Title: Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia

Authors: Derek H. Janssens[1], Michael P. Meers[1], Steven J. Wu[1,2], Ekaterina Babaeva[1], Soheil Meshinchi[3,4], Jay F. Sarthy[1,4], Kami Ahmad[1], & Steven Henikoff[1,5,*]

Affiliations:

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 N. Fairview Ave, Seattle, WA, 98109, USA.

[2]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, 98195, USA.

[3]Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA, USA.

[4]Cancer and Blood Disorder Center, Seattle Children's Hospital, 4800 Sand Point Way, Seattle, WA, 98105, USA.

[5]Howard Hughes Medical Institute, Chevy Chase, MD, USA.

*Address correspondence to steveh@fredhutch.org.

Running title: Chromatin heterogeneity in leukemia

Keywords:

Abstract: Acute myeloid and lymphoid leukemias often harbor chromosomal translocations involving the *Mixed Lineage Leukemia-1* gene, which encodes the KMT2A lysine methyltransferase. The most common translocations produce in-frame fusions of KMT2A to trans-activation domains of chromatin regulatory proteins. Here we develop a strategy to map the genome-wide occupancy of oncogenic KMT2A fusion proteins in primary patient samples regardless of fusion partner. By modifying the versatile CUT&Tag method for full automation we identify common and tumor-specific patterns of aberrant chromatin regulation induced by different KMT2A fusion proteins. Integration of automated and single-cell CUT&Tag uncovers lineage heterogeneity within patient samples and provides an attractive avenue for future diagnostics.

Introduction

Ten percent of acute leukemias harbor chromosomal translocations involving the *Lysine Methyl-transferase 2A* (*KMT2A*) gene (also referred to as *Mixed Lineage Leukemia-1*)[1]. In its normal role, KMT2A catalyzes methylation of the K4 residue of the histone H3 nucleosome tail and is required for fetal and adult hematopoiesis[2]. The N-terminal portion of KMT2A contains a low complexity domain that mediates protein-protein interactions, an AT-hook/CXXC domain that binds DNA, and multiple chromatin-interacting domains (PHD domains and a bromo domain), whereas the C-terminal portion contains a trans-activation domain that interacts with histone acetyl-transferases and a SET domain that catalyzes histone H3K4 methylation[3,4]. The KMT2A pre-protein is cleaved to form a 320-kDa N-terminal fragment (KMT2A-N) and a 180-kDa C-terminal fragment (KMT2A-C) that form a stable dimer[5,6].

KMT2A contributes to leukemogenesis through oncogenic chromosomal rearrangements involving the DNA-binding domain in the N-terminal portion of KMT2A with a diverse array of other chromatin regulatory proteins[7,8]. Although more than 80 translocation partners have been identified in *KMT2A*-rearranged (*KMT2A*r) leukemias, fusions involving AF9, ENL, ELL, AF4 or AF10 transcriptional elongation factors account for the majority of cases[1,8]. These fusion partners regulate RNA Polymerase II (RNAPII) elongation (AF9, ENL, ELL and AF4) or recruit the Dot1L-H3K79 histone methyltransferase (ENL, AF9 and AF10)[9-12]. Additionally, ENL and AF9 interact with the CBX8 chromobox protein to neutralize the PRC1 gene silencing complex[13-16].

Previous work has suggested that KMT2A fusion proteins bind different genomic loci depending on the fusion partner to drive different leukemia subtypes[17,18]. For example, AF4 fusions are more common in acute lymphoid leukemia (ALL), and AF9 fusions are associated with acute myeloid leukemia (AML)[1]. In addition, *KMT2A* rearrangements are also prevalent in mixed lineage leukemia (MPAL), and numerous examples of tumors that interconvert between lineage types have been documented[17,19-21]. However, because methods for efficiently and reliably profiling KMT2A fusion binding sites in patients samples are lacking, the relationship between KMT2A fusions, associated chromatin proteins, leukemia subtypes, and lineage plasticity has been challenging to fully characterize. Here, we establish a chromatin profiling platform that efficiently profiles oncogenic fusion proteins, transcription-associated complexes, and histone modifications in cell lines and patient samples. By integrating these results with related single-cell methods we characterize the regulatory dynamics of *KMT2A*r leukemias and find that distinct fusion partners display differential affinity for various transcriptional cofactors and may influence lineage plasticity.

Results

A strategy for mapping the binding sites of diverse KMT2A fusion proteins.

Characterizing the chromatin localization of oncogenic fusion proteins has often been limited by the inability of ChIP-seq to be used with small amounts of patient samples. To efficiently compare the binding sites for wildtype KMT2A and the fusion proteins, we applied AutoCUT&RUN[22] across a panel of four *KMT2A*r leukemia cell lines and five primary *KMT2A*r patient samples sorted for CD45-positive blasts. This collection spans the spectrum of *KMT2A*r leukemia subtypes with diverse *KMT2A* translocations that create oncogenic fusion proteins with the transcriptional elongation factors AF4 (SEM, RS4;11, 1^0 ALL-1 and 1^0 MPAL-2), AF9 (1^0 MPAL-1), ENL (KOPN-8, 1^0 AML-2), AF6 (ML-2), or a relatively rare fusion to the cytoplasmic GTPase Sept6 (1^0 AML-1) (**Supplementary Table 1**). With the exception of ML-2, these samples also contain a wildtype copy of the *KMT2A* locus. Antibodies to the C-terminal portion recognize only wildtype KMT2A-C, while antibodies to the N-terminal portion recognize both wildtype KMT2A-N and the fusion protein (**Fig. 1a**). Therefore, binding sites unique to the oncogenic fusion protein can be identified by comparing chromatin profiling of C-terminal and N-terminal KMT2A antibodies. We used an automated CUT&RUN platform to profile replicate samples of cell lines with two different antibodies to the N-terminus and two to the C-terminus of KMT2A, and correlation analysis of sequencing results showed high reproducibility (**Supplementary Figure 1**).

By examining the profiling landscapes we identified many sites where both the N-terminus and the C-terminus of KMT2A coincide in both H1 hESCs and *KMT2A*r leukemia cells (**Fig. 1b**). Other sites are apparent where only the N-terminus of KMT2A is detected and only in leukemia cells: these must be sites of fusion protein binding (**Fig. 1b**). To define fusion protein binding sites, we used Gaussian mixture modeling to partition KMT2A peaks into two different distributions in the total enrichment-normalized ratio of N-terminus to C-terminus KMT2A signal (KMT2A N/C score) (**Fig. 1c**). In the control H1 cells, all of the KMT2A N/C scores fall within a single Gaussian distribution and so a two-component model fails to partition the data, whereas in each of the *KMT2A*r leukemia samples two-component Gaussian mixture modeling identifies one group of sites with appreciably higher N/C scores than the other, allowing us to set a threshold to call oncogene target sites (**Fig. 1c, Supplementary Fig. 2a-h**). For example, in the MPAL-1 cell line 7,264 sites are identified as binding KMT2A-N and KMT2A-C, whereas 1,517 sites are enriched only for the N-terminus of KMT2A and are therefore called as oncogene binding sites (**Fig. 1d).** While ~60% of full-length KMT2A binding sites coincide with gene

