A histone methyltransferase point mutation demonstrates a mechanism for translocation-driven acute leukemias

Laila Kobrossy[a], Weiyi Xu[b,c], Chunling Zhang[d], Christopher E. Turner[b], and Michael S. Cosgrove[a,*]

[a] Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, New York 13210, United States
[b] Department of Cell and Developmental Biology, SUNY Upstate Medical University, Syracuse, New York 13210, United States
[c] Current Address: Cell Biology Department, Harvard Medical School, Boston, MA
Department of Cell and Developmental Biology, SUNY Upstate Medical University, Syracuse, New York 13210, United States
[d] Department of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York 13210, United States

Key Words: Histone, methylation, leukemia, 11q23, translocation, epigenetic plasticity, H3K4 methylation, trimethylation, MLL, MLL1, KMT2A, SETd1A, EMT, WDR5, Win motif, cancer stem cell, leukemia stem cell, induced pluripotent stem cells, CRISPr/Cas9, ChIP-Seq, RNA-Seq, HOXA9, Meis1

*To whom correspondence should be addressed:

Michael S. Cosgrove, Ph.D.
SUNY Upstate Medical University
Department of Biochemistry and Molecular Biology
750 East Adams Street
Syracuse, New York 13210
Email: cosgrovm@upstate.edu
Phone: (315) 464-7751

Summary

Epigenomic and gene expression changes in iPS cells with a mutated MLL1 histone methyltransferase suggest an oncogenic mechanism for MLL1-translocations.

Abstract

Much of our understanding of the pathology of acute leukemias is based on studies of 11q23 chromosomal translocations of the gene encoding the mixed lineage leukemia-1 (MLL1) histone H3 lysine 4 (H3K4) methyltransferase. Translocations of the MLL1 gene result in MLL1-fusion ($MLL1_F$) proteins that replace the catalytic SET domain with 1 of more than 80 fusion partners that are thought to drive leukemogenesis through a gain-of-function mechanism. However, this mechanism does not explain how loss of the catalytic SET domain results in upregulated H3K4 trimethylation and target gene expression in $MLL1_F$-leukemic stem cells. In this investigation, we introduced a homozygous loss-of-function point mutation in MLL1 in human induced pluripotent stem cells. The resulting cells mislocalized a different H3K4 trimethyltransferase, SETd1a, to MLL1 target gene promoters and recapitulated a phenotype commonly observed in $MLL1_F$ leukemic stem cells, with aberrant histone methylation and gene expression signatures characteristic of "mixed lineage" cells. These results suggest that loss of MLL1 histone methyltransferase activity, and the associated impact on the epigenome, are common molecular features underlying $MLL1_F$ pathology.

Translocations that affect MLL1, a histone methyltransferase, are found in 70% of infant acute lymphoblastic leukemias (*2-4*). The resulting MLL1-fusion ($MLL1_F$) proteins replace the MLL1 catalytic SET (suppressor of variegation, enhancer of zeste, trithorax) domain with one of more than 80 fusion partners that are thought to drive leukemogenesis (*5*). In addition, multiple missense mutations and deletions across the MLL1 coding sequence appear in a growing subset of MLL1-associated cancers (*6-8*), highlighting the importance of understanding how MLL1 mutations drive tumorigenesis. This investigation sought to explain a paradox in the oncogenic mechanism of $MLL1_F$ leukemias: How removal of the SET domain with methyltransferase activity results in upregulated histone trimethylation and target gene expression in $MLL1_F$-leukemic stem cells.

The MLL1 protein (also known as KMT2A, ALL1, HRTX, HRX, MLL) catalyzes histone H3 lysine 4 (H3K4) methylation, an epigenetic mark associated with expression of genes required for neurogenesis, hematopoiesis, and embryonic development (*9-13*). H3K4 can be mono-, di- or trimethylated, with each state associated with distinct enzymatic complexes and biological functions (*14*). H3K4 trimethylation is enriched in the promoters of active genes (*15-17*) and is required to recruit nucleosome remodeling complexes that regulate promoter accessibility (*18-22*). $MLL1_F$ leukemic stem cells (LSCs) have a hyper-H3K4 trimethylation phenotype that results in upregulation of LSC maintenance genes including HOXA9, MEIS1, MEF2C, and ZEB1 in the majority of $MLL1_F$ leukemias (*23-26*). The molecular mechanisms involved are not well understood.

We investigated the genome-wide impact of a loss of MLL1 enzymatic activity on cellular identity and differentiation in human induced pluripotent stem (iPS) cells. We found that introducing a homozygous loss-of-function point mutation in MLL1 resulted in mislocalization of a different methyltransferase, the SETd1a H3K4 trimethyltransferase, to promoters of MLL1 LSC maintenance genes. This mislocalization recapitulated a phenotype commonly observed in $MLL1_F$ leukemic stem cells: "mixed lineage" expression of multiple lineage markers in a single cell (*27, 28*), with similar aberrant histone methylation and gene expression signatures. These results suggest that $MLL1_F$ pathology leading to leukemias involves a dominant negative loss of MLL1 histone methyltransferase activity and a broad associated impact on the epigenome.

Results

Generating a homozygous point mutation in MLL1 in human iPS cells

H3K4 mono-, di-, or trimethylation by MLL family members including MLL1 requires $WRAD_2$ (for WDR5, RbBP5, ASH2L, and DPY-30 proteins), which associates with the MLL SET domain to create a core enzymatic complex (Figure 1A) (*29-31*)(*32*). WDR5 binds arginine 3765 of MLL1 in an evolutionarily conserved *Win* motif (for WDR5 interaction) (Fig. 1A, inset) (*1, 33-35*). Replacement of R3765 with alanine abolishes MLL1 core complex assembly and methylation activity without altering protein folding (*33, 36-38*).

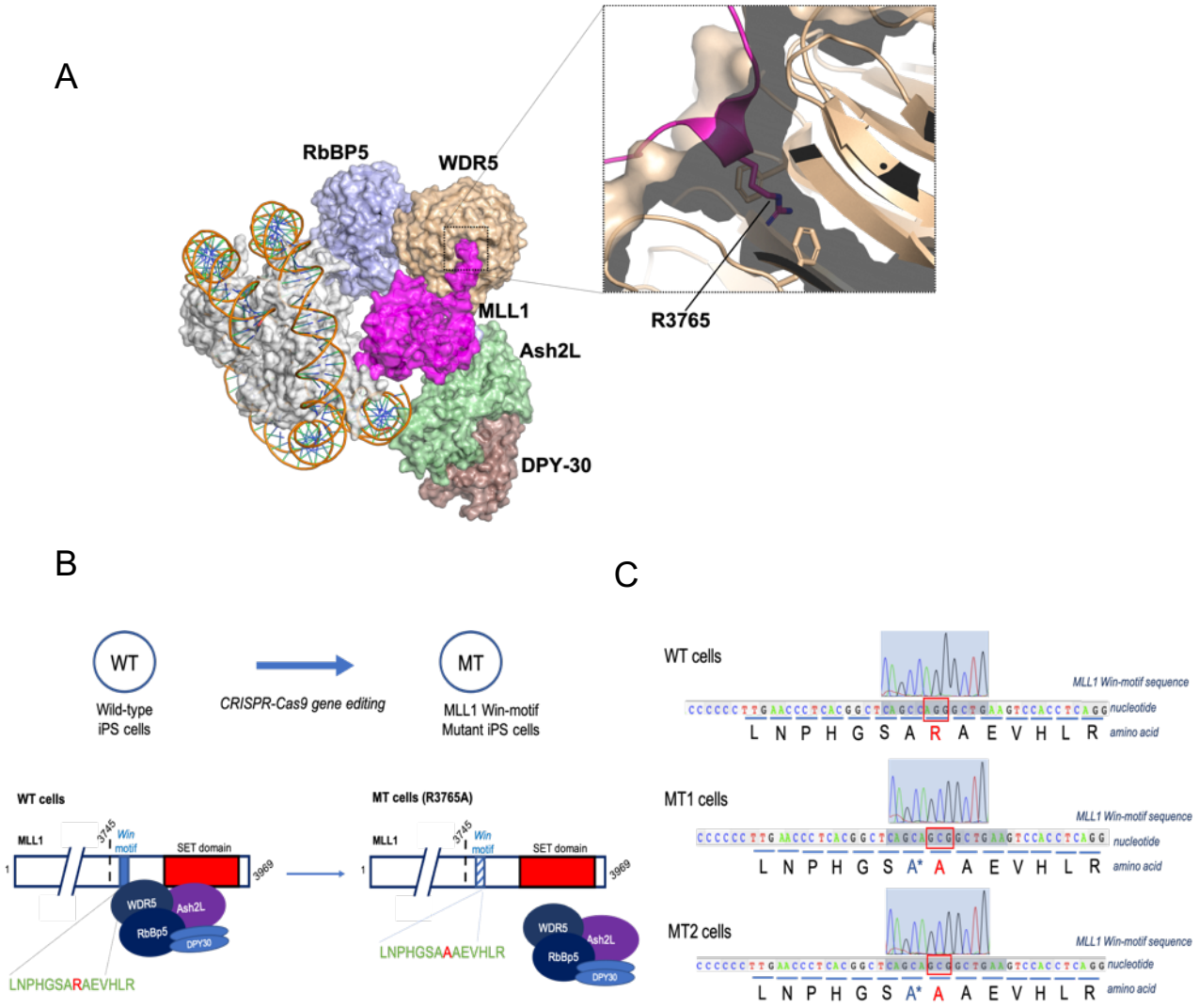


Figure 1. Generating a homozygous R3765A point mutation of a key arginine residue in the MLL1 Win-motif critical for MLL1 core complex assembly and enzymatic activity. (A) Three-dimensional structure of the MLL1 core complex bound to a nucleosome with DNA in orange and histones white (PDB code: 7UD5). Inset: X-ray structure of WDR5 (tan) bound to MLL1 Win motif peptide (magenta) (PDB code: 4ESG). The position of the crucial Win motif arginine 3765 in MLL1 is indicated. (B) Upper: schematic of mutant (MT) iPS cell line generation from wild-type (WT) cells using the CRISPR-Cas9 system. Lower: loss-of-function of MLL1 Win-motif through arginine>alanine point mutation leads to complex disassembly and loss of H3K4 methylation activity (1). (C) Chromatograms of next generation sequencing results confirming R3765A mutation in exon 32 of the KMT2A locus in MT cells. Two clones, MT1 and MT2, were generated. Genomic DNA from WT (upper), mutant clone MT1 (middle), and clone MT2 (lower) iPS cells was extracted and PCR used to amplify the MLL1 Win-motif region. To prevent recutting the modified genome after repair, a silent mutation GCC>GCA was introduced in the single-stranded oligodeoxynucleotide donor site that was the DNA template for homology-directed repair.

We introduced the R3765A loss-of-function point mutation in the MLL1 *Win*-motif in human ASE-9203 iPS cells. We targeted a 20 bp region in the MLL1/KMT2A gene to introduce a R3765>A3765 point mutation (AGG>GCG) in both alleles of the KMT2A exon 32 (Fig. 1B, C). Two single cell-derived clones were identified, 14 (MT1) and 39 (MT2), with homozygous mutations confirmed by in house DNA sequencing (Fig. 1C). Sequencing also confirmed the absence of unintended mutations in the *Win*-motifs of the other five MLL/SET1 family members and eliminated the possibility of off-target effects in the top 10 predicted CRISPR/Cas9 off-target loci (Figure S1A, B).

