Dimitrova N, Naume B, Borresen-Dale AL, Hicks JB. (2011) DNA methylation patterns in luminal breast cancers differ from non- luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol Oncol.* 5, 77-92.

61. Werner S, Brors B, Eick J, Marques E, Pogenberg V, Parret A, Kemming D, Wood AW, Edgren H, Neubauer H, Streichert T, Riethdorf S, Bedi U, Baccelli I, Jücker M, Eils R, Fehm T, Trumpp A, Johnsen SA, Klefström J, Wilmanns M, Müller V, Pantel K, Wikman H (2015) Suppression of early hematogenous dissemination of human breast cancer cells to bone marrow by retinoic Acid-induced. *Cancer Discov.* 5, 506-19.

62. Werner S, Stamm H, Pandjaitan M, Kemming D, Brors B, Pantel K, Wikman H (2015) Iroquois homeobox 2 suppresses cellular motility and chemokine expression in breast cancer cells *BMC Cancer* doi: 10.1186/s12885-015-1907- 4.

63. Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, Bozzoni I (2005) A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell* 123, 819-831.

64. Fazi F, Racanicchi S, Zardo G, Starnes LM, Mancini M, Travaglini L, Diverio D, Ammatuna E, Cimino G, Lo-Coco F, Grignani F, Nervi C. (2007); Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. *Cancer*

Cell 12, 457-466.
65. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, Sultan C. (1985); Proposed revised criteria for the classification of acute myeloid leukemia: a report of the French-American-British Cooperative Group. *Ann Intern Med* 103, 620-625.
66. Zardo G, Ciolfi A, Vian L, Starnes LM, Billi M, Racanicchi S, Maresca C, Fazi F, Travaglini L, Noguera N, Mancini M, Nanni M, Cimino G, Lo-Coco F, Grignani F, Nervi C (2012) Polycombs and microRNA-223 regulate human granulopoiesis by transcriptional control of target gene expression. *Blood* 119, 4034-46.

Figure legends.

Figure 1. RLGS DNA methylation data. Each row represents a RLGS spot corresponding to a specific CpGi listed in the far right column. Each red rectangle represents a methylation event occurring in the corresponding spot.. PBL: peripheral blood lymphocytes; CD34+: CD34+ hHSC/MPP; CD34-: CD34- mature cells; AML: AML patient sample.

Figure 2. Integration of RLGS DNA methylation data with the corresponding histone mark in hESC. Each row represents a specific RLGS spot. Annotation of the spot and the corresponding mark are indicated in the two right columns. Each red rectangle in the diagram represents a methylation event occurring in the corresponding spot. Yellow rectangle indicates that the corresponding spot is marked by H3K4me3 in hESC; orange rectangle indicates that the corresponding spot is marked by H3K27me3 in hESC; blue rectangle indicates that the corresponding spot is marked by a bivalent domain in hESC. AML: AML patient sample.

Figure 3. Frequency of aberrant CpG island methylation in AML samples and relative histone marks in hESC.

A and B. Panels show the number of aberrantly methylated CpG islands and their relative chromatin marks in hESC (yellow, H3K4me3; orange, H3K27me3; blue, bivalent domain) in AML patient samples in panel A and AML cell lines in panel B. In each panel, columns A, B, and C identify the number of samples presenting the same aberrantly methylated CpG island. In panel A, column A indicates that a specific CpG

island is aberrantly methylated in at least 1 and up to 6 out of 19 AML patient samples; column B indicates that a specific CpG island is aberrantly methylated in at least 7 and up to 12 out of 19 AML patient samples; column C indicates that a specific CpG island is aberrantly methylated in at least 13 and up to 19 out of 19 AML patient samples. In panel B, column A indicates that a specific CpG island is aberrantly methylated in at least 1 and up to 2 out of 6 AML cell lines; column B indicates that a specific CpG island is aberrantly methylated in at least 3 and up to 4 out of 6 AML cell lines; column C indicates that a specific CpG island is aberrantly methylated in at least 5 and up to 6 out of 6 AML cell lines.

Figure 4. Genomic bisulfite sequencing of selected target genes. Bisulfite sequencing performed on control CD34+ hHSC/MPP and CD34- mature cells and 19 AML patient samples to assess the methylation status of specific CpG islands in the 5' promoter regions of FOXE1, IRX2, NKX6-1, NRG1 and CABYR genes. Each circle represents a specific CpG dinucleotide as sequenced in 8-10 clones from the same AML sample. The color intensity scale (from 0 to 100%) indicates the percentage methylation of each CpG dinucleotide in the sequenced clones.

Figure 5. Profiles of H3K4me3 and H3K27me3 enrichments in normal hematopoietic cell types. Chromatin immunoprecipitation experiments showing H3K4me3 and H3K27me3 relative enrichments at the 5' promoter region and at an intragenic *locus*, of selected targets in normal CD34+ HSC/MPP and mature CD34- cells, amplified region are identified by nucleotide position (nt) in the schematic diagram. Three independent ChIp experiments were performed and qRT-PCR amplification was performed in triplicate on the precipitated genomic DNAs. Error bars: SD, standard deviation.