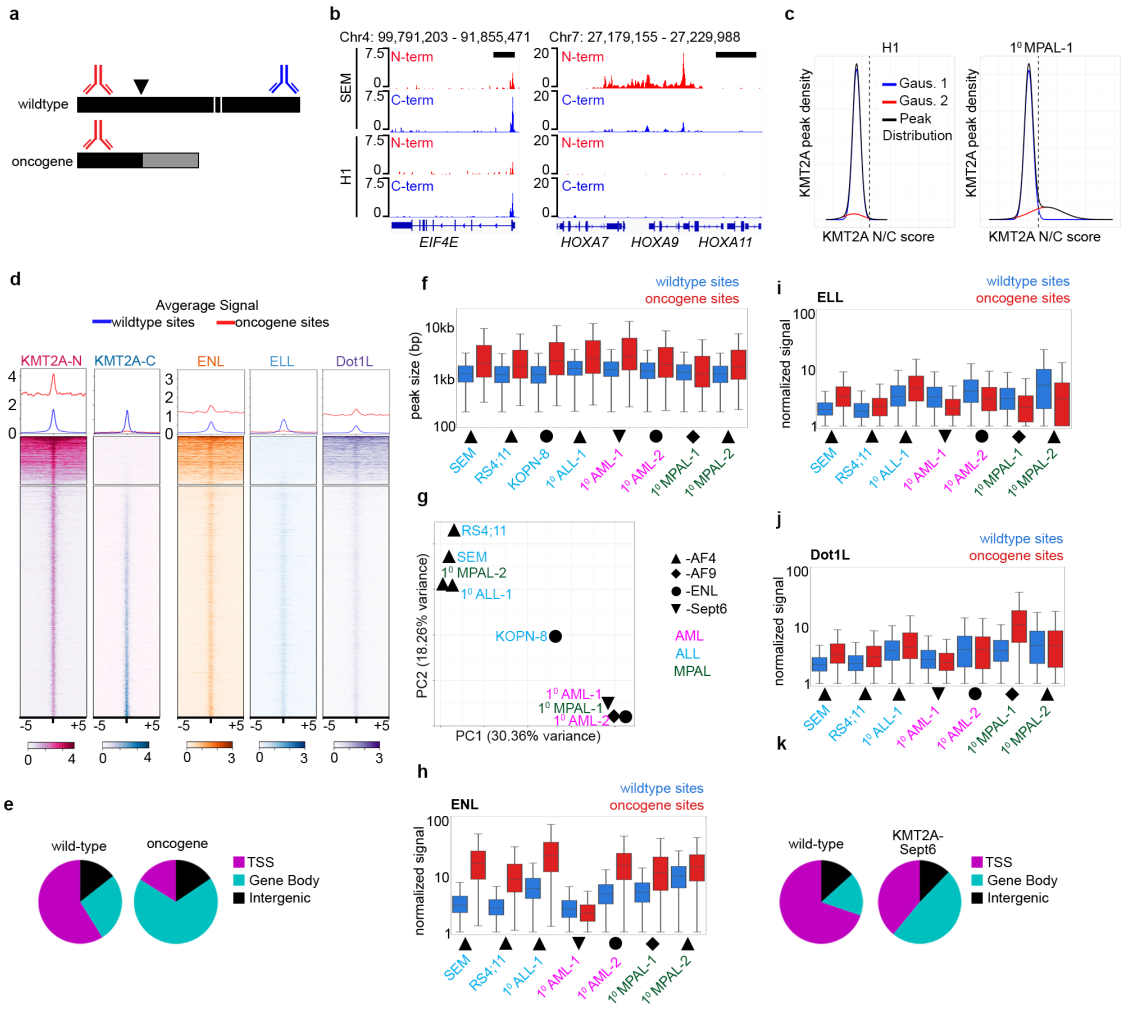


Figure 1. AutoCUT&RUN profiling of KMT2A-fusion protein binding. a, A general strategy for mapping KMT2A-fusion proteins. The wildtype KMT2A (black) is cleaved (white lines) into KMT2A-N and KMT2A-C proteins. Common oncogenic lesions (black arrowhead) result in in-frame translation of oncogenic KMT2A with numerous fusion partners (grey). C-terminal KMT2A antibodies (blue) recognize wildtype KMT2A-C. N-terminal KMT2A antibodies (red) recognize wildtype KMT2A-N and the oncogene KMT2A-fusion proteins. **b,** Example of a wildtype KMT2A binding site *(EIF4E)* and an oncogene binding site *(HOXA locus)*. Black scale bars = 10kb. **c,** A two component Gaussian distribution model of the ratio of KMT2A N-terminal enrichment to C-terminal enrichment (N/C score) identifies only one gaussian distribution in control H1 (left), but identifies two distinct distributions in *KMT2A*r leukemia samples (right) and was used for thresholding (dotted line) to call KMT2A oncogene binding sites. **d,** Oncogene (top) and wildtype binding sites (bottom), in the primary MPAL-1 sample. The KMT2A-fusion protein is widespread at these sites, while the KMT2A-C signal is largely absent. In the primary MPAL-1 tumor oncogene binding sites are enriched for ENL and Dot1L but not ELL. **e,** KMT2A-fusion oncogenes are enriched in gene bodies. **f,** KMT2A-fusion oncogenes bind over broad domains. **g,** PCA of oncogene binding sites in *KMT2A*r samples. The first two components are shown. **h**, Enrichment of ENL at fusion protein binding sites. **i**, Relative enrichment of ELL at fusion protein binding sites. **j**, Enrichment of Dot1L at fusion protein binding sites. **k,** The KMT2A-Sept6 fusion oncogene in the primary AML-1 sample is less enriched in gene bodies than the other fusion proteins.

promoters, ~70% of fusion protein binding sites occupy gene bodies and often span broad -domains up to 10 kb over transcribed regions (**Fig. 1e,f**). This pattern of oncogenic KMT2A fusion localization is consistent with previous reports[18,23].

Only ~1% of fusion protein binding sites are shared across leukemia samples, and these shared sites represent 6% of total sequence space (506kb/8377kb) bound by *KMT2A*r in any cell type (**Supplementary Table 2**). The group of 15 genes that is targeted by the fusion protein in all *KMT2A*r leukemia samples is highly enriched for master regulators of hematopoiesis as well as genes that are required for *KMT2A*r leukemia[24-28]. Interestingly, several of the shared KMT2A oncogene targets had not been investigated as downstream mediators of leukemia. By examining the DepMap database of CRISPR-Cas9 screens targeting *KMT2A*r leukemia cell lines[29,30], we identified *SENP6* and *ARID2* as oncoprotein targets that are novel dependencies in KMT2Ar leukemia (**Supplementary Table 2**).

Principal Component Analysis (PCA) of oncogene binding sites indicates that the specific partner in each fusion protein likely influences the tumor-specific localization. KMT2A-AF4 samples cluster together and include both an ALL patient sample as well as an MPAL patient sample (**Fig. 1g**). The ALL cell line KOPN-8 carries a KMT2A-ENL fusion and has distinct oncogene binding sites. In contrast, despite the primary AML-1, AML-2, and MPAL-1 samples producing distinct KMT2A fusion proteins, the oncogene binding profile is similar. This suggests that KMT2A fusion partners are not the sole determinants of oncogene landscapes, and that lineage-specific features of the chromatin landscape also contribute to oncogene targeting.

We reasoned that the distinct binding sites of KMT2A fusion proteins may in part be driven by the unique cofactors with which the fusion partners associate. Therefore, we mapped the distributions of ENL, ELL, and Dot1L, three chromatin proteins that have previously been shown to interact with KMT2A fusion proteins[31]. Overall, regions bound by KMT2A fusion proteins are also highly enriched for ENL in most of the samples we profiled, but are only slightly enriched for ELL binding in three of the four KMT2A-AF4 fusion lines (**Fig. 1 h,i; Supplementary Fig. 2b-e**).

Whereas Dot1L has been proposed to be a central component of oncogenic transformation by a variety of KMT2A fusion proteins, this histone methyltransferase is most enriched at the oncogene binding sites of the primary MPAL-1 patient sample (**Fig. 1d, j**). This leukemia carries a KMT2A-AF9 fusion protein, and AF9 is normally a component of the DotCom complex[32]. Thus, as for the ENL and ELL transcriptional elongation factors, the interaction of the oncogenic KMT2A fusion partner with an elongation-coupled histone methyltransferase appears to drive localization to gene bodies (**Supplementary Fig. 1i**). Finally, the 1^0 AML-1 sample harbors a relatively rare KMT2A-Sept6 fusion, and this leukemia cell line has a distinctive set of fusion protein binding sites and lacks the characteristic wide spreading of fusion protein into

transcribed genes (**Fig. 1k**). This suggests that the Sept6 fusion protein is mechanistically distinct from other leukemias we profiled. Thus, our profiling strategy has successfully distinguished loci of fusion protein mislocalization from endogenous sites that can be used to delineate differences between *KMT2Ar* cell lines that contribute to leukemogenesis.

Mapping chromatin features of KMT2A fusion protein binding sites.

We next aimed to characterize chromatin features around the fusion protein binding sites that we had identified in each *KMT2A*r cell line. To do this economically and at a scale that could be generally applied to patient samples, we developed AutoCUT&Tag, a modification of our previous AutoCUT&RUN robotic platform[22]. CUT&Tag takes advantage of the high efficiency and low background of antibody-tethered Tn5 tagmentation-based chromatin profiling relative to previous methods, such as ChIP-seq and CUT&RUN[33]. The standard CUT&Tag protocol requires DNA extraction before library enrichment by PCR. However, we recently developed conditions for DNA release and PCR enrichment without extraction (CUT&Tag-Direct)[34]. In this improved protocol a low concentration of SDS is sufficient to displace bound Tn5 from tagmented DNA, and the subsequent addition of the non-ionic detergent Triton-X100 is sufficient to quench the SDS and allow for efficient PCR. This streamlined protocol makes CUT&Tag compatible with robotic handling of samples in a 96-well format in a single plate and generates profiles with data quality comparable to those produced by benchtop CUT&Tag (**Supplementary Fig. 3**).

To define the chromatin features around KMT2A fusion protein binding sites, we used AutoCUT&Tag to profile the active histone modifications H3K4me1, H3K4me3, and H3K36me3, and the silencing histone modifications H3K27me3 and H3K9me3. Together, these five histone modifications distinguish active promoters, enhancers, transcribed regions, developmentally silenced, and constitutively silenced chromatin[35], and provide a straightforward picture of the regulatory status of a genome (**Supplementary Fig. 3c**). Replicate profiles for each mark in leukemia cell lines were very similar and were merged for further analysis (**Supplementary Fig. 4**).

We first examined the chromatin features of wildtype KMT2A and oncoprotein binding sites. Consistent with KMT2A catalyzing trimethylation of the H3K4 residue, binding sites for wildtype KMT2A are heavily marked with H3K4me3, whereas oncogene binding sites are relatively depleted for H3K4me3 (**Fig. 2a**). This difference in chromatin marking supports the observation that oncogenes bind at new sites in the genome without the accompanying wildtype methyltransferase. Interestingly, at a limited subset of broad KMT2A-AF4, KMT2A-ENL, and