MLL1 R3765A substitution leads to redistribution of H3K4 trimethylation to transcription start sites

To investigate the impact of the *Win*-motif mutation on global H3K4 methylation levels in MT cells, we used western blots with whole cell extracts from wild type (WT) and both MT clones. The loss-of-function mutation might have caused a global decrease in H3K4 methylation levels in MT cells (*39, 40*). Instead, H3K4 monomethylation and dimethylation levels in both MT clones were similar to WT, as were global levels of unmodified histone H3. Unexpectedly, both MT1 and MT2 showed *increased* levels of trimethylated H3K4 compared to WT (Fig. 2A), suggesting that loss of MLL1 enzymatic activity leads to hyper-H3K4 trimethylation throughout the genome. Since a similar H3K4 hypertrimethylation defect was noted in mouse MLL-AF10 LSCs (with AF designating the fusion partner) (*41*) and in patient-derived MLL-AF4 blast cells (*42*), these results suggested that the enzymatic activity of MLL1 may be required to limit the spread of the gene-activating H3K4 trimethyl mark.

To further investigate this hypothesis, we performed chromatin immunoprecipitation followed by sequencing (ChIP-Seq) to determine how loss of MLL1 enzymatic activity affects H3K4 trimethylation across the genome in iPS cells. Similar to observations in MLL-AF10 LSCs (*41*) and patient-derived MLL-AF4 leukemia cells (*42*), we observed a significant shift in the H3K4 trimethylation peaks towards transcriptional start sites in both MT1 and MT2 iPS cells compared to WT (Figure 2B) with a corresponding decrease in signal across intergenic and intronic regions (Fig. 2C). Peak annotation and gene enrichment analysis revealed significant overlap in the genomic regions of MT1 and MT2 cells that acquired novel H3K4 trimethylation peaks compared to WT (Fig. 2D), confirming consistent changes in these peaks in the two MT clones. Strikingly, increased H3K4 trimethylation occurred at several of the same genes that are frequently upregulated in $MLL1_F$ leukemias including HoxA9, HoxA10, Meis1, ZEB1 and HMG2A. Kyoto Encyclopedia for Genes and Genomes enrichment analysis of the 5091 overlapping H3K4 trimethylation peaks in MT1 and MT2 cells revealed a significant enrichment for genomic regions encoding biomarkers of multiple cancer networks (see Fig. S2A), WNT signaling

molecules, cell cycle machinery, cell adhesion, and the epithelial-mesenchymal transition (EMT) signaling pathway (Fig. 2E). Taken together, these results suggest that disruption of the MLL1 Win motif-WDR5 interaction results in increased H3K4 trimethylation in promoter regions and a potential shift to a more open, transcriptionally permissive chromatin state.

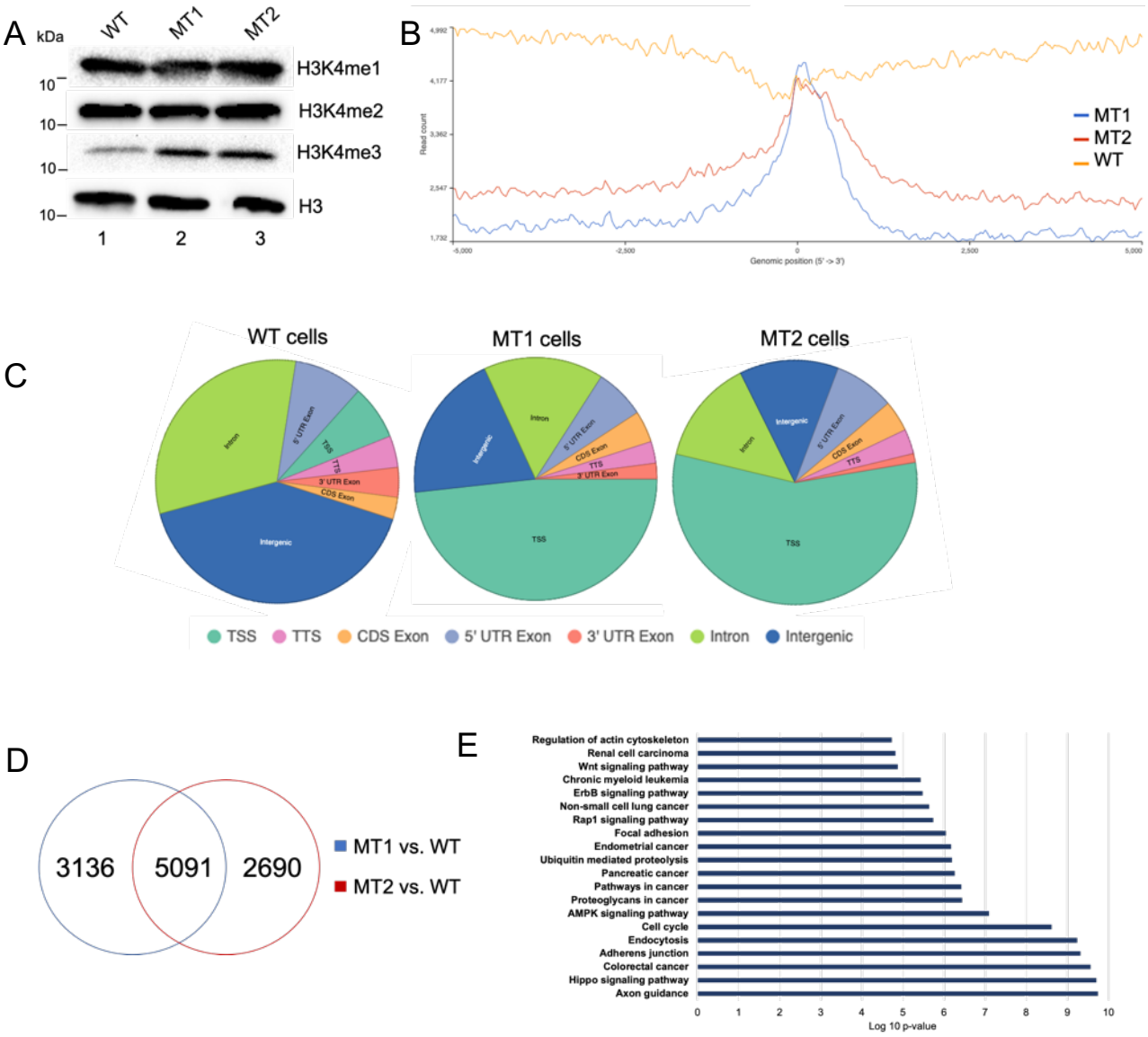


Figure 2. R3765A substitution in MLL1 induces relocalization of the H3K4-trimethyl mark to transcription start sites. (A) Levels of H3K4me1-3 in wildtype (WT) and mutant (MT) cells (clones MT1 and MT2). Whole cell extracts from WT (lane 1) and MT (lanes 2 and 3) were immunoblotted using antibodies specific for monomethylated (H3K4me1), dimethylated (H3K4me2), and trimethylated H3 lysine 4 (H3K4me3). Histone H3 levels were used as a loading control. (B) Mean H3K4me3 ChIP-seq read counts (normalized to IgG) near transcription start sites (TSS) (genomic position 0) in WT (orange) vs. MT cells (blue and red). (C) Pie charts showing relative enrichment of H3K4me3 read counts in WT vs. MT cells in different genomic regions. (D) Venn diagram showing the overlap of differential ChIP-seq H3K4me3 peaks detected in MT1 and MT2 clones, relative to WT. (E) KEGG enrichment analysis of top canonical pathways in overlapping ChIP-seq peaks from MT1 and MT2 clones normalized to WT (see Fig. 2C). Log 10 of enrichment p-values ($p < 0.05$) for each pathway is on the x-axis.

MLL1 R3765A substitution increases expression of $MLL1_F$ LSC maintenance genes

To determine if the changes in global H3K4 trimethylation levels were associated with changes in gene expression, we used RNA-Seq to compare transcriptional dynamics of MT and WT cells and found ~2800 differentially expressed genes (Fig. 3A). Consistent with a more permissive chromatin environment, the majority of genes were upregulated in MT compared to WT cells (1914 genes, or 68%) (Fig. 3A). Gene Set Enrichment Analysis revealed that of the 25 highest scoring (most similar) gene sets, about 50% were associated with aggressive malignancies and multicancer invasiveness (Fig. 3B). The other half correlated with gene sets involved in chromatin remodeling (25%), extracellular matrix dynamics and EMT signaling (17%), and stem cell pluripotency (8%). The relevance of these signaling cascades in cancer was confirmed via ingenuity pathway analysis, as highlighted in the top 10 canonical pathways differentially regulated in MT vs. WT iPS cells (Fig. 3C).

Further evaluation of the cancer-specific gene sets revealed significant enrichment of MLL1 target genes (*39*), $MLL1_F$-induced gene expression patterns (*43*), and cancer invasiveness transcriptional signatures (*44*) (Fig. 3D). Consistent with the ChIP-Seq results, we observed significant upregulation of several leukemia- and stem cell-associated genes including HOX genes and cancer associated microRNAs, many of which are key players in the development of $MLL1_F$ leukemias (*45-52*) (Fig. 3E). These observations were not due to differences in MLL1 mRNA or protein levels (Fig. S3A and S3B), or differences in WDR5 and RbBP5 protein levels in MT cells (Fig. S3B). These results point to a shift to a more permissive epigenetic state that supports aberrant transcriptional activation of key regulatory genes in MT cells. Furthermore, the occurrence of this aberrant gene expression signature in cells without an MLL1 fusion protein suggests that loss of MLL1 enzymatic activity drives overexpression of the core genes associated with $MLL1_F$ leukemias.