Figure 6. Increased H3K4me3 and H3K27me3 enrichments at selected targets in AML patient samples and cell lines. ChIp analysis assessing H3K4me3, H3K27me3 and histone 3 (H3, nucleosomal density control), enrichments in the 5' promoter region of the indicated target genes in 6 AML patient samples and 5 AML cell lines compared to the relative enrichments in normal CD34+ HSC/MPP and mature CD34- cells (3 samples from 3 independent separations). qRT-PCR was performed to detect ChIp DNAs. Statistical analysis was performed using the nonparametric Wilcoxon rank sum test. *P*-

value < 0.05 =*; data were represented as a Box Wisker Plot for H3K27me3 and H3K4me3 modifications, the ends of the box represent the upper and lower quartiles, the median is marked by a vertical line inside the box and the whiskers represent the minimum and maximum data values. Error bars in H3 data representation: SD, standard deviation.

Figure 7. H3K4me3 and H3k4me0 enrichments at selected targets. Chromatin immunoprecipitation comparing H3K4me3 and H3K4me0 enrichments at the 5' promoter region of indicated targets (FOXE1, NKX6-1, IRX2, CABYR and NRG1) in 3 independent selections of normal CD34+ HSC/MPP and mature CD34- cells from cord blood and in 4 patient AML samples. Each qRT-PCR amplification was performed in triplicate. Statistical analyses was performed using the non parametric Wilcoxon rank sum test: pvalue < 0,05= *. Error bars: SD, standard deviation.

Figure 8. Expression of human histone methylases and demethylases is altered in AML relative to normal mature CD34- cells. Panel A: qRT-PCR to detect expression levels of *JARID1A*, *JARID1B*, *LSD1*, *EZH2*, *JMJD3*, and *JARID2* in normal mature CD34- cells and in 9 AML patient samples. Panel B: qRT-PCR to detect expression levels of *JARID1A*, *JARID1B*, *LSD1*, *EZH2*, *JMJD3*, and *JARID2* in normal mature CD34- cells and in 5 human AML cell lines (HL60, ME-1, ML-2, AML-193, HEL). Each sample was analyzed in triplicate. Statistical analysis was performed in triplicate using the nonparametric Wilcoxon rank sum test. *P*-value < 0.05 =*; *P*-value < 0.01 =**. Error bars: SD, standard deviation.

Figure 9. Histone methylases and demethylases occupancy at 5' promoter region of selected targets.

Chromatin immunoprecipitation experiments showing the enrichments of EZH2, JMJD3, JARID1B and LSD1 at the 5' promoter region (Proximal) and at a control region (Distal)of CABYR, FOXE1, IRX2, NKX6-1 and NRG1. Amplified proximal and distal DNA regions corresponds to those amplified to detect histone modifications shown in figure 5, 6 and 7). Experiment was conducted on four different purifications of mature CD34- cells and on four preparations of HL60, ML-2 and AML193 cell lines. Each qRT-PCR amplification was executed in triplicate. IgG: control immunoglobulin. Statistical analyses were performed using the non parametric Wilcoxon rank sum test; *P*-value <

0.05 =*; *P*-value < 0.01 =**; *P*-value < 0,001 =***, and data were represented as a Box Wisker Plot where the ends of the box represent the upper and lower quartiles, the median is marked by a vertical line inside the box and the whiskers represent the minimum and maximum data values.

Figure 10. mRNA expression analyses of selected targets in CD34-, AMLs and AML cell lines. qRT-PCR analyses of mRNA expression levels of the indicated targets (IRX2, FOXE1, NKX6-1, NRG1 and CABYR) in normal control mature CD34- cells (CD34-), patient AML samples (from #1 to #10) and in 4 AML cell lines (HL60, ME-1, ML-2 and AML193). Error bars: SD, standard deviation.

Figure 11. Differentiation promoting effect of restored IRX2 expression in AML. Panel A: qRT-PCR to detect expression of IRX2 in HL60-IRX2 and AML-193-IRX2, and in one newly diagnosed AML patient sample (AML-IRX2) 48 hours after lentiviral infection (see Methods for experimental details). Expression was compared to mRNA levels in control samples including uninfected parental cell types as well as cells infected with empty vector (EV). Panel B: Wright-Geimsa staining of uninfected, EV, and *IRX2* expressing cell types. Panel C: qRT-PCR to detect expression levels of CSF1, CSF2, and CSF3 in cell types indicated. qRT-PCR analyses are the result of 3 independent experiments performed in triplicate. Panel D: % of CD11b and CD14 positive cells in HL60-EV and HL60-IRX2 cells 4 and 7 days following lentiviral infection. The results presented are the average of 3 independent experiments. Panel E: % of CD11b and CD14 positive cells in HL60-EV and HL60- IRX2 after treatment with RA for the hours indicated. Statistical analysis was performed using the nonparametric Wilcoxon rank sum test. *P*-value < 0.05 =*. Error bars: SD, standard deviation.

Acknowledgments:

This research was funded by**:** University of Rome Sapienza, Progetti di Ricerca C26A15S3JR. We thank Prof. Robin Foa', Head of Division of Hematology, and Dr. Sabina Chiaretti for providing AML human samples.

Author contributions: S.S., C.M., C.C., R.M. performed the experiments and collected the data. G.Z., C.C., C.G., A.R., M.Z, analyzed the data. G.Z. designed and supervised

the study. G.Z. wrote the manuscript. All the authors contributed in manuscript writing and editing and approved the final version of the manuscript.

Conflicts of interest: The authors declare no conflict of interest.