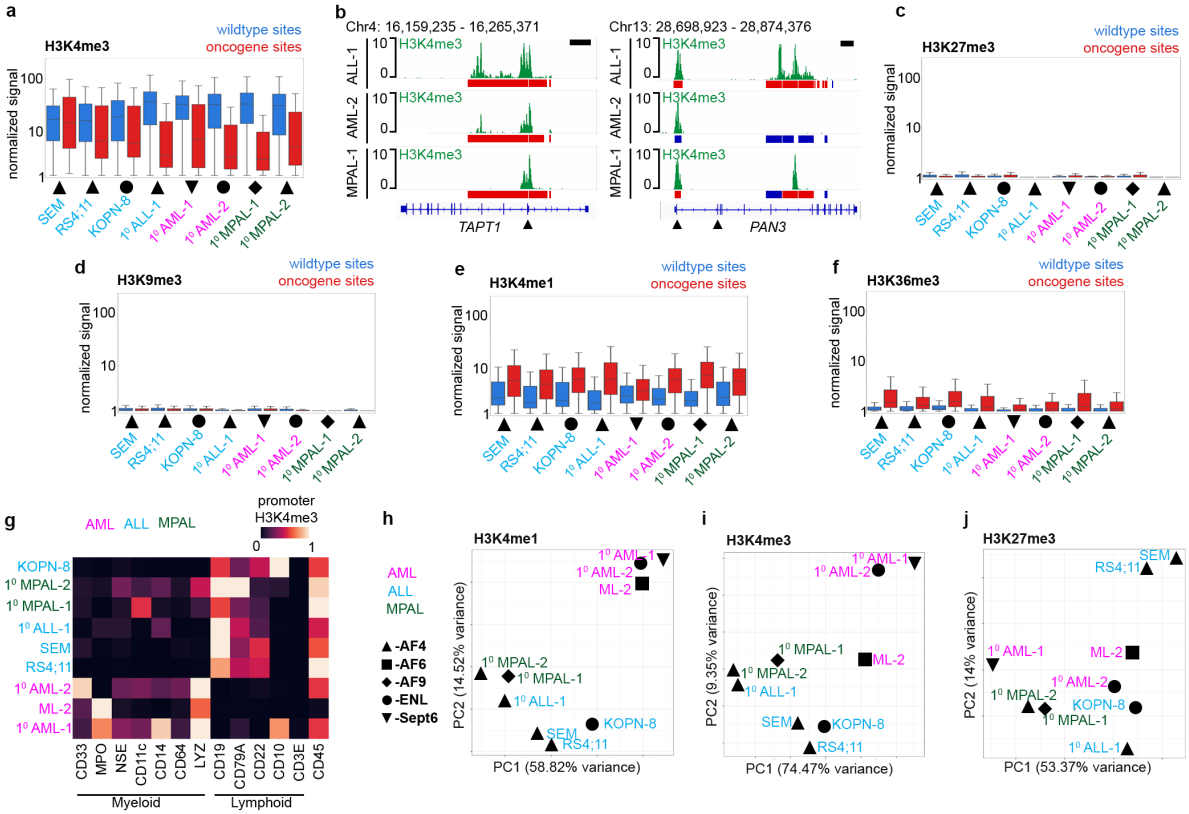


Figure 2. Chromatin features of KMT2A fusion-protein binding sites. a, The KMT2A fusion protein lacks the H3K4me3 methyltransferase domain, and this mark is depleted at oncogene binding sites relative to wildtype binding at promoters. **b**, Genome browser tracks showing relatively rare examples of aberrant H3K4me3 enrichment away from the TSS (Arrowhead) at fusion binding sites (red bars). H3K4me3 is enriched in the gene body of *TAPT1* in the primary ALL-1 sample (KMT2A-AF4) and the primary AML-2 sample (KMT2A-ENL) but not the primary MPAL-1 sample (KMT2A-AF9). H3K4me3 is enriched in the *PAN3* gene body in ALL-1 and MPAL-1. In AML-2 this regions is called a wildtype KMT2A binding site (blue bars). Black scale bars = 10kb. **c,** Both the wildtype and KMT2A fusions bind at sites of active chromatin devoid of H3K27me3-marked developmentally repressed domains. **d**, The wildtype and KMT2A fusions binding sites lack H3K9me3, which marks constitutively repressed heterochromatin.**c,** The oncogene is enriched at distal regions marked by H3K4me1, and actively transcribed gene bodies. **e,** The oncogene is enriched at distal regions marked by H3K4me1. **f,** The KMT2A fusion protein is enriched in actively transcribed gene bodies marked by H3K36me3. **g,** H3K4me3 signal at the promoters of diagnostic cell surface markers accurately classifies AML, ALL and MPAL leukemias. **h,** PCA of genome-wide H3K4me1 signal separates samples by leukemia subtype. **i**, PCA of genome-wide H3K4me3 signal. **j**, PCA of genome-wide H3K27me3 signal distinguishes two different ALL groups and separates the primary AML-1 sample from primary AML-2 and ML-2.

KMT2A-AF9 binding sites we see that H3K4me3 is deposited away from the gene promoter (**Fig. 2b**). This suggests that these oncogenic fusions may direct the aberrant localization of an alternative H3K4-methyltransferase under certain circumstances. Oncoprotein binding sites lack H3K27me3 or H3K9me3 (**Fig. 2c,d**), but are enriched in H3K4me1 and H3K36me3, both of which mark transcribed gene bodies (**Fig. 2e,f**). Such enrichment of gene-body marks is as expected for mis-targeting of transcriptional elongation complexes via KMT2A fusions[31].

Histone modification profiling holds the potential to reveal similarities and distinctions between leukemias by reporting their transcriptional regulome status. For example, H3K4me3 reports activity of gene promoters, and the signal for this modification at blood cell marker gene

promoters resembles the immunophenotype characteristic of each leukemia (**Fig. 2g**). Correlation matrices for each histone mark across the genome showed the greatest variance in H3K4me1, H3K4me3 and H3K27me3 modifications (**Supplementary Fig. 4**), and we examined these profiles in more detail. We first identified leukemia enriched regions for each modification using the SEACR peak-calling method[36], and performed PCA to determine modification-specific similarities between samples. Overall, ALL, AML, and MPAL leukemias clustered together by their H3K4me1 and H3K4me3 profiles (**Fig. 2h,i**), consistent with similar repertoires of lineage-specific transcriptionally active regions in each leukemia type. In contrast, PCA based on profiling of H3K27me3 partitions samples into groups largely unrelated to their leukemia subtype, suggesting that there are distinctions between leukemias in silenced regions (**Fig. 2j**). H3K27me3 is an epigenetically inherited histone modification that is linked to developmental progression as cells determine their identities. Thus, these distinct H3K27me3 leukemia landscapes may be indicative of the hematopoietic transitions that are defective in each tumor.

Clustering of regulatory elements between leukemias.

To identify common groups of regulatory elements that are shared between leukemia subtypes we compiled merged lists of H3K4me1 peaks from all samples, quantified the sample-specific signal over all peaks (**Figure 3a**) and performed t-distributed stochastic neighbor embedding (t-SNE) of these elements, followed by density peak clustering (**Figure 3e**)[37]. A majority of features are shared across all leukemias ("common"), while a smaller number are specific to each sample. As expected, we also identified groups of H3K4me1-enriched regions in AML ("myeloid elements"), ALL ("lymphoid elements"), or shared between AML and MPAL ("mixed myeloid elements"), or ALL and MPAL ("mixed lymphoid elements"). Thus, this regulatory analysis implies that MPAL leukemias share features with both ALL and AML.

PCA analysis indicates that the other histone modifications we profiled are also able to distinguish between *KMT2A*r leukemias and so we extended our t-SNE and clustering analysis to identify groups of regions enriched for H3K4me3, H3K36me3, and H3K27me3 that are shared between *KMT2A*r leukemia subtypes. Most H3K4me3 and H3K36me3 peaks are common across leukemias, indicating that they largely share gene expression repertoires (**Figs. 3b-c)**. Grouping H3K4me3-marked promoter regions by t-SNE also identified myeloid, lymphoid, mixed myeloid and mixed lymphoid elements (**Figs. 3f**), however as compared to H3K4me1, a smaller proportion of H3K4me3 marked features show any lineage specificity. This is consistent with previous reports that regulatory elements marked by H3K4me1 generally show more cell-type specificity than the promoter elements marked by H3K4me3[38,39]. H3K27me3 peaks show

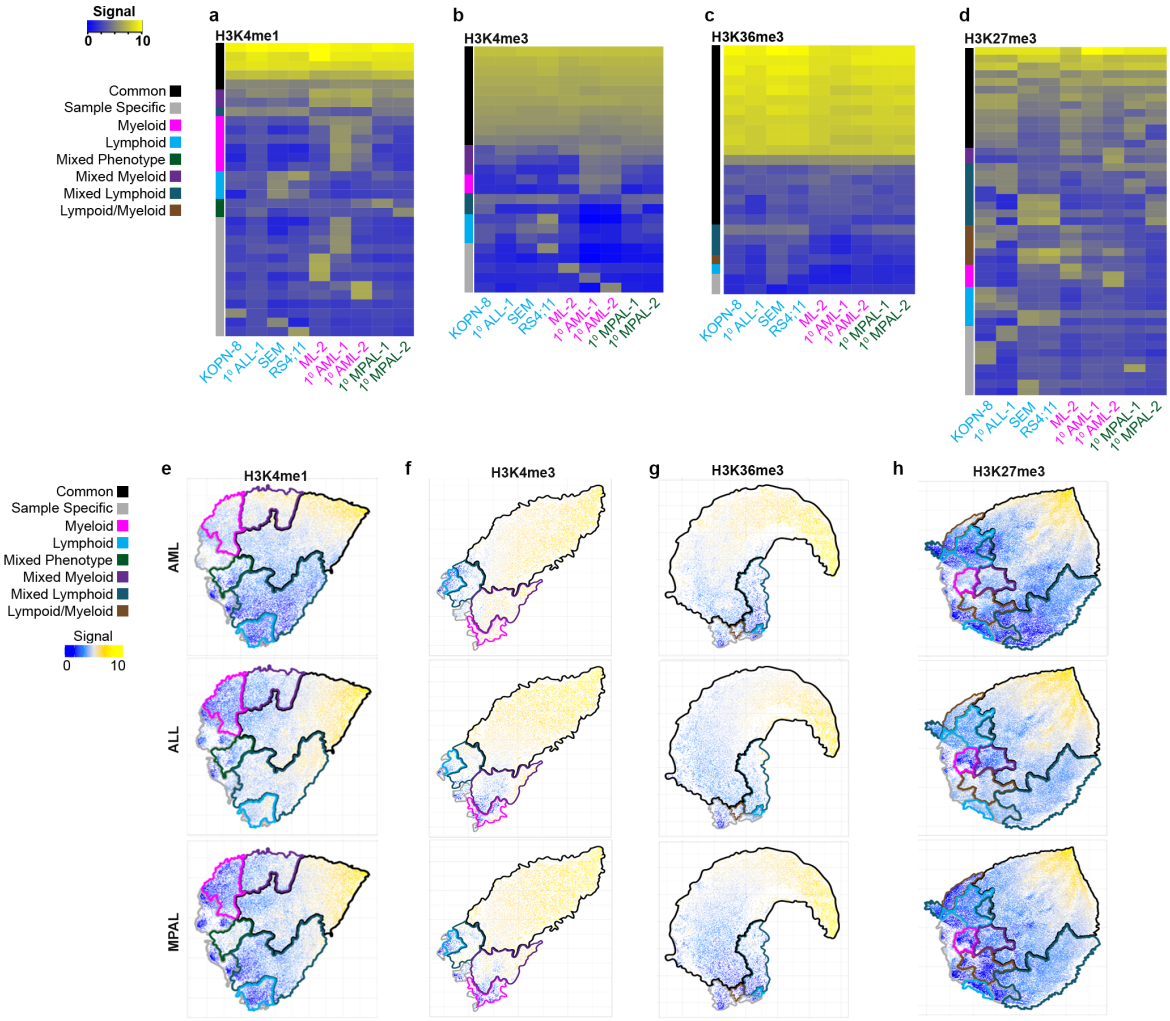


Figure 3. Clustering regulatory features distinguishes common and restricted elements in leukemia samples. a, Clustering analysis separates the H3K4me1-marked regions into 32 groups, and the heatmap shows the average signal intensity of H3K4me1 in each of these groups (y-axis) for each *KMT2A*r leukemia sample (x-axis). The colors alongside the heatmap show the subtype designation of each group as common across samples (black), mixed myeloid (plum), mixed lymphoid (teal), myeloid (magenta), lymphoid (cyan), mixed phenotype (green), or sample specific (grey). **b,** Same as (a) for H3K4me3 clustered into 25 groups. **c,** Same as (a) for H3K36me3 clustered into 25 groups. **d,** Same as (a) for H3K27me3 clustered into 45 groups. **e,** Two-dimensional t-SNE projections separate lineage-specific H3K4me1 marked regions. Each colored pixel corresponds to a single H3K4me1 peak, and t-SNE defined clusters are outlined to show lineage specificity, where outline colors correspond to subtype specific groups of elements (a). H3K4me1 enriched regions (yellow) in AML (top row; magenta outline) are separated from the regions that show ALL specific enrichment (middle row; cyan outline). MPAL specific regions (bottom row; green outline) fall in the middle. **f**, Same as (e) for H3K4me3. **g**, Same as (e) for H3K36me3. **h**, Same as (e) for H3K27me3. H3K27me3 marked regions enriched in ALL fall on separate sides of the t-SNE graph, whereas AML and MPAL H3K27me3 marked regions fall in the middle.

diversity similar to H3K4me1 (**Fig. 3d**). As suggested by the H3K27me3 PCA analysis, t-SNE of the H3K27me3 developmentally repressed landscape is uniquely able to subdivide lymphoid specific elements and mixed lymphoid specific elements into two spatially separated groups (**Fig. 3h**). We conclude that high-throughput CUT&Tag profiling of active and repressed chromatin landscapes provides a powerful tool to characterize *KMT2A*r leukemias, and that

profiling the developmentally repressed genome reveals tumor-specific differences that are not apparent by profiling the active genome.