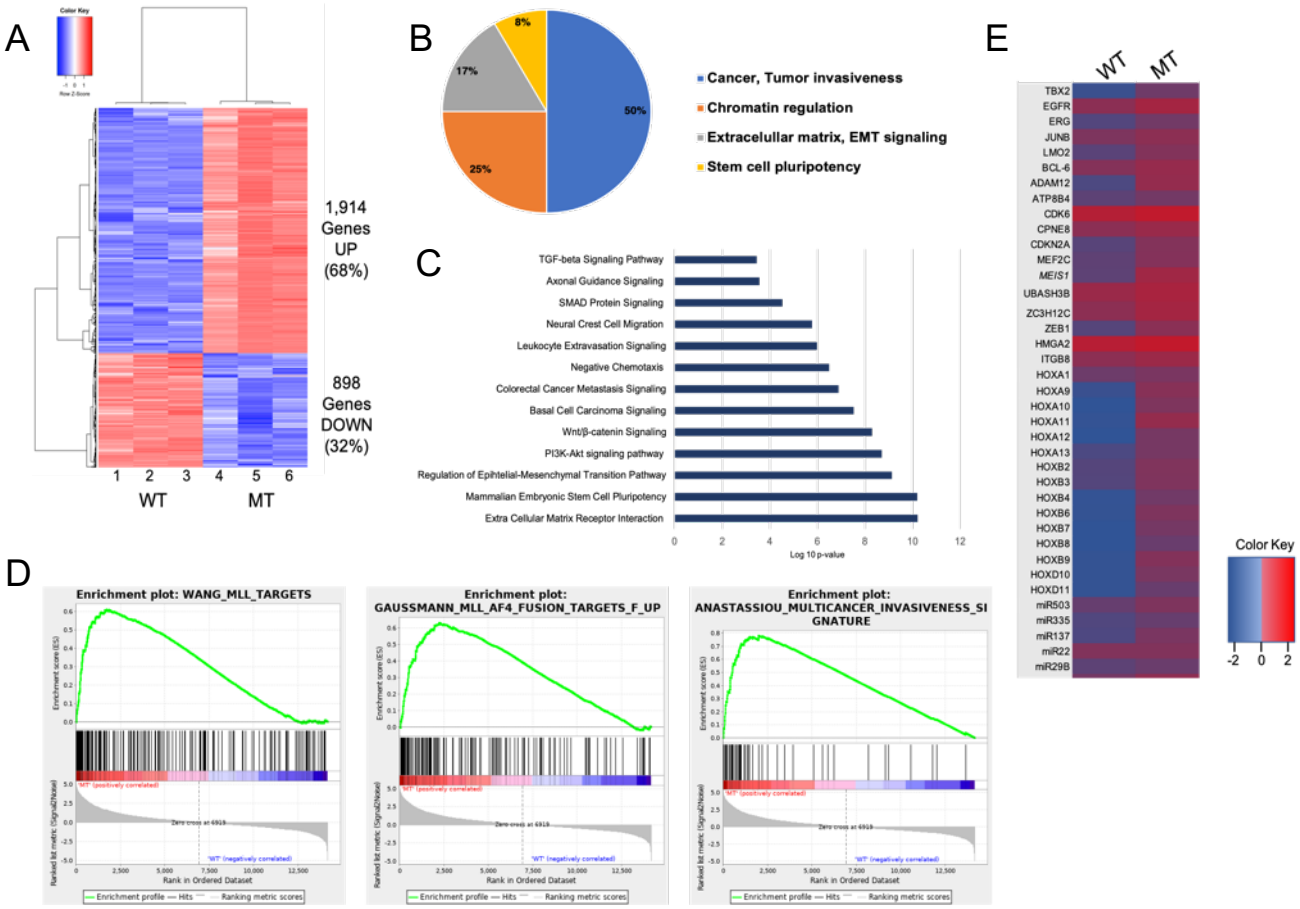


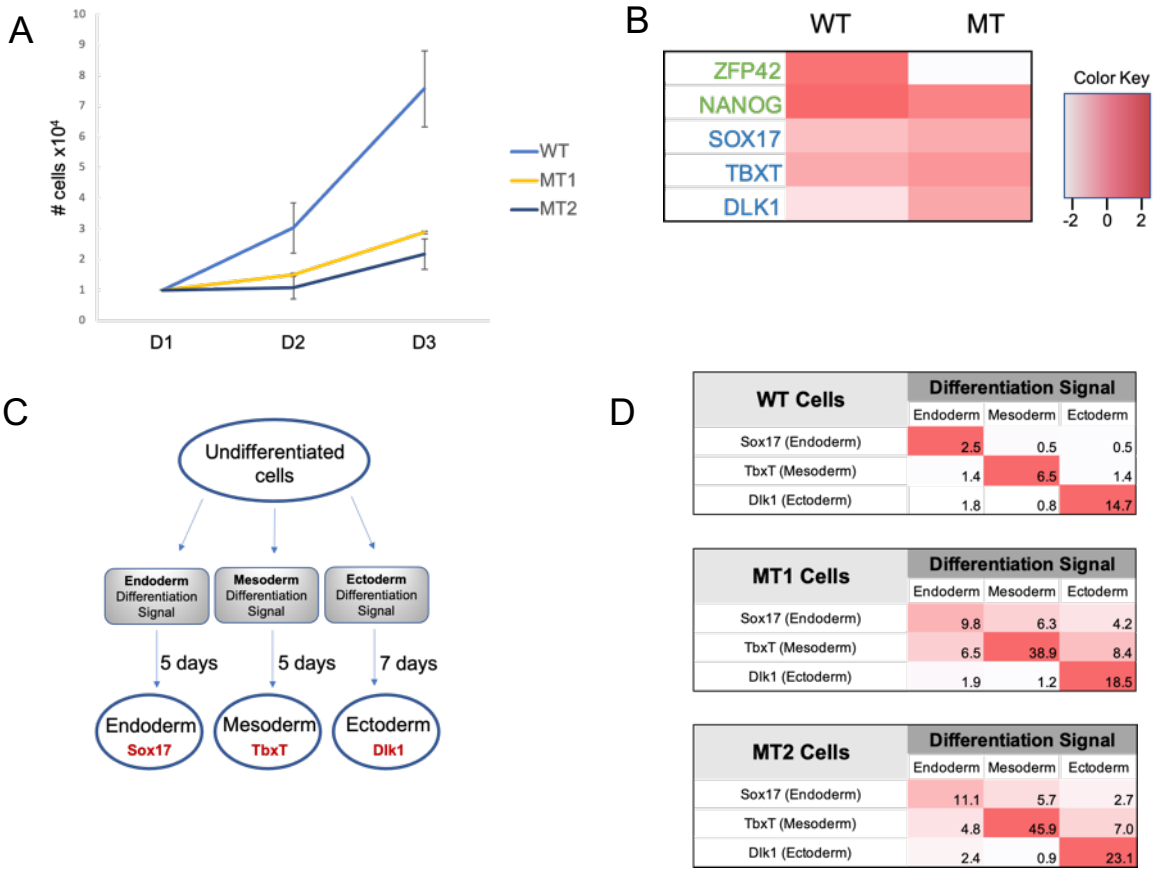
Figure 3. R3765A substitution in MLL1 induces an MLL1F-like gene expression phenotype in human iPS cells. (A) Heatmap showing differentially expressed genes in mutant (MT) vs. wild-type (WT) cells (clone MT2) in RNA-Seq: 68% of differentially expressed genes were upregulated (UP) in MT cells, compared to 32% that were downregulated (DOWN). The heatmap was generated using normalized counts from three independent replicates of total RNA from WT (columns 1-3) and MT cells (columns 4-6). (B) Top 25 enriched gene sets in MT cells identified through gene set enrichment analysis (GSEA) were stratified according to biological pathways they represent. (C) Gene ontology enrichment analysis showing top canonical pathways in differentially expressed genes in MT cells. X-axis, log10 p-values; y-axis, list of top overrepresented pathways based on all differentially expressed genes in MT vs. WT cells. (D) Enriched gene sets in MT cells. GSEA revealed that MLL1 target genes (left), MLL-AF4 fusion protein-regulated genes (middle), and multicancer invasiveness signature genes (right) were enriched in MT cells (clone MT2). Enrichment score (ES) and individual genes (vertical black lines) are on the y- and x-axes, respectively. Green connects points of ES and genes. ES is the maximum deviation from zero, calculated for each gene going down the ranked genes in order of signal/noise ratio. Heat map shows the degree of correlation of genes enriched in MT cells. Significance threshold was < 0.05. (E) Heatmap of RNA-Seq results for differentially expressed genes with transcripts significantly enriched in MT cells (clone MT2). Significantly altered transcripts were selected based on adjusted p value < 0.05.

MLL1 R3765A promotes a "mixed lineage" differentiation phenotype in iPS cells

Both MT clones were highly sensitive to external growth conditions and consistently grew slower than WT cells when seeded at the same density (Fig. 4A). Together with insights from RNA-Seq and ChIP-Seq, these observations suggested the possibility of pluripotency and/or cell identity defects in MT clones. Consistent with this hypothesis, "human embryonic stem cell pluripotency" was the second most significantly altered canonical pathway in MT cells by RNA-Seq (Fig. 3C). ZFP42, encoding

pluripotency marker Rex1, was the most downregulated gene in MT cells compared with WT in all triplicate repeats (Fig. 4B). This was accompanied by downregulation of Nanog, another key pluripotency marker, as well as increased expression of SOX17, TBXT, and DLK1, which are key regulators of lineage-specific differentiation of iPS and stem cells (*53-55*) (Fig. 4B).

We therefore investigated if MT cells were capable of differentiating into the three germ layers (endoderm, mesoderm, and ectoderm) compared to WT. We used commercially available differentiation



WT Cells	Differentiation Signal		
	Endoderm	Mesoderm	Ectoderm
Sox17 (Endoderm)	2.5	0.5	0.5
TbxT (Mesoderm)	1.4	6.5	1.4
Dlk1 (Ectoderm)	1.8	0.8	14.7

MT1 Cells	Differentiation Signal		
	Endoderm	Mesoderm	Ectoderm
Sox17 (Endoderm)	9.8	6.3	4.2
TbxT (Mesoderm)	6.5	38.9	8.4
Dlk1 (Ectoderm)	1.9	1.2	18.5

MT2 Cells	Differentiation Signal		
	Endoderm	Mesoderm	Ectoderm
Sox17 (Endoderm)	11.1	5.7	2.7
TbxT (Mesoderm)	4.8	45.9	7.0
Dlk1 (Ectoderm)	2.4	0.9	23.1

Figure 4. R3765A substitution in MLL1 induces a mixed lineage phenotype in iPS cells. (A) Mutant (MT) cells grow more slowly than wild-type (WT). WT and MT cells (Clones MT1 and MT2) were plated in triplicate in 24-well plates and counted at 24 (D1), 48 (D2) and 72 hours (D3). Data are relative increase in mean cell number compared to cells counted at D1. Error bars are 1 standard deviation with p=0.05. (B) Heatmap of pluripotency (green) and lineage-specific (blue) markers differentially expressed in RNA-Seq of MT vs. WT cells. Significantly altered transcripts were selected based on a false-discovery rate (FDR) adjusted p-value and FDR ≤ 0.05. Color key shows differential expression based on log2 scale. (C) Differentiation. Undifferentiated WT and MT (MT1 and MT2 clones) iPS cells were treated with STEMdiff endoderm, mesoderm, or ectoderm differentiation media for 5-7 days, followed by RNA extraction and RT-qPCR analysis for expression of endoderm (Sox17), mesoderm (TbxT), and ectoderm (Dlk1) lineage-specific biomarkers. (D) MT cells display a mixed-lineage phenotype when prompted to differentiate into one of three germ layers. Heatmap color intensity corresponds to the relative increase in lineage-specific biomarker gene expression within each cell type upon treatment with lineage-specific differentiation media (WT, upper; MT1, middle; MT2, lower). Columns correspond to the STEMdif differentiation signal, and rows have relative increase in biomarker gene expression relative to undifferentiated controls. Fold-increase in biomarker gene expression was quantified by RT-qPCR using cDNA collected at experimental endpoints (Day 5 for endoderm and mesoderm, Day 7 for ectoderm), and after normalization to GAPDH levels. Data represent two independent experiments with qPCR using three technical replicates; standard deviation for mean fold-enrichment varied from ± 6-29% for p=0.05.

media that are used to assess stem cell pluripotency (Fig. 4C). WT cells differentiated into each of the three germ layers, as demonstrated by the significantly increased expression of lineage-specific makers in differentiated WT cells compared to undifferentiated controls (Fig. 4D). MT cells also differentiated into each of the three germ layers, with a significantly higher potency, as shown by the relative fold-increase in gene expression in MT cells compared to WT cells. However, differentiating MT cells consistently displayed increased transcriptional or epigenetic plasticity, resulting in a "*mixed-lineage*" phenotype throughout all three germ layers, demonstrated by the misexpression of two or more lineage-specific markers in cells treated with single-lineage-specific growth media. These results suggested a role for the MLL1 core complex in conveying or ascribing precise tissue-specific patterning across germ layers, possibly by limiting epigenetic plasticity at key MLL1 target genes.

MLL1 R3765A substitution promotes EMT and increases invasiveness

The activation of the EMT pathway is a hallmark of cancer invasion and metastasis, specifically for the development and maintenance of cancer stem cells (*56-58*). The EMT pathway is mediated by specific transcription factors such as Zeb1, Twist1, and Slug, and involves the repression of epithelial-specific adhesion molecules (such as E-cadherin) (*59, 60*). This occurs simultaneously with the transcriptional activation of mesenchymal structural components (N-cadherin, vimentin), and extracellular matrix molecules such as fibronectin (*61-64*). Transcripts of several of these mesenchymal signature biomarkers were significantly increased in MT cells (Fig. 5A). We investigated if loss of MLL1 activity was associated with cancer stem cell formation and increased invasiveness.

We explored newly acquired mesenchymal and oncogenic characteristics of the MT cells. First, we compared MT and WT cells for protein levels and localization of E-cadherin and Hic-5, which are, respectively, signature biomarkers for epithelial and mesenchymal cells (*59, 65, 66*). Immunohistochemistry revealed increased staining at focal adhesions of the promigratory, adhesion adapter protein Hic-5 (Fig. 5B), and low levels of epithelial cell-specific E-cadherin at the cell periphery in both MT clones relative to WT (Fig. 5C). Phalloidin-conjugate staining revealed actin stress fibers in both MT clones (Fig. 5B), suggesting a possible shift in cytoskeletal and cell migration dynamics. These results were confirmed by immunoblots, which revealed a decrease in E-cadherin protein in both MT clones, with a corresponding increase in Hic-5 (Fig. 5D).

To determine if cell movement dynamics in MT cells were typical of cells that have undergone an EMT, we conducted time-lapse imaging of WT and MT cells and characterized cell motility parameters including migration velocity. Motility assays revealed a significant increase in migration of both MT clones compared to WT iPS cells (Fig. 5E), further suggesting increased invasive potential with the loss of MLL1 activity.

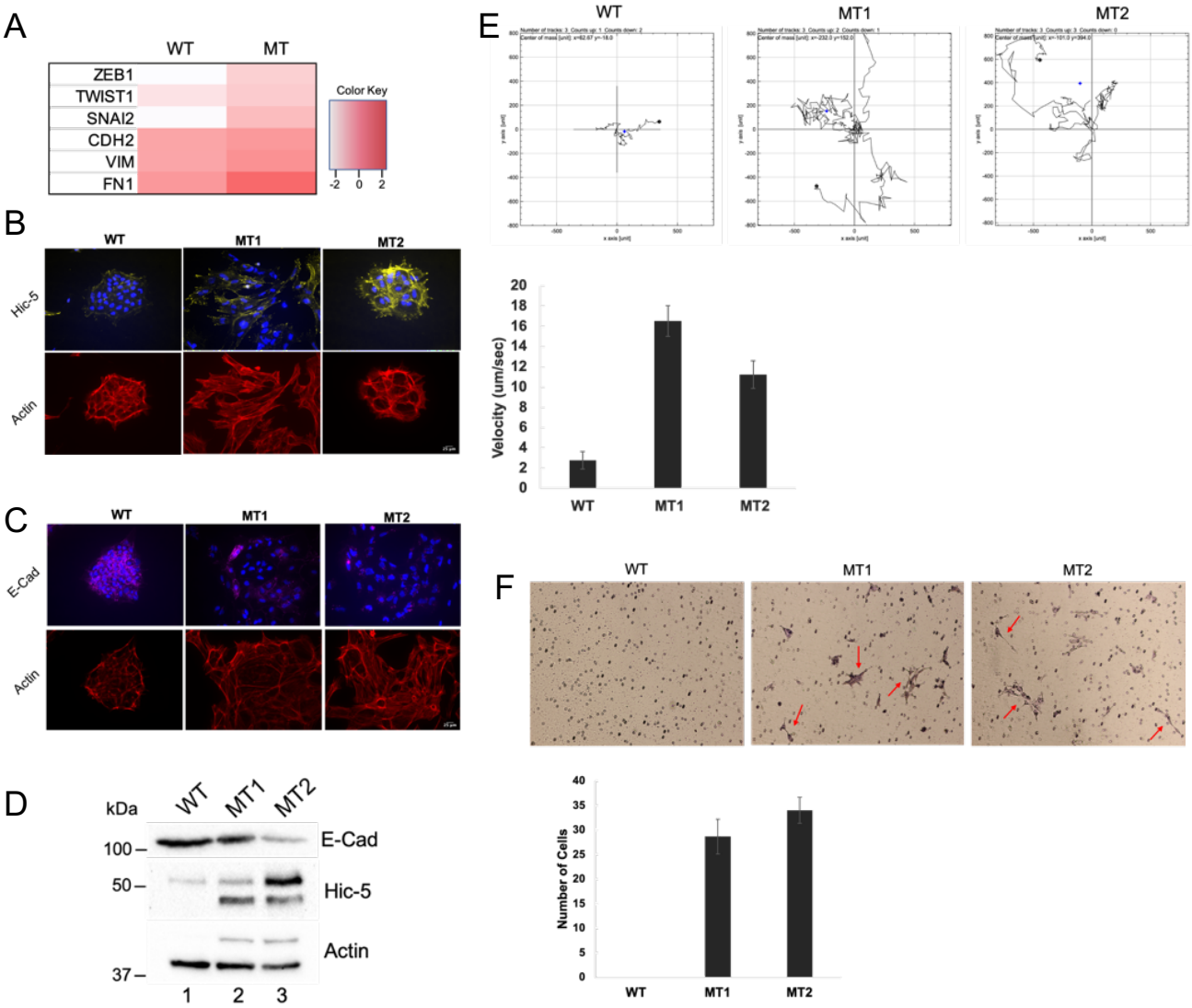


Figure 5. R3765A substitution in MLL1 induces EMT and invasivness. (A) Heatmap of mesenchymal signature biomarkers differentially expressed in RNA-Seq of mutant (MT) vs. wild-type (WT) cells. Significantly altered transcripts were selected based on adjusted p-value and FDR ≤ 0.05. Color key shows differential expression based on a log2 scale. (B) Immunofluorescence of WT and MT1 and MT2 cells showing differential staining for Hic-5 (yellow). DAPI (blue), nuclear regions; phalloidin (red) actin filaments. (C) Immunofluorescence of WT and MT1 and MT2 cells showing differential staining for E-cadherin (E-Cad, magenta). DAPI (blue), nuclear; phalloidin (red), actin filaments. (D) Immunoblots of whole cell extracts from WT (lane 1) and MT1, lane 2, and MT2, lane 3 cells using antibodies specific for E-cadherin (E-Cad) and Hic-5. Beta-actin was the loading control. (E) Upper: Rose diagrams showing cell motility assays in WT and MT1 and MT2 cells by time-lapse microscopy imaging. Lower: Bar chart of cell migration velocity. Error bars are standard deviation with p-value 0.05. (F). Upper: Photographs of representative Matrigel invasion assays, performed as two biological replicates. Red arrows, Giemsa-stained invading MT cells. Lower: Cells counted in three separate vision fields for each biological replicate, with average number of cells/view for each cell type. Error bars are 1 standard deviation with p-value 0.05.

We next investigated if MT cells acquired invasive characteristics, as suggested by the RNA-Seq and ChIP-Seq pathway enrichment. In transwell invasion assays, multiple MT cells invaded extracellular matrix-coated filters (Fig. 5F), suggesting substantial invasive potential. In comparison, no control WT iPS cells invaded, as expected from nondifferentiated iPS cells with more epithelial cellular characteristics, including robust cell-cell adherens junctions (*60*) (Fig. 5B). These results were consistent with the hypothesis that MLL1 histone methyltransferase activity is important for restricting the EMT phenotype, and loss of activity promotes cancer stem cell formation and invasiveness.

MLL1 R3765A promotes redistribution of SETd1a trimethyltransferase to $MLL1_F$ target gene promoters

$MLL1_F$ leukemia cells show increased H3K4 trimethylation that is thought to result from the $MLL1_F$ protein binding to MLL1 target genes and activating wild type MLL1 expressed by the unmutated allele of these heterozygous cells (*67*). However, the similar ectopic H3K4 trimethylation phenotype in MT cells with a homozygous loss-of-function point mutation indicated involvement of a different enzyme. Humans have 6 MLL family members: MLL1-4, SETd1a and SETd1b (also known as SET1a,b). Biochemical reconstitution revealed that SETd1a and SETd1b are efficient trimethyltransferases responsible for the bulk of H3K4 trimethylation in cells while MLL1-4 are relatively poor trimethyltransferases (*31, 68, 69*). Furthermore, small hairpin RNA knockdown of SETd1a but not SETd1b transcripts induces apoptosis in murine and human $MLL1_F$-leukemia cells (*70*). We therefore hypothesized that increased H3K4 trimethylation is due to relocalization of SETd1A to MLL1 target gene promoters.

Consistent with this hypothesis, both MT clones had increased SETd1A protein relative to WT (more noticeable in MT2) (Fig. 6A). In contrast, WDR5 levels in both MT clones were unchanged (Fig. 6A). ChIP with SETd1A and H3K4 trimethylation antibodies determined if their localization was altered at promoters of MLL1 target genes in MT cells. SETd1a (Fig. 6B) and H3K4 trimethylation (Fig. 6C) levels increased at promoters of genes for MMP1, Meis1 and HoxA9, which RNA-Seq showed were all highly upregulated in MT cells. In contrast, the REX1 promoter showed decreased SETd1a localization and the ACT promoter was unchanged, consistent with, respectively, downregulated and unchanged mRNA expression by RNA-Seq. These results suggested a key role for SETd1A relocalization in creating a more permissive chromatin environment that promotes the mixed lineage cancer stem cell phenotype after loss of MLL1 enzymatic activity.

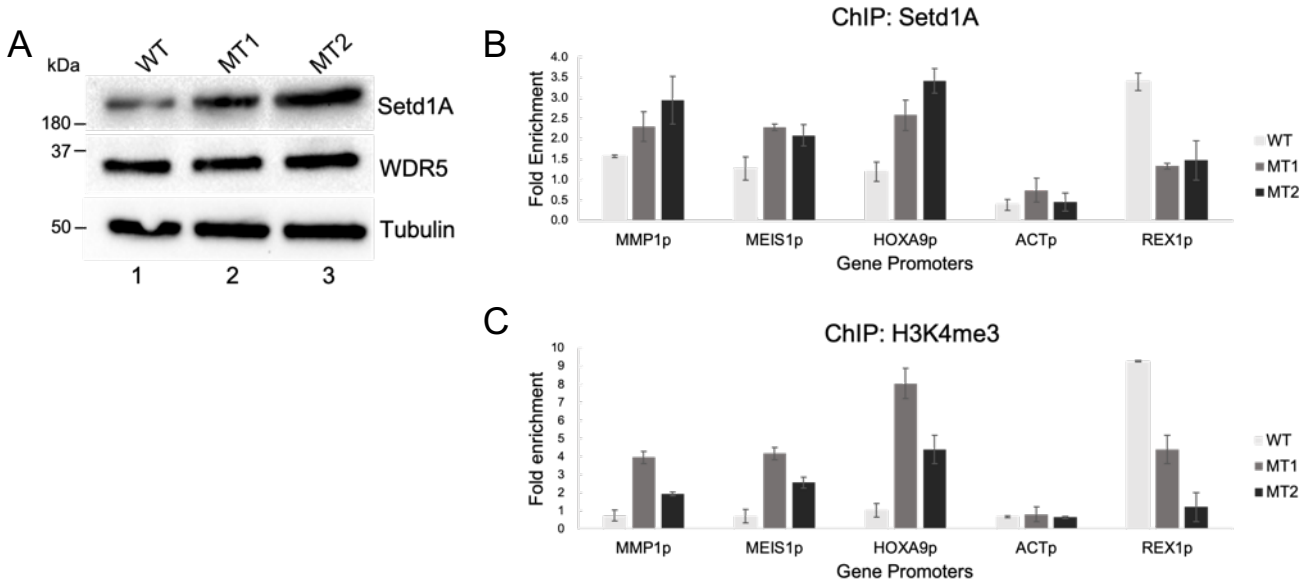


Figure 6. SET1/MLL family member SETd1a is stabilized and localized to MLL1 target genes in MT cells. (A) Increased levels of SETd1A in mutant (MT) cells. Immunoblot analysis was performed on whole cell extracts from wild-type (WT) (lane1) and MT (MT1, lane2; MT2, lane 3) cells using antibodies specific for SETd1A and WDR5. Tubulin was the loading control. (B) ChiP-qPCR results for SETd1A localization to promoters of upregulated genes (MMP1, HOXA9, MEIS1), a downregulated gene (Rex1p) and a control gene with unchanged expression in WT vs. MT cells. Bars are mean values from in two independent biological replicates, each as three technical replicates. Fold-enrichment was relative to an IgG ChIP control for each sample, and all values were normalized to the GAPDH housekeeping gene. Error bars show standard deviation with p=0.05. (C) as in (B) except ChIP used an H3K4 trimethylation antibody.