Integration of autoCUT&Tag with scCUT&Tag reveals tumor heterogeneity.

Given our ability to identify regulatory regions that discriminate leukemia types, we reasoned that the heterogeneous usage of those elements within the same leukemia might underlie the phenotypic plasticity of *KMT2A*r leukemia. To test this, we performed CUT&Tag on single *KMT2A*r leukemia cells, where antibody binding and pA-Tn5 tethering is performed on bulk samples, and then individual cells are arrayed on an ICELL8 platform for barcoded PCR library enrichment[33]. After optimizing the SDS and Triton X-100 inputs to CUT&Tag-Direct for single

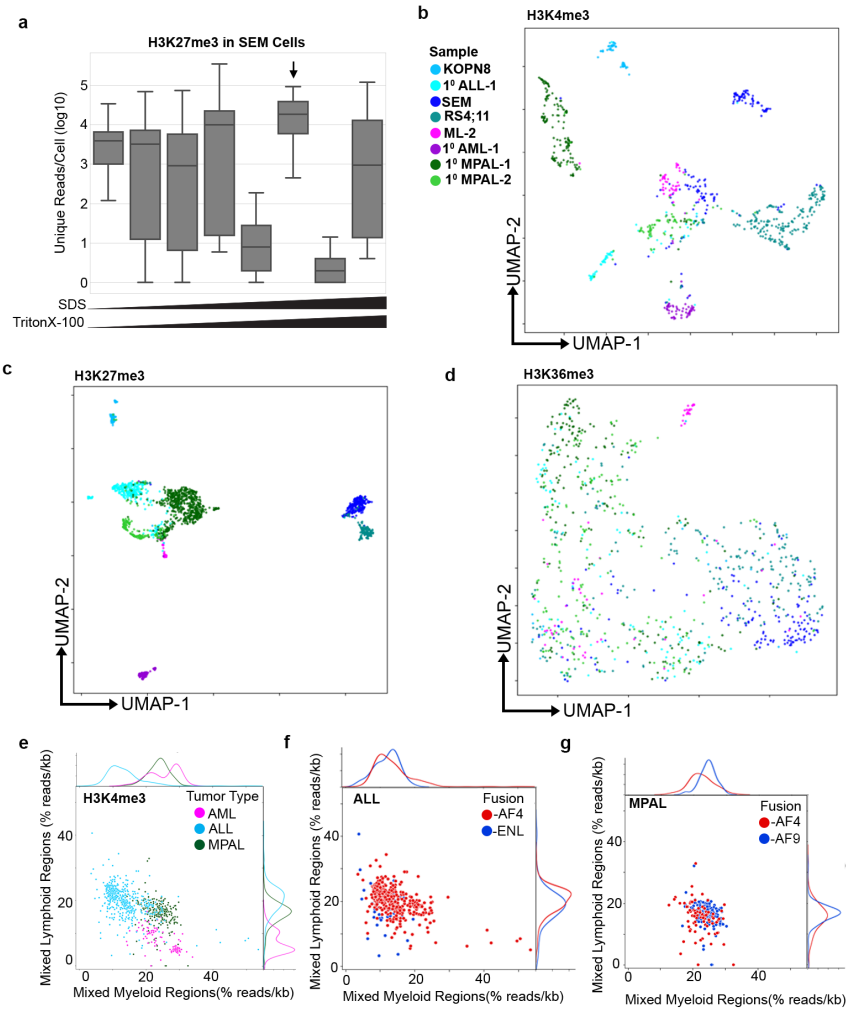


Figure 4. Histone methylation-specific subtype clustering of single cells in leukemia samples. **a**, Titrating the concentration of SDS and Triton-X in the nanowell increases the library yield for individual cells, and identifies optimum conditions for Tn5 release and PCR enrichment (Arrow). **b**, UMAP scatterplot of H3K4me3 histone modification patterns in single leukemia cells. Hierarchical clustering was performed using the groups of H3K4me3 features defined by bulk profiling experiments. **c**, Same as (b) for H3K27me3. **d**, Same as (b) for H3K36me3. **e**, Scatterplot of read counts in lineage-specific H3K4me3 promoter modules. Single cells are colored by leukemia subtype and positioned according to the percent of cellular reads that fall in either the mixed myeloid or mixed lymphoid regions as defined by bulk H3K4me3 profiling. Density plots indicate the distribution of cells from samples characterized as AML, ALL or MPAL **f**, Same as (e), but showing only the ALL cells, colored according the KMT2A fusion partner in the sample. **g**, Same as (e), but showing only the MPAL cells, colored according the KMT2A fusion partner in the sample.

cell applications, we were able to increase the median number of unique reads per cell from -~6000 to ~24,000 (**Fig. 4a**), while maintaining a high fraction of reads in peaks (**Supplementary Fig. 5a**).

To examine the cellular heterogeneity of active and repressed chromatin in *KMT2A*r leukemia we applied this modified protocol to profile between 1137-3611 cells from our collection of samples for the H3K4me3, H3K27me3, and H3K36me3 histone modifications. After cells with fewer than ~300 fragments were excluded, single-cell CUT&Tag for H3K4me3, H3K36me3, and H3K27me3 yielded a median of 4972, 3962, and 13025 unique reads per cell, respectively (**Supplementary Fig. 5b**). Profiles for each single cell were first binned across the groups of regulatory features we identified by clustering analysis of our bulk profiling data (**Supplementary Fig. 5c,d,e**), and cells were projected in UMAP space based on that binning (**Fig. 4b-d**). Discrete sample-specific clusters were resolved by UMAP projection of cells profiled for H3K4me3 and H3K27me3 (**Fig. 4b,c**) but not H3K36me3 (**Fig. 4d**), indicating that the differences in the H3K4me3 and H3K27me3 landscapes of *KMT2A*r leukemia cells of the same samples are generally less than the differences between samples.

To directly examine whether individual cells in *KMT2A*r leukemia samples show differential enrichment for H3K4me3 at lineage-specific elements, we compared the percentage of fragments within individual cells for each leukemia type that fell within the myeloid- or lymphoid-enriched features as defined by bulk CUT&Tag profiling (**Supplementary Fig. 5f**). Whereas the ALL and AML cells generally exhibit mutually exclusive enrichment for mixed lymphoid or myeloid elements, respectively, the majority of MPAL cells are enriched for both (**Fig. 4e**). Interestingly, the distribution of cells in the lymphoid-myeloid space differs between samples defined by different fusions. A small subset of KMT2A-AF4 ALL cells exhibit bias toward myeloid features (**Fig. 4f**), and cells in the primary MPAL-2 sample containing KMT2A-AF4 are more dispersely arrayed than cells from the KMT2A-AF9 containing primary MPAL-1 (**Fig. 4g**). This suggests that KMT2A-AF4 containing leukemias may have greater lineage plasticity than the other KMT2A fusions proteins we profiled. Thus, single-cell CUT&Tag profiling is able to resolve heterogenous lineage biases within primary pediatric leukemia samples providing a powerful tool for these cancers.

Discussion

Here we have applied high-throughput chromatin profiling to *KMT2A*r leukemias to delineate fusion protein-specific targets and to identify chromatin features that are characteristic of myeloid, lymphoid and mixed-lineage leukemias. To profile these features with high signal-to-

noise requiring only low sequencing depths for maximum economy, we modified CUT&Tag-direct[34] for full automation in 96-well format on a standard robot. As CUT&Tag-direct requires only hundreds to thousands of cells for informative histone modifications[40], AutoCUT&Tag is suitable for profiling of samples for a wide range of studies, including developmental and disease studies and screening patient samples.

By also performing AutoCUT&RUN on KMT2A fusions and components of the SuperElongation and DotCom complexes we have elucidated mechanistic details that likely contribute to the heterogeneity of these tumors. First, we found the most common KMT2A-fusion proteins, including KMT2A-AF4, KMT2A-AF9 and KMT2A-ENL all colocalize with the ENL protein in gene bodies, whereas a relatively rare KMT2A-Sept6 fusion protein does not colocalize with ENL and also tends to be more tightly associated with promoters. This suggests that the interaction of the C-terminal domain of AF4, ENL and ELL with transcriptional elongation complexes likely recruits the fusion protein from the promoter into the gene-body. Consistent with the possibility that these interactions play a pivotol role in oncogenic transformation, the wildtype ENL allele is required for tumor growth in numerous *KMT2A*r cell lines[41].

How do KMT2A fusion proteins promote lineage plasticity in *KMT2A*r leukemia? AF4 is the most common KMT2A fusion partner in pediatric leukemias[1,7], and KMT2A-AF4 fusions are associated with lineage switching, where an ALL at diagnosis presents as AML upon tumor relapse[17,21]. Our single cell CUT&Tag profiling data suggests that of the samples included in our study KMT2A-AF4 leukemias are also likely to be the most plastic since they contain the most diverse set of active regulatory elements.

The enhanced throughput and consistency of the AutoCUT&RUN and AutoCUT&Tag platforms for chromatin profiling makes these technologies suitable for profiling patient specimens. Recent advances in our understanding of *KMT2A*r leukemia has allowed for the development or repurposing of numerous targeted compounds as therapeutics[42,43], and the set of 14 genes bound by the oncoprotein across *KMT2A*r leukemias may represent promising potential therapeutic targets. Incorporating AutoCUT&RUN and AutoCUT&Tag into longitudinal clinical trials will provide unprecedented resolution to assess the efficacy of novel epigenetic medicines. In addition these technologies are extremely scalable and cost effective, meaning the information obtained from these trials could also be used to apply chromatin profiling for patient diagnosis in the future.