Discussion

Leukemias with MLL1 mutations have especially poor prognosis (*71*). However, despite 30 years of knowing that cells harboring MLL1 translocations display both lymphoid and myeloid lineage markers (*27, 28*), we lack a basic understanding of the molecular basis and oncogenic mechanism of this "mixed lineage" phenotype. While several mechanisms are postulated for how $MLL1_F$ proteins drive leukemogenesis, conflicting evidence raises questions about the role of MLL1's histone methyltransferase activity. In $MLL1_F$ cells, the sequence encoding the catalytic SET domain in one of the two MLL1 alleles is replaced by a noncatalytic fusion partner, suggesting a potential haploinsufficiency mechanism. While experimental deletion of the SET domain in mice results in altered HOX gene expression and developmental abnormalities (*72*), conditional deletion of the SET domain in hematopoietic stem and progenitor cells does not alter global levels H3K4 mono-, di-, or trimethylation, which in part has led to the suggestion that MLL1's histone methyltransferase activity is dispensable for hematopoiesis and leukemogenesis (*73*). However, this observation may be cell-type specific, as enriched populations of murine and human $MLL1_F$ LSCs display a hyper H3K4 trimethylation phenotype that promotes misexpression $MLL1_F$ LSC maintenance genes (*41, 42*). Furthermore, these cells can be antagonized

by overexpression of the histone H3K4 demethylase, KDM5B, suggesting that inhibiting the enzyme(s) responsible for the altered H3K4 trimethylation pattern in $MLL1_F$ LSCs may be a potential treatment strategy for $MLL1_F$-leukemias (*41*).

Fully realizing this therapeutic potential, however, requires understanding how $MLL1_F$ proteins promote the hyper-H3K4 trimethylation phenotype in $MLL1_F$ LSCs. While it has been suggested that the aberrant pattern of H3K4 trimethylation results from $MLL1_F$ interaction with transcriptional elongation machinery and activation of the wildtype MLL1 protein encoded by the unaffected allele (*67, 74, 75*), there is currently no direct experimental evidence supporting this hypothetical activation mechanism. Furthermore, our results argue against this $MLL1_F$ activation hypothesis since we observed increased H3K4 trimethylation and gene expression similar to LSCs when both MLL1 alleles were inactivated in MT cells.

An alternative possibility is that the hyper-trimethylation phenotype of LSCs could be due to the enzymatic activity of other H3K4 methyltransferases that may be mis-localized in the absence of MLL1 enzymatic activity. Indeed, while it is commonly thought that MLL1 catalyzes H3K4 trimethylation in cells, biochemical reconstitution experiments reveal that it is a relatively poor trimethyltransferase, instead catalyzing predominantly H3K4 mono- and dimethylation (*30, 31*). In contrast, similar in vitro experiments show that the human SETd1a/b complexes are relatively efficient trimethyltransferases (*31*), which is consistent with experiments showing they catalyze the bulk of H3K4 trimethylation in mammalian cells (*68, 69*). Together, these results led us to test the hypothesis that the H3K4 mono- and dimethyltransferase activity of MLL1 may be required to limit the spread of H3K4 trimethylation catalyzed by other members of the MLL family (Fig. 7A). Consistent with this hypothesis, our results show that homozygous loss of MLL1 enzymatic activity results in redistribution of SETd1a and the H3K4 trimethyl mark to the promoters of MLL1 target genes (Fig. 7B). A similar redistribution of the H3K4 trimethyl mark is observed in enriched populations of MLL1-AF10 LSCs (*41*) and in patient-derived MLL1-AF4 blast cells (*42*).

Reconciling this data with a potential mechanism for leukemogenesis in LSCs, we suggest that $MLL1_F$ proteins lacking the catalytic SET domain might act in a dominant negative manner by displacing MLL1 protein encoded by the remaining wild type allele or altering its activity, creating the phenotype of MLL1 homozygous loss-of-function (Fig. 7c). A dominant negative mechanism for $MLL1_F$ proteins would explain the puzzling result of why an MLL1 N-terminus fused to the transcriptionally inert LacZ protein produces acute myeloid leukemia when knocked into the MLL1 locus in mice (*76*), and is consistent with studies showing specific sequences from both the MLL1-N terminus or the fusion protein are required for cellular transformation (*77*). In addition, a loss-of-function mechanism may also explain

the large number of cancer-associated missense mutations in the MLL1 gene in cancer databases. Indeed, we recently showed that up to 85% of the 29 cancer-associated missense mutations localized in or around the MLL1 SET domain are defective for histone H3K4 mono- and/or dimethylation (*8*).

Together, these results suggest that targeting the H3K4 trimethylation activity catalyzed by the SETd1a complex may be a rational therapeutic strategy for $MLL1_F$-leukemias and possibly other cancers with MLL1 mutations. In partial agreement with this hypothesis, knockdown of SETd1a, but not other MLL family members, antagonizes growth of murine MLL1-AF9 and human MOLM-13 leukemic cells (*70*). While it has been suggested that this effect is independent of SETd1A catalytic activity (*70*), inhibitors that targets the SETd1A (and MLL1) Win motif-WDR5 interaction, but not that of other MLL family members (MLL2-4, SETd1B) (*78*), have been shown to reduce H3K4 trimethylation and selectively antagonize growth of $MLL1_F$ cells, but not other leukemia cell types lacking an MLL1 translocation (*79*). A model in which relocalization of SETd1a to the promoters of LSC maintenance genes provides a rationale for understanding how Win motif inhibitors selectively antagonize $MLL1_F$ leukemic cells.

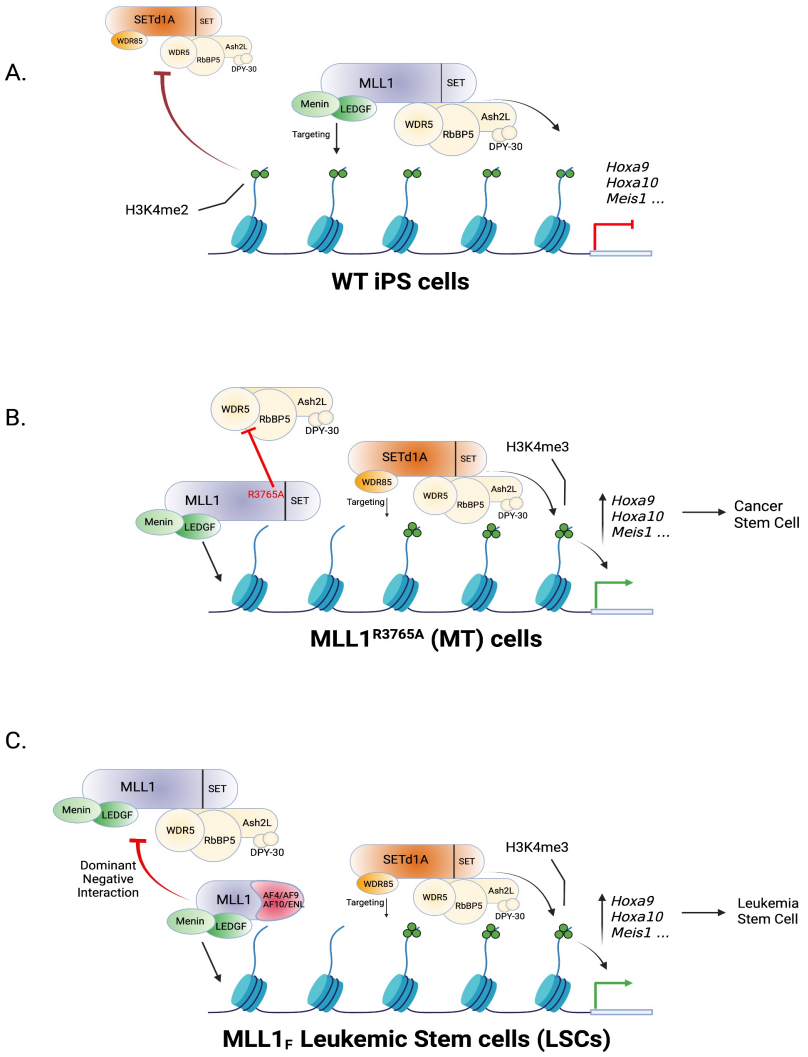


Figure 7. Model for the role of MLL1 histone methyltransferase activity in normal and leukemic stem cells. A. Wild type (WT) MLL1-catalyzed H3K4 dimethylation (H3K4me2) limits the spread of the gene-activating H3K4 trimethylation (H3K4me3) mark catalyzed by SETd1a in WT iPS cells **B.** Homozygous loss-of-function point mutation of MLL1 prevents deposition of the H3K4me2 mark at MLL1 target genes. This allows spreading of the SETd1a-catalyzed H3K4me3 mark and activation of LCS maintenance genes Hoxa9, HoxA10, Meis1, etc. **C.** Model for $MLL1_F$-induced leukemogenesis. In this model, the $MLL1_F$ protein acts in a dominant negative manner by interaction with or displacement of WT MLL1 and prevents H3K4me2 methylation at MLL1 target genes. As in the homozygous loss-of-function point mutant cells (B), lack of MLL1 catalytic activity allows spreading of the SETd1a-catalyzed H3K4me3 mark to MLL1 target gene promoters resulting in expression of LSC maintenance genes. Created with BioRender.com.

These results also suggest that inhibitors that specifically target MLL1 may induce formation of LCS cells that are difficult to detect and treat due to their slow growth and

chemotherapeutic resistance (*80*). This is important as current structure-activity relationship (SAR) studies predominantly use MLL1 for inhibitor design, which over time, is expected to select for compounds with increased MLL1 specificity. Our results may suggest that SAR studies should be focused on inhibiting the SETd1A enzyme.

Taken together, these results suggest a new cancer model in which loss of MLL1 enzymatic activity, either by homozygous point mutation in our case (Fig. 7B), or by dominant negative $MLL1_F$ protein in $MLL1_F$ leukemias (Fig. 7C), promotes redistribution of the activating H3K4 trimethylation mark to the promoters of LSC maintenance genes. Thus, instead of creating a more restrictive environment, loss of MLL1 enzymatic activity creates a more permissive chromatin environment with increased epigenetic plasticity that allows greater selection of more differentiated progeny cells that evade normal growth controls, become more invasive, and more easily acquire drug resistance.

Acknowledgements and Funding

The authors would like to thank Ashley Canning and Michael Connelly for helpful comments on the manuscript and Anne Smardon for helpful discussions. We thank Steve Hanes and Leszek Kotula for qPCR instrument access, and Ryan Palumbo and Baylee Porter for help with qPCR experiments. We thank Karen Gentile at SUNYMAC for her help with ChIP-Seq and RNA-seq data collection, and Frank Middleton for help with data deposition.

This work was supported by NIH R01 CA140522 to M.S.C., NIH R35 GM131709 to C.E.T., and by a Richard & Jean Clark Pediatric Research Funds Award to M.S.C. and C.E.T.

Competing Interests

M.S.C. owns stock and serves on the Consultant Advisory Board for Kathera Biocience Inc., the makers of antifungal technologies. M.S.C. also holds US patents (8,133,690), (8,715,678) and (10,392,423) for compounds and methods for inhibiting SET1/MLL family complexes.

Data Availability

ChIP-seq and RNA-seq datasets have been deposited in the NCBI GEO database under accession numbers GSE213239 and GSE213237, respectively.

References

1. A. Patel, V. Vought, V. Dharmarajan, M. S. Cosgrove, A conserved arginine containing motif crucial for the assembly and enzymatic activity of the Mixed Lineage Leukemia protein-1 core complex. *J Biol Chem* **283**, 32162-32175 (2008).
2. C. L. Dobson *et al.*, The mll-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *Embo J* **18**, 3564-3574 (1999).
3. A. V. Krivtsov, S. A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* **7**, 823-833 (2007).
4. A. G. Muntean, J. L. Hess, The pathogenesis of mixed-lineage leukemia. *Annu Rev Pathol* **7**, 283-301 (2012).
5. A. C. Winters, K. M. Bernt, MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4 (2017).
6. S. Kudithipudi, A. Jeltsch, Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochim Biophys Acta* **1846**, 366-379 (2014).
7. S. Weirich, S. Kudithipudi, A. Jeltsch, Somatic cancer mutations in the MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex. *Mol Oncol* **11**, 373-387 (2017).
8. A. J. Canning, S. Viggiano, M. E. Fernandez-Zapico, M. S. Cosgrove, Parallel functional annotation of cancer-associated missense mutations in histone methyltransferases. *Scientific Reports* **12**, 18487 (2022).
9. Y. Dou *et al.*, Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873-885 (2005).
10. T. Nakamura *et al.*, ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell* **10**, 1119-1128 (2002).
11. B. D. Yu, J. L. Hess, S. E. Horning, G. A. Brown, S. J. Korsmeyer, Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505-508 (1995).
12. B. D. Strahl, R. Ohba, R. G. Cook, C. D. Allis, Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc Natl Acad Sci U S A* **96**, 14967-14972 (1999).
13. T. Gan, C. D. Jude, K. Zaffuto, P. Ernst, Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia* **24**, 1732-1741 (2010).
14. A. J. Ruthenburg, C. D. Allis, J. Wysocka, Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**, 15-30 (2007).
15. C. L. Liu *et al.*, Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* **3**, e328 (2005).
16. D. K. Pokholok *et al.*, Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527 (2005).
17. H. Santos-Rosa *et al.*, Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411 (2002).
18. T. Kim, S. Buratowski, Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137**, 259-272 (2009).
19. M. G. Pray-Grant, J. A. Daniel, D. Schieltz, J. R. Yates, 3rd, P. A. Grant, Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434-438 (2005).
20. H. Santos-Rosa *et al.*, Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* **12**, 1325-1332 (2003).
21. R. J. Sims, 3rd *et al.*, Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* **280**, 41789-41792 (2005).

22. J. Wysocka *et al.*, A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).
23. C. T. Collins, J. L. Hess, Deregulation of the HOXA9/MEIS1 axis in acute leukemia. *Curr Opin Hematol* **23**, 354-361 (2016).
24. V. Stavropoulou *et al.*, MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer Cell* **30**, 43-58 (2016).
25. H. J. Lawrence *et al.*, Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. *Leukemia* **13**, 1993-1999 (1999).
26. Z. Li *et al.*, Consistent deregulation of gene expression between human and murine MLL rearrangement leukemias. *Cancer Res* **69**, 1109-1116 (2009).
27. R. C. Stong, S. J. Korsmeyer, J. L. Parkin, D. C. Arthur, J. H. Kersey, Human acute leukemia cell line with the t(4;11) chromosomal rearrangement exhibits B lineage and monocytic characteristics. *Blood* **65**, 21-31 (1985).
28. S. Ziemin-van der Poel *et al.*, Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc Natl Acad Sci U S A* **88**, 10735-10739 (1991).
29. Y. Dou *et al.*, Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **13**, 713-719 (2006).
30. A. Patel, V. Dharmarajan, V. E. Vought, M. S. Cosgrove, On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **284**, 24242-24256 (2009).
31. S. A. Shinsky, K. E. Monteith, S. Viggiano, M. S. Cosgrove, Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. *J Biol Chem* **290**, 6361-6375 (2015).
32. S. Rahman *et al.*, Multistate structures of the MLL1-WRAD complex bound to H2B-ubiquitinated nucleosome. *Proc Natl Acad Sci U S A* **119**, e2205691119 (2022).
33. V. Dharmarajan, J. H. Lee, A. Patel, D. G. Skalnik, M. S. Cosgrove, Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J Biol Chem* **287**, 27275-27289 (2012).
34. A. Patel, V. Dharmarajan, M. S. Cosgrove, Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* **283**, 32158-32161 (2008).
35. J. J. Song, R. E. Kingston, WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* **283**, 35258-35264 (2008).
36. A. Patel, V. E. Vought, V. Dharmarajan, M. S. Cosgrove, A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem* **283**, 32162-32175 (2008).
37. A. Patel, V. E. Vought, V. Dharmarajan, M. S. Cosgrove, A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **286**, 3359-3369 (2011).
38. S. A. Shinsky *et al.*, A non-active-site SET domain surface crucial for the interaction of MLL1 and the RbBP5/Ash2L heterodimer within MLL family core complexes. *J Mol Biol* **426**, 2283-2299 (2014).
39. T. A. Milne *et al.*, MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* **10**, 1107-1117 (2002).
40. A. Yokoyama *et al.*, Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* **24**, 5639-5649 (2004).
41. S. H. Wong *et al.*, The H3K4-Methyl Epigenome Regulates Leukemia Stem Cell Oncogenic Potential. *Cancer Cell* **28**, 198-209 (2015).
42. M. G. Guenther *et al.*, Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* **22**, 3403-3408 (2008).
43. A. Gaussmann *et al.*, Combined effects of the two reciprocal t(4;11) fusion proteins MLL.AF4 and AF4.MLL confer resistance to apoptosis, cell cycling capacity and growth transformation. *Oncogene* **26**, 3352-3363 (2007).

44. D. Anastassiou *et al.*, Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC cancer* **11**, 529 (2011).
45. S. Ide *et al.*, MicroRNA-503 promotes tumor progression and acts as a novel biomarker for prognosis in oesophageal cancer. *Anticancer Res* **35**, 1447-1451 (2015).
46. B. Argiropoulos, R. K. Humphries, Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26**, 6766-6776 (2007).
47. S. Bhatlekar, J. Z. Fields, B. M. Boman, Role of HOX Genes in Stem Cell Differentiation and Cancer. *Stem Cells Int* **2018**, 3569493 (2018).
48. C. W. Chen, S. A. Armstrong, Targeting DOT1L and HOX gene expression in MLL-rearranged leukemia and beyond. *Exp Hematol* **43**, 673-684 (2015).
49. R. de Barros E Lima Bueno *et al.*, HOX genes: potential candidates for the progression of laryngeal squamous cell carcinoma. *Tumour Biol* **37**, 15087-15096 (2016).
50. D. H. Spencer *et al.*, Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia* **29**, 1279-1289 (2015).
51. T. C. Somervaille *et al.*, Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* **4**, 129-140 (2009).
52. A. Ramassone, S. Pagotto, A. Veronese, R. Visone, Epigenetics and MicroRNAs in Cancer. *Int J Mol Sci* **19**, (2018).
53. N. Irie *et al.*, SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253-268 (2015).
54. A. S. Bernardo *et al.*, BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144-155 (2011).
55. A. Yevtodiyenko, J. V. Schmidt, Dlk1 expression marks developing endothelium and sites of branching morphogenesis in the mouse embryo and placenta. *Dev Dyn* **235**, 1115-1123 (2006).
56. J. Li, B. P. Zhou, Activation of β-catenin and Akt pathways by Twist are critical for the maintenance of EMT associated cancer stem cell-like characters. *BMC Cancer* **11**, 49 (2011).
57. H. Peinado, D. Olmeda, A. Cano, Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**, 415-428 (2007).
58. X. Ye *et al.*, Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256-260 (2015).
59. A. Cano *et al.*, The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol* **2**, 76-83 (2000).
60. N. Maherali *et al.*, A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3**, 340-345 (2008).
61. S. A. Mani *et al.*, The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715 (2008).
62. C. Scheel, R. A. Weinberg, Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links. *Semin Cancer Biol* **22**, 396-403 (2012).
63. M. C. Brown, C. E. Turner, Paxillin: adapting to change. *Physiol Rev* **84**, 1315-1339 (2004).
64. N. O. Deakin, C. E. Turner, Paxillin comes of age. *J Cell Sci* **121**, 2435-2444 (2008).
65. J. Pignatelli, D. A. Tumbarello, R. P. Schmidt, C. E. Turner, Hic-5 promotes invadopodia formation and invasion during TGF-β-induced epithelial-mesenchymal transition. *J Cell Biol* **197**, 421-437 (2012).
66. D. A. Tumbarello, C. E. Turner, Hic-5 contributes to epithelial-mesenchymal transformation through a RhoA/ROCK-dependent pathway. *J Cell Physiol* **211**, 736-747 (2007).
67. A. T. Thiel *et al.*, MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. *Cancer Cell* **17**, 148-159 (2010).
68. J. H. Lee, D. G. Skalnik, Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol* **28**, 609-618 (2008).

69. M. Wu *et al.*, Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**, 7337-7344 (2008).
70. T. Hoshii *et al.*, A Non-catalytic Function of SETD1A Regulates Cyclin K and the DNA Damage Response. *Cell* **172**, 1007-1021 e1017 (2018).
71. C. H. Pui *et al.*, Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region. *Lancet* **359**, 1909-1915 (2002).
72. R. Terranova, H. Agherbi, A. Boned, S. Meresse, M. Djabali, Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103**, 6629-6634 (2006).
73. B. P. Mishra *et al.*, The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. *Cell reports* **7**, 1239-1247 (2014).
74. T. A. Milne *et al.*, Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol Cell* **38**, 853-863 (2010).
75. Y. Dou, J. L. Hess, Mechanisms of transcriptional regulation by MLL and its disruption in acute leukemia. *Int J Hematol* **87**, 10-18 (2008).
76. C. L. Dobson, A. J. Warren, R. Pannell, A. Forster, T. H. Rabbitts, Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. *Embo J* **19**, 843-851 (2000).
77. R. K. Slany, C. Lavau, M. L. Cleary, The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Mol Cell Biol* **18**, 122-129 (1998).
78. N. L. Alicea-Velazquez *et al.*, Targeted Disruption of the Interaction between WD-40 Repeat Protein 5 (WDR5) and Mixed Lineage Leukemia (MLL)/SET1 Family Proteins Specifically Inhibits MLL1 and SETd1A Methyltransferase Complexes. *J Biol Chem* **291**, 22357-22372 (2016).
79. F. Cao *et al.*, Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell* **53**, 247-261 (2014).
80. D. Vetrie, G. V. Helgason, M. Copland, The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML. *Nat Rev Cancer* **20**, 158-173 (2020).

Supporting Information

Materials and Methods

Generating the MLL1 Win-motif homozygous mutation in human iPS cells

ASE9203 human iPS cells (WT cells) were from Applied Stem Cell (ASC), and CRISPR-Cas9-mediated gene editing was by ASC using their proprietary CRISPR-Cas9 protocol. Puromycin-resistance screening was performed on single cell-derived colonies and two clones (MT1 and MT2) homozygous for the MLL1 Win motif R3765A were identified and sequenced.

Cell Culture

Both WT and MT iPS cells were maintained in mTeSR™1 medium with 1x Supplement media (Stem Cell Technologies, Catalog #85850) and supplemented with 10 μM Y27632 (Stemgent) only upon thawing and passaging. Cells were plated and maintained on Matrigel hESC-qualified Matrix (Corning) according to the Stem Cell Technologies "Maintenance of Human Pluripotent Stem Cells in mTeSR™1" manual. Passaging was with ReLeSR passaging reagent (Stem Cell Technologies). Due to their highly adhesive characteristics, MT cells were periodically gently scraped when passaged to release them from Matrigel-coated wells, and passaged at higher density (1:3 dilution per passage) relative to WT cells (1:10 dilution per passage). Cells were maintained for a maximum of three passages due to growth constraints of MT cells following the third passage.