Data Accession Gene Expression Omnibus GSEXXXXX

Author contributions

DHJ and SH optimized the CUT&Tag method for automation, and DHJ adapted these modifications for single cell CUT&Tag profiling. SM provided clinical samples and helpful discussion. DHJ, JFS, KA, and SH designed experiment. DHJ, EB and JFS performed experiments. DHJ, MPM, SJW and JFS performed data analysis. DHJ, MPM, KA, and SH wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the Fred Hutchinson Genomics Shared Resource Facility for technical support, particularly Phil Corrin and Jeff Delrow for help with AutoCUT&RUN profiling of KMT2A. We thank Terri Bryson and Trizia Llagas for help with cell culture, and Jorja Henikoff and Matt Fitzgibbon for preparing the sequencing data for analysis. In addition, we thank Jitendra Thakur for helpful discussions related to data analysis and presentation. We thank Charles Mullighan, Jenny Lill and Marie Bleakley for generously sharing the *KMT2A*r samples and cell lines used in this study. This work was supported by NIH grants R01 HG010492 (S.H.), 4DN TCPA A093 (S.H.) and F32 GM129954 (M.M.), by the Howard Hughes Medical Institute (S.H.), by a pilot project grant from the Chan-Zuckerberg Initiative (S.H.), by a Damon Runyon-Sohn Foundation Fellowship (J.F.S.) and by an Alex's Lemonade Stand Foundation Young Investigator Award (J.F.S.).

Methods

Cell Culture: Human K562 cells were purchased from ATCC (Manassas, VA, Catalog #CCL-243) and cultured according to supplier's protocol. H1 hESCs were obtained from WiCell (Cat# WA01-lot# WB35186) and cultured in Matrigel™ (Corning) coated plates in mTeSR™1 Basal Media (STEMCELL Technologies cat# 85851) containing mTeSR™1 Supplement (STEMCELL Technologies cat# 85852). The *KMT2A*r cell lines ML-2, KOPN-8, RS4;11 and SEM were obtained from the Bleakley lab at the Fred Hutchinson Cancer Research Center.

Primary Patient Samples: Cryopreserved CD45 leukmia blasts for primary MPAL-1 (Sample ID: SJMPAL012424_D1, Alias TB-11-3295) and primary ALL-1 (Sample ID: SJALL048347_D1, Alias TB-13-0939) were obtained from St. Jude Children's Research Hospital in accordance with institutional regulatory practices. Cryopreserved CD45 leukmia blasts for primary AML-1

(Sample ID: A40725), primary AML-2 (Sample ID: A67194) and primary MPAL-2 (Sample ID: A58548) were obtained from the Meshinchi lab at the Fred Hutchinson Cancer Research Center. Diagnosis of clinical samples as ALL, AML or MPAL was based on flow cytometry of samples stained with CD45-APC-H7 (BD Cat# 560178), cytoplasmic CD3-PE (BD Cat# 347347), CD34-PerCP Cy5.5 (BD Cat# 347203), CD18-APC (BD Cat# 340437), cytoplasmic MPO-FITC (Dako Cat# F071401-1), and CD33-PE-Cy7 (BD Cat# 333946). The KMT2A fusion present in each sample was determined by RNA-sequencing.

Antibodies: For profiling the wildtype and oncogenic KMT2A protein we used two monoclonal antibodies targeting the KMT2A N-terminus: Mouse anti-KMT2A (1:100, Millipore Cat #05-764) refered to as KMT2A-N1, and Rabbit anti-KMT2A (1:100, Cell Signaling Tech Cat #14689S) refered to as KMT2A-N2; as well as two monoclonal anitbodies targeting the KMT2A C-terminus: Mouse anti-KMT2A (1:100, Millipore Cat #05-765) refered to as KMT2A-C1, and Mouse anti-KMT2A (1:100, Santa Cruz Cat #sc-374392) refered to as KMT2A-C2. Since pA-MNase does not bind efficiently to many mouse antibodies, we used a rabbit anti-Mouse IgG (1:100, Abcam Cat# ab46540) as an adapter; this antibody was also used in the absence of a primary antibody as the IgG negative control. For profling the SEC and Dotcom components via manual and and AutoCUT&Tag we used rabbit anti-ENL (Cell Signaling Tech Cat# 14893S), rabbit anti-ELL (Cell Signaling Tech Cat# 14468S) and rabbit anti-Dot1L (Cell Signaling Tech Cat# 90878S). For profiling histone marks via manual and AutoCUT&Tag, as well as single-cell CUT&Tag we used Rabbit anti-H3K4me1 (1:100 Thermo Cat# 710795), Rabbit anti-H3K4me3 (1:100 for bulk profiling or 1:10 for single-cell experimetns, Active Motif Cat# 39159), Rabbit anti-H3K36me3 (1:100 for bulk profiling or 1:10 for single-cell experimetns, Epicypher Cat# 13-0031), Rabbit anti-H3K27me3 (1:100 for bulk profiling or 1:10 for single-cell experimetns, Cell Signaling Technologies Cat# 9733S), and Rabbit anti-H3K9me3 (1:100, Abcam Cat# ab8898). To increase the local concentration of pA-Tn5, all CUT&Tag reactions also included the secondary antibody Guinea Pig anti-Rabbit IgG (1:100, antibodies-online Cat# ABIN101961).

AutoCUT&RUN: Primary patient samples were thawed at room temperature, washed and bound to Concanavalin-A (ConA) paramagnetic beads (Bangs Laboratories Cat# BP531) for magnetic separation. Samples were then suspended in Antibody Binding Buffer and split for incubation with either the KMT2A N- or C-terminus specific antibodies or the IgG control antibody overnight. Sample processing was performed by the CUT&RUN core facility at the

Fred Hutchinson Cancer Research Center according to the AutoCUT&RUN protocol available through the Protocols.io website (dx.doi.org/10.17504/protocols.io.ufeetje).

CUT&Tag: Manual CUT&Tag reactions were performed according to the CUT&Tag-Direct protocol[34]. Briefly, nuclei were prepared by suspending cella in NE1 Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 0.5mM Spermidine, 0.1% Triton X-100, 20% Glycerol) for 10 min on ice. Samples were then spun down and resuspended in Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, Roche Complete Protease Inhibitor EDTA-Free) and lightly cross-linked by addition of 16% fomaldehyde to 0.1%. After 2 min, cross-linking was stopped by addition of 2.5 M glycine to a final concentration of 75 mM. Nuclei were washed and either cryopreserved in a Mr. Frosty Chamber for long term storage, or bound to ConA magnetic beads for further processing. ConA-bound nuclei were suspended in Antibody Binding Buffer (Wash Buffer containing 2 mM EDTA) and split into individual 0.5 mL tubes for antibody incubation at room temperature for 1 hr or 4°C overnight. Samples were then washed to remove unbound primary antibody, brought up in Wash buffer containing the secondary antibody, and incubated at 4°C for 1 hr. Samples were then washed and brought up in 300-Wash Buffer (Wash Bufer with 300 mM NaCl), containing pA-Tn5 (1:150 dilution), and incubated at 4°C for 1 hr. Samples were then washed in 300-Wash Buffer, and brought up in Tagmentation Buffer (300 Wash Buffer plus 10 mM MgCl2), and incubated at 37°C for 1 hr to allow the Tn5 tagmentation reaction to go to completion. Samples were then washed with TAPS wash buffer (10 mM TAPS with 0.2 mM EDTA), and brought up in 5 µL of Release Solution (10 mM TAPS with 0.1% SDS). Samples were then incubated in a thermocycler with heated lid at 58 degrees for 1 hr to release Tn5 and prepare tagmented chromatin for PCR. Neutralizing Solution (15 µL 0.67% Triton-X100) was added followed by 2 µL barcoded i5 primer (10 µM), 2 µL barcoded i7 primer (10 µM) and 25 µL of NEBNext PCR mix. Samples were then placed in a thermocycler and PCR amplification was performed using 12-14 rapid cycles. CUT&Tag libraries were then cleaned up with a single round of SPRIselect beads at a 1.3 : 1 v/v ratio of beads to sample, quantified on a Tapestation bioanalyzer instrument and pooled for sequencing.

AutoCUT&Tag: A detailed protocol complete with program downloads has been made publicly available on protocols.io for implementing AutoCUT&Tag on a Beckman Coulter Biomek liquid handling robot (https://www.protocols.io/view/autocut-amp-tag-streamlined-genome-wide-profiling-bgztjx6n). To facilitate adaptation of the method to other standard liquid handling modules, the complete specifications for each step in the automated procedure are outlined in

guidelines section. Briefly, nuclei were extracted, lightly cross-linked, bound to ConA beads and incubated with primary antibody as in manual CUT&Tag. Up to 96 samples were then arrayed in a 96 well PCR plate and positioned on a a stationary ALP on the Beckman Coulter Biomek FX Robot equipped with an ALPAQUA Magent Plate for standard magnetic separation, an ALPAQUA LE Magent Plate for low volume elution, and a thermal block for temperature controlled inbuation. Wash Buffer and 300-Wash Buffer were loaded in Deep Well Plates, Secondary Antibody Solution, pA-Tn5 solution, Tagmentation Buffer, TAPS Buffer and Release Buffer were all loaded into V-Bottom Plates and were positioned on Stationary ALPs in accordance with the preprogrammed AutoCUT&Tag method. The AutoCUT&Tag processing was conducted over the course of 4 hours. The sample plate containing ConA-bound tagmented nuclei in 10 µL 0.1% SDS was then removed, sealed and placed on a thermocycler with heated lid for a 1 hour incubation at 58°C. Using a reservoir and multichannel pipettor, 54 µL of 0.15% SDS neutralization solution was added to each well, followed by 4 µL of premixed i5/i7 barcoded primers, and 36 µL of premixed KAPA PCR Master Mix. The plate was then sealed and returned to a thermocycler for 14 rapid PCR cycles. Following PCR amplification, the sample plate was returned to the Biomek for one round of post-PCR cleanup on the Biomek deck set up in accordance a preprogrammed post-PCR cleanup method, including a second 96-well plate preloaded with SPRISelect Ampure beads, a Deep Well Plate loaded with 80% Ethanol for bead washes, and two V-Bottom Plates preloaded with 10 mM Tris-HCl pH 8.0 for tip washes and elution. Upon completion of the 1 hr cleanup the samples were then quantified using a Tapestation bioanalyzer instrument and pooled for sequencing.

Single-cell CUT&Tag: Nuclei were extracted and lightly cross-linked using the same strategy as for manual CUT&Tag. The nuclei concentration was then quantified to allow for accurate dilution prior to dispensing into nanowells on the ICELL8. For each antibody 10 µL of ConA beads were washed in Binding Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$) and bound to the sample for 10 min. Samples were the split into 0.5 mL Lobind tubes, one for each antibody, and resuspended in 25 µL of Antibody Buffer containing primary antibody at a 1:10 dilution. Samples were incubated at 4°C overnight, washed twice with 100 µL of Wash Buffer, and then resuspended in 50 µL Wash Buffer containing secondary antibody at a 1:50 dilution. Samples were incubated at 4°C for 1 hr, washed twice with 100 µL of Wash Buffer, and then resuspended in 50 µL 300-Wash Buffer with 1:50 diltuion of pA-Tn5. Samples were incubated at 4°C for 1 hr, washed 2X with 100 µL of 300-Wash Buffer, and then resuspended in 50 µL of Tagmentation Solution (300-Wash Buffer with 10 mM $MgCl_2$). Samples were incubated

at 37°C in a thermocycler with heated lid for 1 hr to allow the tagmentation reaction to go to completion. Samples were washed with 10 mM TAPS to remove any residual salt, and then resuspended in 10 mM TAPS pH8.5 containing 1X DAPI and 1X secondary diluent reagent (Takara Cat# 640196) at a concentration of 400 nuclei/µL. 80 µL of cell suspension was loaded into 8 wells of the 384 cell plate, together with 25 µL of the fiducial reagent (Takara Cat# 640196) according to the manufacturer's instructions. Sample suspension (35 nL) was dispensed on the ICELL8 into the nanowells of a 350v Chip (Takara Cat# 640019). The 350v Chip was dried and sealed, and cells were centrifuged at 1200xg for 3 min. The Chip was then imaged to identify wells containing a single nuclei and a filter file was prepared. During image processing, 35 nL of 0.19% SDS in TAPS was added to all nanowells on the ICELL8 using an unfilitered dispense. The Chip was then dried, sealed and centrifuged at 1200xg for 3 min and then heated at 58°C in a thermocycler with heated lid for 1 hr to release the pA-Tn5 and prepare the tagmented chromatin for PCR. Before opening, the Chip was centrifuged at 1200xg, and 35 nL of 2.5% Triton-X100 neutralization solution was added to all wells containing a single nuclei via a filtered dispense on the ICELL8. The Chip was then dried and 35 nL of i5 indices was added via a filtered dispense. The Chip was then dried and 35 nL of i7 indices was added via a filtered dispense. The Chip was then dried, sealed and centrifuged at 1200xg for 3 min. Then 100 nL of KAPA PCR mix (2.775 X HiFi Buffer, 0.85 mM dNTPs, 0.05 U KAPA HiFi polymerase / µL)(Roche Cat# 07958846001) was added to all wells containing a single nucleus via two 50 nL filtered dispenses. The Chip was centrifuged at 1200xg for 3 min, sealed and placed in a thermocycler for PCR amplification using the following conditions: 1 cycle 58 ^{0}C 5 min; 1 cycle 72 ^{0}C 10 min; 1 cycle of 98 ^{0}C 45 sec; 15 cycles of 98 ^{0}C 15 sec, 60 ^{0}C 15 sec, 72 ^{0}C 10 sec; 1 cycle 72 ^{0}C 2 min. The Chip was then centrifuged at 1200xg for 3 min into a collection tube (Takara Cat# 640048). To remove residual PCR primers and detergent, the sample was then cleaned up using two rounds of SPRISelect Ampure bead cleanup at a 1.3 : 1 v/v ratio of beads to sample. Samples were resuspended in 30 uL of 10 mM Tris-HCL pH 8.0, quantified on a Tapestation bioanalyzer instrument, and pooled with bulk samples for sequencing.

DNA sequencing and Data processing: The size distributions and molar concentration of libraries were determined using an Agilent 4200 TapeStation. Up to 48 barcoded CUT&RUN libraries or 96 barcoded CUT&Tag libraries were pooled at approximately equimolar concentration for sequencing. Paired-end 25 × 25 bp sequencing on the Illumina HiSeq 2500 platform was performed by the Fred Hutchinson Cancer Research Center Genomics Shared Resources. This yielded 5-10 million reads per antibody. Single-cell CUT&Tag libararies were

prepared using unique i5 and i7 barcodes and pooled with bulk samples for sequencing. For 500-100 cells 20 million reads was sufficient to obtain an average of approximately 80% saturation of the estimated library size for each single cell. Paired-end reads were aligned using Bowtie2 version 2.3.4.3 to UCSC HG19 with options: --end-to-end --very-sensitive --no-mixed --no-discordant -q --phred33 -I 10 -X 700.

Identifying *KMT2A*r oncoprotein targets: To identify unique *KMT2A*r targets, we first generated a merged set of 18087 SEACR peaks originating from either N-terminal or C-terminal KMT2A antibody-targeted CUT&RUN in any cell type assayed. We quantified the number of fragments mapping to each peak i from each dataset j, and summed reads mapped from the two antibodies targeting the same KMT2A terminus in the same dataset to yield N-terminal (n_{ij}) and C-terminal (c_{ij}) fragments mapped in each peak, existing in cell type sets Nj and Cj, respectively. We calculated the cell type-specific "N over C ratio" (NCR) for each peak as follows:

(1) $NCR_{ij} = \log_{10}((((n_{ij})+\min(N_j))/((c_{ij})+\min(C_j)))*ECDF((N_j+C_j)/2))((n_{ij}+c_{ij})/2))$,

where min(x) = minimum value of x across the peak set; and ECDF(y)(x) = Empirical Cumulative Distribution Function of set y evaluated at x, as implemented in R (https://www.r-project.org/) using the ecdf() function. As illustrated in equation (1), ECDF was used to shrink NCR values towards zero in inverse proportion with the mean $n_{ij}+c_{ij}$ signal observed in the peak. Each peak-cell type combination P_{ij} was assigned a True or False value for peak identity and *KMT2A*r identity. Peak identity was asserted as True and added to peak set P_j if the mean $n_{ij}+c_{ij}$ was at least 1.96 standard deviations above the mean N_j+C_j. For all P_j, *KMT2A*r identity was evaluated by fitting a two-component Gaussian Mixture Model to all NCRj corresponding to P_j, and asserting as True any NCR_{ij} that are greater than the NCR value greater than the mean of NCRj at which the two fitted Gaussian distributions intersect. Gaussian Mixture Modeling was implemented in R using the normalMixEM() function from the "mixtools" library. For all peaks assigned as *KMT2A*r in any cell type, NCR scores were hierarchically clustered using the hclust() function in R on a euclidean distance matrix generated by the dist() function.

t-SNE embedding of the active and repressed chromatin regions: For histone modification data, peaks were called from merged replicate datasets using SEACR[36], and peak sets were merged for each modification across all cell types. We generated matrices of raw read counts mapping in each cell type (columns) to merged peaks (rows) for each modification, and we

filtered out instances were counts were lower than any count value whose evaluated Empirical Cumulative Distribution Function was more than 5% diverged from the predicted ECDF value based on a lognormal fit of the data distribution, using the fitdistr() function from the MASS library with “densfun” set to “lognormal”. We then $\log_{10}$-transformed the results and rescaled columns to z-scores. Principal component analysis (PCA) was performed on the resulting transformed matrices using the prcomp() function in R. For t-SNE analysis, all principal components contributing greater than 1% variance were used as input to the Rtsne() function from the Rtsne library, with perplexity set as the nearest integer to the square root of the number of peaks, and check_duplicates set as FALSE. We used the resulting two-dimensional t-SNE values as input to the densityClust() function from the densityClust library, and used that output in the findClusters() function, with rho and delta values set to the 95th percentile of all rho and delta values output from densityClust(), respectively. To generate cluster average heatmaps, scaled count values were averaged by cluster and the resulting matrix was used as input to the heatmap.2() function from the gplots library. PCA and t-SNE plots were generated using the ggplot2 library (https://ggplot2.tidyverse.org/).

UMAP embedding of single cells: Cluster-specific regions defined from the bulk t-SNE embeddings were used to generate a single-cell count matrix of N features (Fig 3). These matrices were then normalized by sequencing depth. Next, each feature in the matrix (row) was scaled by subtracting the mean and dividing by the standard deviation (z-scaling). The upper and lower bound values of the matrix were capped at ±1.5[44]. We then used hierarchical clustering on the count-matrix to organize features and cells. While clustering, we confined single cells to each celltype so no cells could be organized outside of their celltype category. In addition, the normalized count-matrix was reduced from N dimensions to two dimensions using UMAP and plotted.

Comparison of myeloid and lymphoid enriched H3K4me3 signal in single cells: We quantified the number of unique fragments from each cell that fell in the myeloid, mixed myeloid, lymphoid, mixed lymphoid, common or lineage specific clusters of H3K4me3 peaks as defined by analysis of bulk data. The number of unique fragments that fell in each cluster was then divided by the number of base pairs covered by the set of peaks in a given cluster. The percent of bp normalized myeloid, mixed myeloid, lymphoid, and mixed lymphoid fragments for each cell was then determined relative to the total bp normalized signal in all peaks.

Preparation of Figure Panels: All heat maps were generated using DeepTools[45]. All of the data were analyzed using either bash or python (https://github.com/python).. The following packages were used in python: Matplotlib, NumPy, Pandas, Scipy, and Seaborn.