Tri-lineage Differentiation

WT and MT cells were propagated separately in lineage-specific differentiation media for endoderm, mesoderm or ectoderm using the STEMdiff Trilineage Differentiation Kit Protocol (Stem Cell Technologies). After differentiation, total RNA was extracted from duplicate cultures, and qRT-PCR was performed to evaluate expression of lineage-specific signature markers for endoderm (Sox17), mesoderm (TbxT), or ectoderm (DLK1). Data were normalized to the housekeeping gene (GAPDH), and the ddCt algorithm was used for differential expression analysis and comparison between cell types.

RNA Extraction and Reverse Transcription (RT)-qPCR

WT and MT cells were plated in 6-well Matrigel-coated plates (Corning) until 80-90% confluent. RNA was extracted using Trizol reagent (Invitrogen) according to manufacturer's instructions. Reverse transcription used SuperScript IV Reverse Transcriptase (Invitrogen) using 1 ug total RNA per reaction and 5% of the synthesized cDNA was used in a 20 ul PCR reaction. Real-time qPCR ysed iQ SYBR Green Supermix (Biorad) in a Realplex 4 qPCR 96-Well Real Time Cycler (Eppendorf). C_t values were internally normalized to GAPDH for

each sample, and the ddCt algorithm was used for differential expression analysis and comparison between cell type. Primer sequences are in the "Antibodies and Primers" section.

RNA Sequencing

Whole transcriptome profiling was performed on technical triplicates of WT and MT cells (Clone MT2) at the SUNY Molecular Analysis Core (SUNYMAC) facility at Upstate Medical University. RNA was isolated using Trizol reagent. RNA quality and quantity were assessed with RNA 6000 Nano Kits on an Agilent Bioanalyzer 2100. Sequencing libraries were prepared with the TruSeq Stranded Total RNA Library Prep Kit RiboZero Gold (Illumina), using 1 ug total RNA as input. Library size was assessed with DNA 1000 Kits on an Agilent Bioanalyzer 2100, and libraries were quantified with Qubit dsDNA HS Assay Kits (Invitrogen). Libraries were pooled and sequenced on the NextSeq 500 instrument (Illumina), with paired end 2 x 75 bp reads using a High Output 150 cycle reagent kit.

An average of 54 M paired-end reads per sample was generated from sequencing. Sequencing quality was accessed by fastQC v0.11.8, low-quality bases/reads and adaptors were removed from reads by trimmomatic v0.38. Trimmed reads were mapped to the Gencode GRCh38 release 29 Human reference genome using STAR aligner v2.7.0d [3]. Reads mapped to genes were summarized by the featureCounts program in subread v1.6.3. Genes were filtered by CPM ≥1 in at least 2 samples, and data were normalized to effective library size using edgeR v3.22.5. Differential gene expression analyses were performed using edgeR programs. RNA-seq data have be deposited into the Gene expression Omnibus (GEO) database.

Genes with false discovery rate (FDR) ≤0.05 and fold-change ≥ 2 were considered significantly differentially expressed genes (DEGs) between MT and WT cells. DEGs were investigated for pathway enrichment by Ingenuity Pathway Analysis (IPA, Qiagen) and functional clustering by Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7. GSEA was performed on MT and WT transcriptome expression data using MSigDB database v6.2.

Chromatin Immunoprecipitation

ChIP experiments were performed in duplicate for WT and MT cells (clones MT1 and MTs) using the High Sensitivity ChIP Kit (Abcam). $5x10^5$ cells were used as input material for each immunoprecipitation. Following crosslinking and cell lysis, DNA shearing was performed using Diagenode Bioruptor® Sonicator at high power output, with 5 pulses of 30 seconds, followed by 30 seconds rest between each pulse. 1ug antibody was used for each reaction. Immunoprecipitated DNA was column purified and resuspended in 20ul DNA elution buffer. The purified DNA was either used for sequencing (ChIP-Seq) or qPCR as indicated.

Library Generation, Sequencing and Analysis for ChIP-Seq

ChIP-Seq was performed on technical triplicates of WT and MT cells (Clones MT1 and MT2) at the SUNY Molecular Analysis Core (SUNYMAC) facility at Upstate Medical University. Due to low input material and therefore low yield following ChIP, all purified ChIP DNA was used as input for NEBNext Ultra II Library Prep Kits for Illumina. Library size was assessed with a DNA 1000 Kit on an Agilent Bioanalyzer 2100, and libraries were quantified with Qubit dsDNA HS Assay Kits (Invitrogen). Libraries were pooled and sequenced on a NextSeq 500 instrument (Illumina), with paired end 2 x 75 bp reads using High Output 150 cycle reagent kits. Processing and alignment of ChIP-seq data used Partek Flow analysis software. Reads were aligned using the BWA algorithm followed by MACS2 analysis with q-value=0.05 for peak detection and annotation. Annotated peaks were used to generate TSS plots and for gene enrichment analysis. Peak comparison analysis pipelines were generated with sample attributes specific for treatment (immunoprecipitate vs. IgG), genotype (WT vs. MT) and pair (MT1 and MT2 vs. WT). Gene and pathway enrichment analyses used the KEGG Mapper online analysis tool, and Venn Diagrams were creating using the Bioinformatics & Evolutionary Genomics online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

ChIP-qPCR

Following ChIP, qPCR was performed using iQ SYBR Green Supermix (Biorad) in either a CFX Opus 384 Real-time PCR system (Biorad) or a realplex 4 qPCR 96-Well Real Time Cycler (Eppendorf) with 5% of the immunoprecipitated DNA used in a 20 ul PCR reaction. For both RT and ChIP , Ct values were internally normalized to GAPDH for each sample. For ChIP, fold-enrichment calculations for Setd1A and H3K4 trimethylation ChIP were normalized relative to IgG ChIP controls for each cell type. Each qPCR reaction was performed with technical triplicates. Primers set sequences are in the “Antibodies and Primers” section.

Immunoblot Analysis

Whole cell extracts were prepared by lysing cells in RIPA buffer (50 mM Tris HCl pH 7.4/150 mM NaCl/0.5% sodium deoxycholate/0.1% sodium dodecyl sulfate/1% NP-40) containing cOmplete protease inhibitor cocktail (Roche), followed by quantification using Bradford assays. Approximately 20 ug protein was analyzed by 4%–15% gradient TGX Precast polyacrylamide gel (Bio-Rad), transferred onto a PVDF membrane (Biorad Transblot Turbo system) and probed for specific proteins (see “Antibodies and Primer Sequences” section). PVDF membranes were blocked for 1 hour with a 5% nonfat milk solution and incubated with primary antibody at 4°C overnight. Blots were washed three times with 0.1% TBS/Tween20 and incubated with HRP-conjugated secondary antibody for 1 hour at room temperature. Blots were washed three times and visualized by chemiluminescence (Clarity Western ECL Substrate, Bio-Rad) on a Bio-Rad ChemiDoc MP Imager using the chemiluminescence setting.

Immunofluorescence

Cells on Matrigel-coated glass coverslips were fixed with 4% paraformaldehyde in PBS for 15 min, permeabilized with 1% Triton X-100 in phosphate buffered saline (PBS) for 15 min, quenched with 0.1 M glycine for 15 min and blocked with 3% bovine serum albumin (BSA) for 1 hour at room temperature. Coverslips were incubated with indicated primary antibodies diluted in 3% BSA for 2 hours at 37°C, followed by 1 hour incubation in fluorescent-tagged secondary antibodies as indicated. F-actin was stained with rhodamine-phalloidin (for cytoskeleton) and nuclei were stained with DAPI (Sigma-Aldrich). Cells were imaged using a Zeiss Axioskop2 plus microscope, fitted with a Q imagin ExiBlue charge-coupled device camera using an Apochromat 20Å~ objective.

Motility Assay

For live cell analyses, WT and MT iPS cells (clones MT1 and MT2) were plated in duplicate in 6-well tissue culture plates, coated with Matrigel and tracked using a Nikon TE2000 inverted microscope equipped with a temperature/CO_2 regulated environmental chamber. Images were acquired at 10 min intervals for 14 hours. The manual tracking plugin of ImageJ was used to track the cell centroid over the duration of the movie. The Chemotaxis and Migration plugins of ImageJ were used to calculate velocity. Motility assays were performed on three independent biological replicates.

Invasion Assay

Invasion assays were performed using Corning BioCoat 24-well Invasion Chambers (Corning). WT and MT cells were plated in duplicate in regular growth media (mTESR1 supplemented with 1x Supplement) at 5 x 10^4cells/ml. Chemoattractant (mTESR1 supplemented with 5x Supplement media) was added to Falcon TC Companion plates, and chambers transferred to wells containing the chemoattractant. Cells were incubated overnight for 22 hours in a humidified tissue culture incubator at 37°C and 5% CO_2 atmosphere. After incubation, non-invading cells were scrubbed from the upper surface of the chamber membranes using a cotton-tipped swab. Cells on the surface were washed twice with PBS, fixed with 100% methanol for 20 minutes at room temperature, washed twice with PBS, and stained with Giemsa (Sigma) for 20 minutes at room temperature. After final washes, cells were visualized under a light microscope and photographed with a Nikon microscope camera (FE 3291306). The total number of migrating cells was counted in three random vision fields per sample.

Antibodies and Primers

Primary Antibodies	**Manufacturer, Catalog Number**	**Dilution Factor/Application**
Mouse anti-E-cadherin	BD Biosciences, 610182	1:1000 WB
Rabbit anti-E-cadherin	Cell Signaling, 3195	1:100 IF
Mouse anti-Hic-5	BD Biosciences, 611165	1:1000 WB, 1:100 IF

Rhodamine-phalloidin (1:500, R415; Cat# PHDR1, Denver, CO)	Cytoskeleton, R415 PHDR1	1:500 IF
Rabbit anti-Histone H3	Millipore, 06-755	1:2000 WB
Rabbit anti-H3K4me1	Abcam, ab8895	1:1000 WB
Rabbit anti-H3K4me2	Abcam, ab32356	1:5000 WB
Rabbit anti-H3K4 trimethylation	Active Motif, 39159	1:1000 WB
Rabbit anti-H3K4 trimethylation ChIP Grade	Epicypher, 13-0041	1 ug per ChIP
Rabbit anti-Setd1A ChIP Grade	Abcam, ab70378	1:2000 WB, 1 ug per ChIP sample
Rabbit anti- beta Actin	Abcam, ab8227	1:1000 WB
Mouse anti-Tubulin	Sigma, T5168	1:1000 WB
Rabbit anti-WDR5	Abcam, ab22512	1:1000 WB
Mouse monoclonal anti-MLL1	Millipore, 05-765	1:1000 WB
Rabbit anti-RbBP5	Bethyl Labs, A300-109A	1:1000 WB
Rabbit IgG negative control antibody for ChIP	Thermo Fisher Scientific, 31235	1 ug per ChIP sample
Secondary Antibodies		
Goat anti-mouse DyLight 488-conjugated	Thermo Fisher, 35502	1:250 IF
Goat anti-mouse DyLight 550-conjugated	Thermo Fisher, 84540	1:250 IF
Goat anti-rabbit DyLight 488-conjugated	Thermo Fisher, 35552	1:250 IF
Goat anti-rabbit DyLight 550-conjugated	Thermo Fisher, 84541	1:250 IF
Secondary anti-rabbit IgG, HRP-conjugated	GE, NA934VS	1:30,000
Secondary anti-mouse IgG, HRP-conjugated	GE, NA931VS	1:10,000