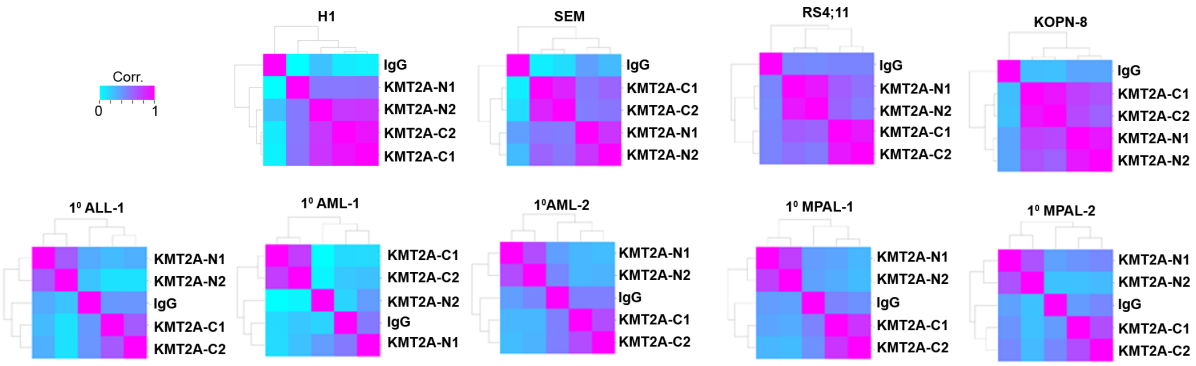
References

1. Winters, A.C. & Bernt, K.M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4 (2017).
2. Rao, R.C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* **15**, 334-46 (2015).
3. Zeleznik-Le, N.J., Harden, A.M. & Rowley, J.D. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proc Natl Acad Sci U S A* **91**, 10610-4 (1994).
4. Slany, R.K. The molecular biology of mixed lineage leukemia. *Haematologica* **94**, 984-93 (2009).
5. Hsieh, J.J., Cheng, E.H. & Korsmeyer, S.J. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**, 293-303 (2003).
6. Hsieh, J.J., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S.J. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* **23**, 186-94 (2003).
7. Slany, R.K. The molecular mechanics of mixed lineage leukemia. *Oncogene* **35**, 5215-5223 (2016).
8. Hu, D. & Shilatifard, A. Epigenetics of hematopoiesis and hematological malignancies. *Genes Dev* **30**, 2021-2041 (2016).
9. Smith, E., Lin, C. & Shilatifard, A. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev* **25**, 661-72 (2011).
10. Monroe, S.C. *et al.* MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia. *Exp Hematol* **39**, 77-86 e1-5 (2011).
11. Zeisig, D.T. *et al.* The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin. *Oncogene* **24**, 5525-32 (2005).
12. Okada, Y. *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* **121**, 167-78 (2005).
13. Maethner, E. *et al.* MLL-ENL inhibits polycomb repressive complex 1 to achieve efficient transformation of hematopoietic cells. *Cell Rep* **3**, 1553-66 (2013).
14. Tan, J. *et al.* CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **20**, 563-75 (2011).
15. Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780 (2016).
16. Douillet, D. *et al.* Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat Genet* **52**, 615-625 (2020).
17. Lin, S. *et al.* Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 737-749 (2016).
18. Prange, K.H.M. *et al.* MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene* **36**, 3346-3356 (2017).

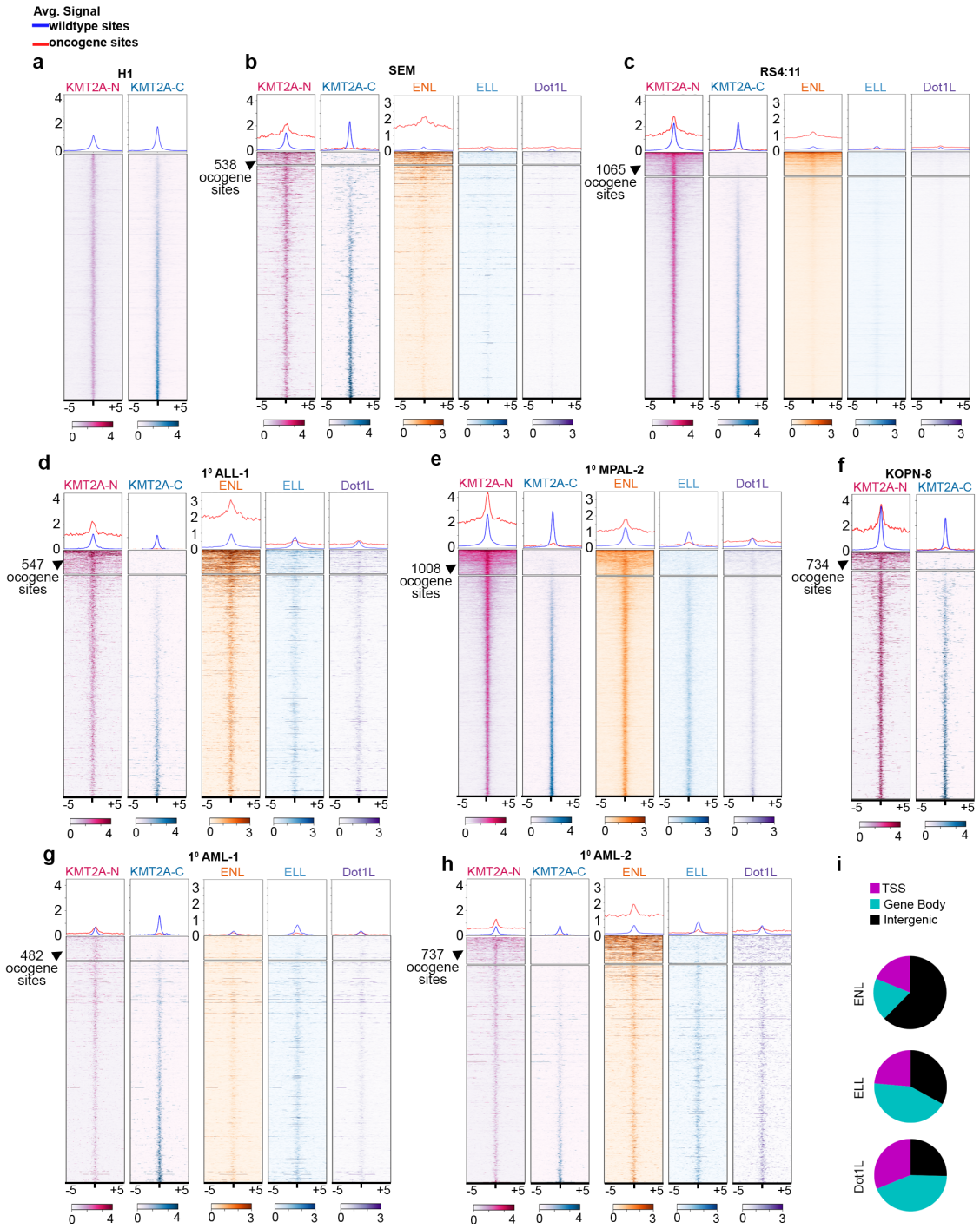
19. Alexander, T.B. *et al.* The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373-379 (2018).
20. Andersson, A.K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* **47**, 330-7 (2015).
21. Rayes, A., McMasters, R.L. & O'Brien, M.M. Lineage Switch in MLL-Rearranged Infant Leukemia Following CD19-Directed Therapy. *Pediatr Blood Cancer* **63**, 1113-5 (2016).
22. Janssens, D.H. *et al.* Automated in situ chromatin profiling efficiently resolves cell types and gene regulatory programs. *Epigenetics Chromatin* **11**, 74 (2018).
23. Guenther, M.G. *et al.* Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* **22**, 3403-8 (2008).
24. Izaguirre-Carbonell, J. *et al.* Critical role of Jumonji domain of JMJD1C in MLL-rearranged leukemia. *Blood Adv* **3**, 1499-1511 (2019).
25. Itskovich, S.S. *et al.* MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia. *Nat Commun* **11**, 2369 (2020).
26. Wong, P., Iwasaki, M., Somervaille, T.C., So, C.W. & Cleary, M.L. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes Dev* **21**, 2762-74 (2007).
27. Zeisig, B.B. *et al.* Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol* **24**, 617-28 (2004).
28. Cante-Barrett, K., Pieters, R. & Meijerink, J.P. Myocyte enhancer factor 2C in hematopoiesis and leukemia. *Oncogene* **33**, 403-10 (2014).
29. Behan, F.M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511-516 (2019).
30. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e16 (2017).
31. Chan, A.K.N. & Chen, C.W. Rewiring the Epigenetic Networks in MLL-Rearranged Leukemias: Epigenetic Dysregulation and Pharmacological Interventions. *Front Cell Dev Biol* **7**, 81 (2019).
32. Kuntimaddi, A. *et al.* Degree of recruitment of DOT1L to MLL-AF9 defines level of H3K79 Di- and tri-methylation on target genes and transformation potential. *Cell Rep* **11**, 808-20 (2015).
33. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
34. Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* (2020).
35. Zhang, J. *et al.* An integrative ENCODE resource for cancer genomics. *Nat Commun* **11**, 3696 (2020).
36. Meers, M.P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
37. Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176**, 897-912 e20 (2019).
38. Gorkin, D.U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744-751 (2020).
39. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-12 (2009).
40. Henikoff, S., Henikoff, J.G., Kaya-Okur, H.S. & Ahmad, K. Efficient transcription-coupled chromatin accessibility mapping in situ. *biorxiv* **https://doi.org/10.1101/2020.04.15.043083**(2020).
41. Erb, M.A. *et al.* Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **543**, 270-274 (2017).
42. Daigle, S.R. *et al.* Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* **20**, 53-65 (2011).

43. Krivtsov, A.V. *et al.* A Menin-MLL Inhibitor Induces Specific Chromatin Changes and Eradicates Disease in Models of MLL-Rearranged Leukemia. *Cancer Cell* **36**, 660-673 e11 (2019).
44. Cusanovich, D.A. *et al.* Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-4 (2015).
45. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).

Supplementary Figures

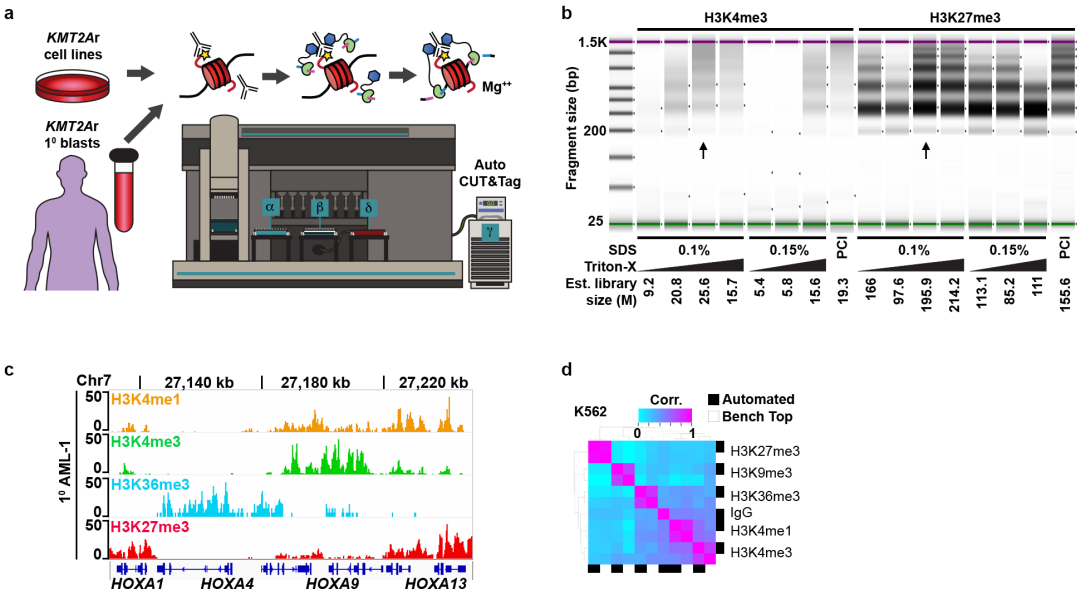


Supplementary Figure 1: KMT2A N-terminus and C-terminus specific antibodies for AutoCUT&RUN chromatin profiling. Pearson correlations between KMT2A N-terminus and C-terminus specific antibodies over the KMT2A merged peak for each sample. In the control H1 cells signals for the KMTA-N1 antibody (Millipore Cat #05-764), KMTA-N2 antibody (Cell Signaling Tech Cat #14689S), KMTA-C1 antibody (Millipore Cat #05-765) and KMTA-C2 antibody (Santa Cruz Cat #sc-374392), are all highly correlated, indicating that in these samples the N-terminal and C-terminal regions of wildtype KMT2A co-localize on chromatin.

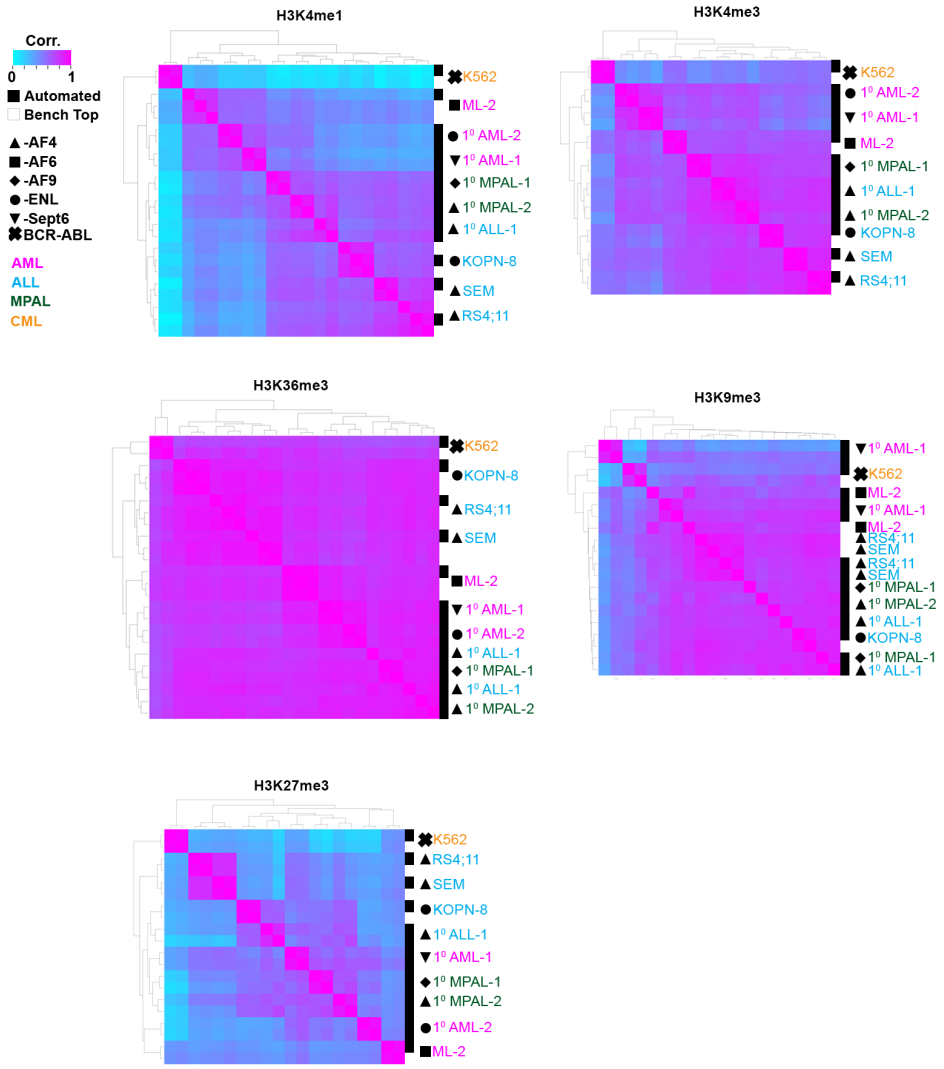


Supplementary Figure 2: Features of oncogene binding sites in all samples. a, Heatmaps show that in the control H1 cells all KMT2A peaks have comparable levels of KMT2A N-terminal signal and KMT2A C-terminal signal. **b-h,** In all of the *KMT2A*r samples the oncogene is broadly enriched over the sites called as oncogene targets as compared to the wildtype KMT2A target sites. Over this same array of sites tumor-specific differences in the enrichment of the SEC and

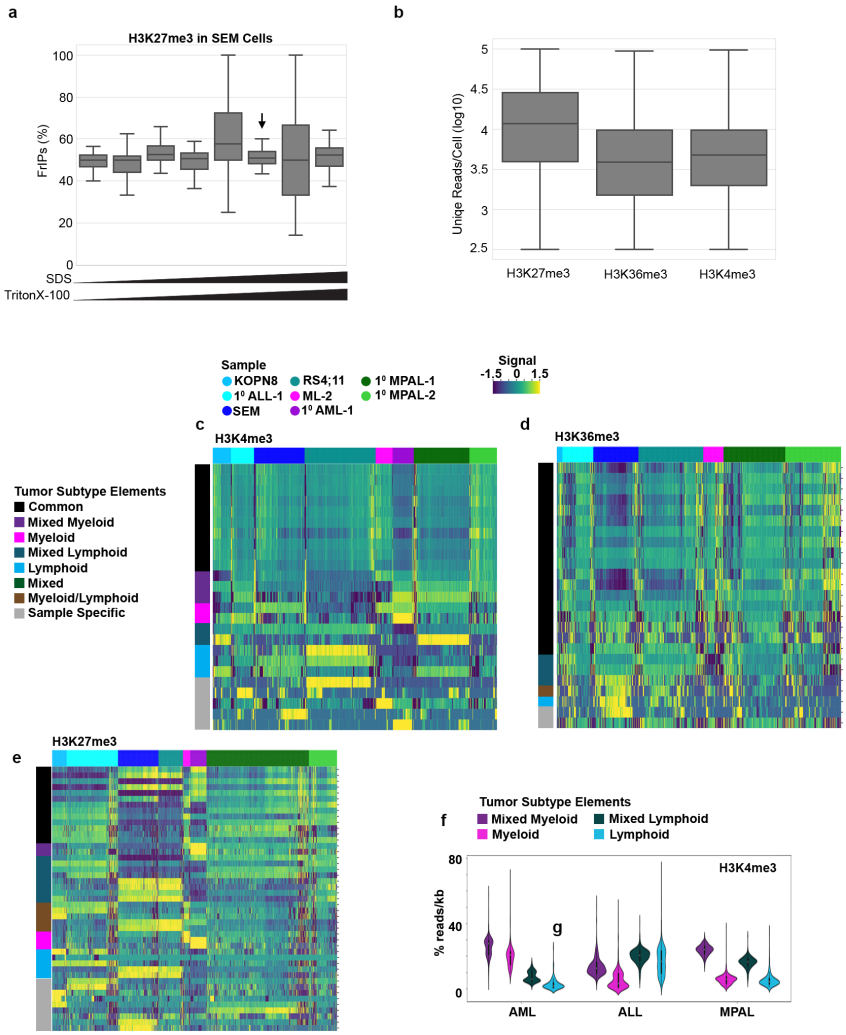
Dotcom components ENL, ELL and Dot1L are seen at the oncogene binding sites, although these factors are enriched at the wildtype binding sites to a similar degree between tumors. **i,** ENL, ELL, and Dot1L peaks are all more enriched over gene bodies that the TSS in control K562 cells that lack a KMT2A fusion oncogene.



Supplementary Figure 3: Adaptation of CUT&Tag for full automation. a, Con-A bound nuclei are incubated with the primary antibody of interest and arrayed for AutoCUT&Tag profiling on a Liquid handling robot equipped for high volume magnetic seperation (α), low volume magnetic seperation (β) and temperature control (δ, γ). This method pepares up to 96 sequencing-ready samples in a single day that can all be pooled on a single HiSeq two-lane or comparable flow cell for sequencing. The automated protocol uses a low concentration of SDS to displace bound Tn5 from tagmented DNA, and Triton-X100 to quench the detergent for PCR. **b,** A titration experiment showing optimization of the DNA release and quenching conditions for AutoCUT&Tag. Varying amounts of SDS and Triton-X100 were tested for library yield. Arrows indicate the optimum condition. **c,** Segmentation of a representative region of the human genome in an AML patient sample by the histone modifications used in this study. **d,** Pearson correlation matrix of reproducibility between benchtop and automated CUT&Tag profiling methods on K562 fixed nuclei with 5 antibodies to histone modifications.



Supplementary Figure 4: AutoCUT&Tag is highly reproducible for profiling primary patient leukemia. a, Pearson correlation matrix of reproducibility between benchtop and automated CUT&Tag profiling methods on fixed nuclei from all 9 *KMT2A*r leukemias as well as a K562 chronic myeloid leukemia control sample using antibodies against H3K4me1, H3K4me3, H3K36me3, H3K9me3 and H3K27me3. H3K4me1, H3K4me3, and H3K27me3 show the greatest variation between samples.



Supplementary Figure 5: Single cell cluster analysis of histone methylation in *KMT2A*r leukemias. a, The Fraction of Reads in Peaks (FRiPs) varies across SDS and Triton-X titration conditions. Arrow indicates the optimum conditions for Tn5 release and PCR enrichment. **b**, Boxplots showing the distributions of unique reads per cell across all of the cells profiled for H3K27me3, H3K36me3 and H3K4me3. **c,** Heatmap of the normalized count matrix of H3K4me3 reads from single cells in the 25 groups of H3K4me3 features identified in bulk (as shown in Fig. 3f). While clustering, single cells were confined to their subtype category. Colors above the matrix correspond to the sample, and colors on the left correspond to the lineage designations of the 25 groups of H3K4me3 features. **d,** Same as (c), for H3K36me3 clustered into 25 groups (as shown in Fig. 3c). **e,** Same as (c) for H3K27me3 clustered into 45 groups (as shown in Fig. 3d). **f**, Violin plot of single cells grouped according to their tumor subtype (x-axis) showing the

percent of H3K4me3 reads that fall into lineage-specific H3K4me3 promoter modules, as defined in Fig. 3b and shown in Fig. 3f.