Primer Name	**Forward Primer (5'>3')**	**Reverse Primer (5'>3')**
ChIP-qPCR		
ZFP42	ATTGGGCATTCCAGCCTACC	GAGTTCAGCTCCTTGGACCC
ACTIN	AGCCTCGCCTTTGCCGA	GCGCGGCGATATCATCATC
MMP1	ATGCTGAAACCCTGAAGGTG	CTGCTTGACCCTCAGAGACC
MEIS1	TGTGTAAGACGCGACCTGTT	CCGTGCGTGTGTAAAGTGTG
GAPDH	TACTAGCGGTTTTACGGGCG	TCGAACAGGAGGAGCAGAGAGCGA
HOXA9	AGTGGTCTCACTTCCTCCCA	GCTGAGCAGTAAAAGGTGGC
RT-qPCR		
DLK1	GGGCACAGGAGCATTCATAG	GACGGGGAGCTCTGTGATAG
TBXT	GCCCTCTCCCTCCCCTCCACGCACAG	CGGCGCCGTTGCTCACAGACCACAGG

SOX17	CGCTTTCATGGTGTGGGCTAAGGACG	TAGTTGGGGTGGTCCTGCATGTGCTG
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG

MLL Family Win-motif Sequence Confirmation in WT and MT cells		
Human MLL Family Member Gene (Protein)	DNA Coding Sequence	Amino Acid Sequence
KMT2D (MLL2)	atcaaccccactggctgtgcc**cga**tcagagcctaaaatcctc	GC**A**RSE
KMT2C (MLL3)	cgttaaccccacaggttgtgcc**cgt**tctgaacctaaaatgagt	GC**A**RSE
KMT2B (MLL4)	ctgaatccccatggggctgct**cgg**gcagaggtctatctccgg	GA**A**RAE
KMT2F (Setd1A)	gagcaccagacaggctcagcc**cgc**agcgaaggctactacccc	GS**A**RSE
KMT2G (Setd1B)	gagcacgtgacgggctgtgcc**cgc**agtgagggcttctacacc	GC**A**RSE

Figure S1A. Integrity of all five Win-motif sequences for the other human MLL family members was confirmed in both WT and MT cells by next generation sequencing. Win-motif coding sequence is in blue (key arginine coding sequences in bold).

Off Target Locus	Sequence	PAM	Similarity	Mismatch	Gene	Locus
1	GAGCCCTCAGGGCTCAGCCA	CAG	6	3,10	No	chr4@ 182905668-182905691
2	GGACCCTGAGGGCTCAGCCA	AGG	2	2,8,10	No	chr15@ 74395949-74395972
3	GGACCCTGAGGGCTCAGCCA	GGG	2	2,8,10	No	chr11@ 61518499-61518522
4	GTCCCCTCACCGCTCAGCCA	GGG	2	2,3,11	No	chr16@ 689862-689885
5	GACCACTCACGGCTCAGCCT	GGG	1	3,5,20	No	chr11@ 46384434-46384457
6	GACCCCTCAGGGATCAGCCA	AGG	1	3,10,13	No	chr9@ 108810230-108810253
7	GATCCATCACAGCTCAGCCA	CAG	1	3,6,11	No	chr4@ 8432046-8432069
8	GAACCCACAAAGCTCAGCCA	GGG	1	7,10,11	No	chr6@ 52013917-52013940
9	GAGCCCTCACAGCTCAGCCG	GGG	1	3,11,20	No	chr18@ 79803717-79803740
10	GAGCCCTCACCCCTCAGCCA	AGG	1	3,11,12	No	chrY@ 17303630-17303653

Figure S1B. Top 10 predicted CRISPR-Cas9 off-target loci identified by Applied Stem Cells. Next generation sequencing confirmed the integrity of all 10 loci in MT cells.

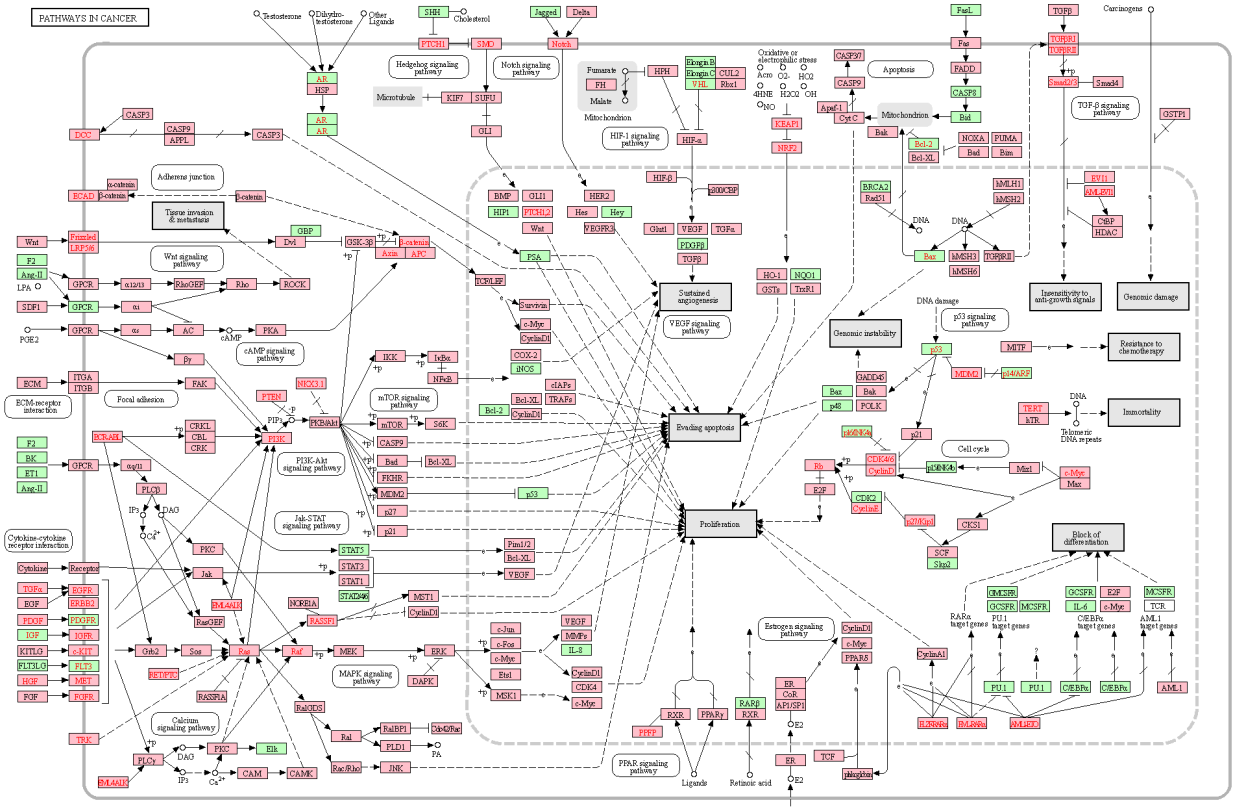


Figure S2. KEGG enrichment analysis of molecules involved in cancer pathways (pink) and represented in overlapping ChIP-Seq peaks in MT1 and MT2 clones normalized to WT cells.

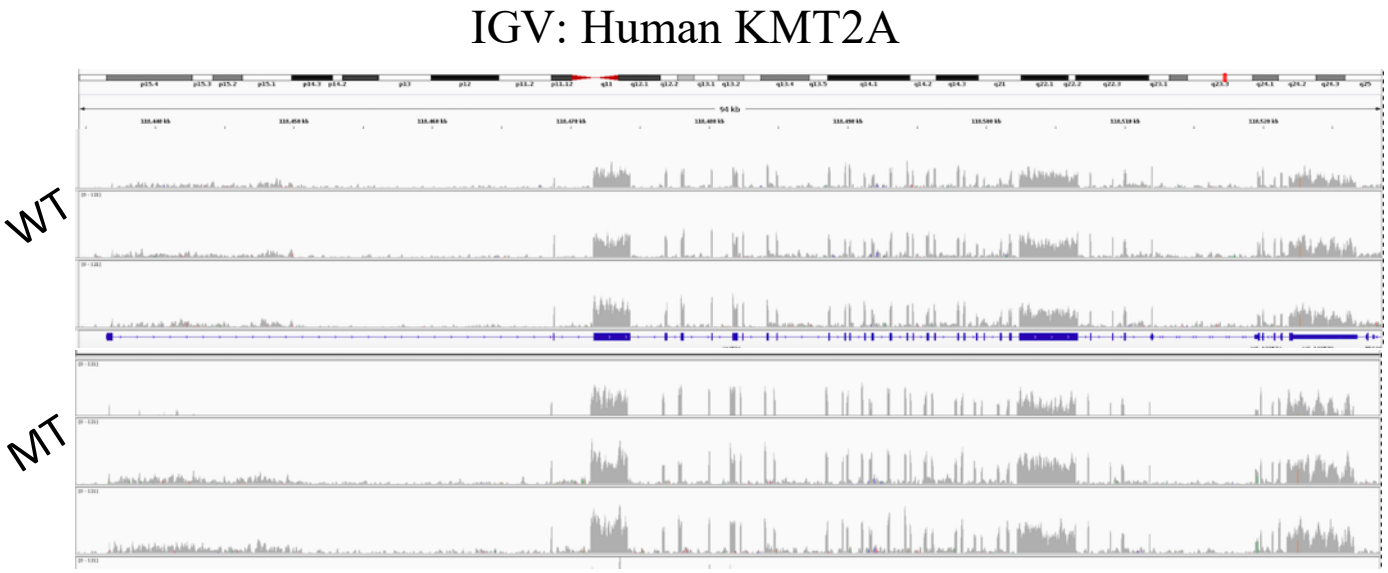
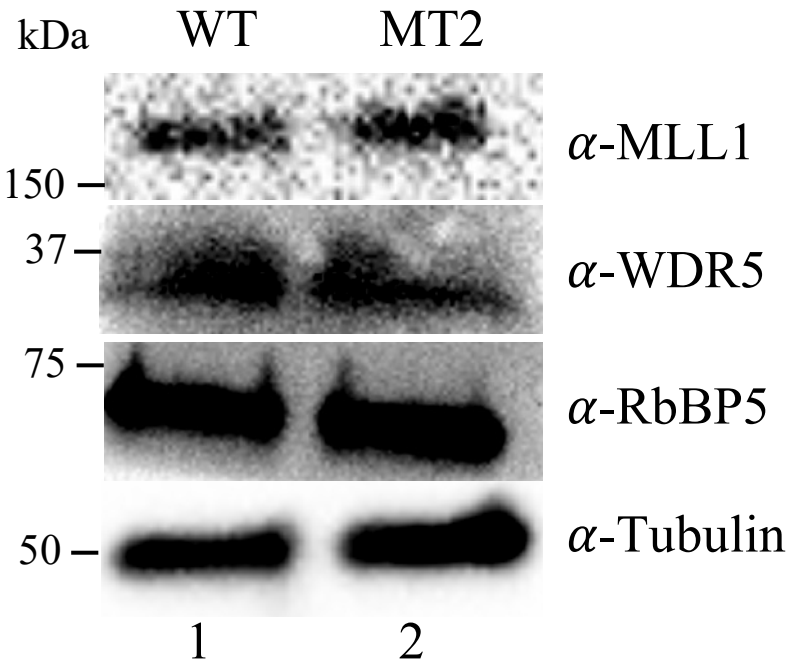


Figure S3A. Integrated genomics view (IGV) of the human KMT2A locus in MT and WT cells, showing identical transcript profile from multiple exons in triplicate RNA-seq samples.



Figures S3B. Protein levels of MWRAD complex components in WT and MT cells. Whole cell extracts from WT (panel 1) and MT (panel 2) cells were immunoblotted using antibodies specific for MLL1, WDR5, and RbBP5. Tubulin was used as a loading control.