KMT2A oncoproteins induce epigenetic resistance to targeted therapies.

Derek H. Janssens[1#], Melodie Duran[1], Dominik J. Otto[1,2], Weifang Wu[1], Yiling Xu[1,9], Danielle Kirkey[3,4], Charles G. Mullighan[5,6], Joanna S. Yi[7], Soheil Meshinchi[3,4], Jay F. Sarthy[8], Kami Ahmad[1+], Steven Henikoff[1,9+].

[1]Basic Sciences Division, [2]Translational Data Science IRC, [3]Translational Science and Therapeutics Division, Fred Hutchinson Cancer Research Center, Seattle, WA. [4]Division of Hematology and Oncology, Seattle Children's Hospital, Seattle, WA. [5]Department of Pathology, [6]Center of Excellence for Leukemia Studies, St. Jude Children's Research Hospital, Memphis, TN. [7]Pediatric Hematology and Oncology, Baylor College of Medicine and Texas Children's Hospital, Houston, TX. [8]Seattle Children's Research Institute, Seattle, WA. [9]Howard Hughes Medical Institute, Chevy Chase, MD, USA.

[#]Current Address. Department of Epigenetics, Van Andel Institute, Grand Rapids, MI.
[+]Corresponding author. Email: kahmad@fredhutch.org; steveh@fredhutch.org.

Abstract

Chromosomal translocations involving the *Lysine-Methyl-Transferase-2A* (*KMT2A*) locus generate potent oncogenic fusion proteins (oncoproteins) that disrupt regulation of developmental gene expression. By profiling the oncoprotein-target sites of 36 broadly representative *KMT2A*-rearranged leukemia samples, including three samples that underwent a lymphoid-to-myeloid lineage-switching event in response to therapy, we find the genomic enrichment of the oncoprotein is highly variable between samples and subject to dynamic regulation. At high levels of expression, the oncoproteins preferentially activate either an acute lymphoblastic leukemia (ALL) program, enriched for pro-B-cell genes, or an acute myeloid leukemia (AML) program, enriched for hematopoietic-stem-cell genes. The fusion-partner-specific-binding patterns over these gene sets are highly correlated with the prevalence of each mutation in ALL versus AML. In lineage-switching samples the oncoprotein levels are reduced and the oncoproteins preferentially activate granulocyte-monocyte progenitor (GMP) genes. In a sample that lineage switched during treatment with the menin inhibitor revumenib, the oncoprotein and menin are reduced to undetectable levels, but ENL, a transcriptional cofactor of the oncoprotein, persists on numerous oncoprotein-target loci, including genes in the GMP-like lineage-switching program. We propose KMT2A oncoproteins promote lineage-switching events through dynamic chromatin binding and can induce epigenetic lesions, marked by ENL, that support resistance to targeted therapies.

Introduction

KMT2A-rearranged (*KMT2A*r) leukemias are defined by chromosomal translocations that produce potent in-frame fusion oncoproteins that associate with chromatin and cause upregulation of target gene expression[1,2]. *KMT2A*r leukemias are highly aggressive and traditionally have been associated with a poor prognosis[3,4]. Advances in targeted immuno-therapies and pharmacological inhibitors have dramatically improved the treatment options for *KMT2A*r leukemias[5,6]. Despite sharing genetically related oncogenic lesions, *KMT2A*r leukemias show highly heterogeneous responses to therapeutic intervention[5,7,8].

The acquisition of secondary mutations and clonal selection are common genetic mechanisms that lead to therapeutic resistance, however, *KMT2A*r leukemias are characterized by an unusually low mutational burden[9], and may also evade targeted therapies through epigenetic mechanisms that are not well understood. In response to therapeutic pressures, *KMT2A*r B-cell acute lymphoblastic leukemias (B-ALLs) occasionally switch to an acute myeloid leukemia (AML) identity and rapidly relapse[10-13]. Several pharmacological agents have been developed that target the interaction of the KMT2A oncoproteins with menin, a chromatin scaffold protein, and cause dissociation of oncoprotein complexes from the chromatin[14,15]. In phase I clinical trials of the menin inhibitor revumenib, reduced oncoprotein target gene expression was correlated with disease remission[5]. Mutations in *MEN1* (encoding menin) that block the interaction of menin with revumenib act as a genetic mechanism of resistance[16]. In addition, genetic mutations, alternative splicing and altered expression of chromatin regulatory proteins such as CHD4 have been identified in *KMT2A*r AML samples after lineage switching[10], and mutations in components of a KMT2A related complex, KMT2C/D-UTX, can lead to resistance to menin inhibitors in cell line and animal models[17]. This indicates that concerted alterations in gene expression can also lead to therapeutic resistance and relapses of *KMT2A*r leukemias.

Single-cell RNA sequencing and chromatin profiling studies have identified considerable intra-tumoral heterogeneity in *KMT2A*r leukemia samples[8,18], but why *KMT2A*r leukemias, as opposed to other genetic subtypes of acute leukemia, display an unusual propensity for lineage switching remains unclear. Previously, we developed a high-throughput method for mapping the genome-wide binding sites of KMT2A oncoproteins in patient samples and found the oncoprotein-target genes themselves are dynamically regulated[19], suggesting the oncoproteins may contribute to the lineage plasticity and epigenetic resistance to targeted therapies.

Here, by directly profiling the KMT2A oncoproteins in primary patient samples, we demonstrate that the oncoprotein fusion partner and its expression levels influence genome-wide binding site selection. We find that both the oncoprotein expression and target binding are dynamic and promote lineage switching by transitioning from a B-ALL-instructive role to activating genes that are normally expressed in granulocyte-monocyte progenitor (GMP) cells. We characterize the first example of a lineage-switching event that occurred during treatment with a menin inhibitor, and find that in the AML patient sample the oncoprotein

cofactor ENL remains bound to a subset of oncoprotein-target genes including the GMP-like program, despite the on-target dissociation of the oncoprotein and menin from chromatin. In addition, we find this same GMP-like program is upregulated in a previously characterized menin-inhibitor-resistant AML, suggesting the persistent activation of this GMP-like program may be a recurring mechanism of epigenetic resistance that allows *KMT2A*r leukemias to evade targeted therapies.

Results

Heterogeneous KMT2A oncoprotein levels

We used AutoCUT&RUN[20] to profile KMT2A-oncoprotein binding sites in 36 *KMT2A*r leukemia samples, including 4 cells lines, 14 infant leukemias and 18 leukemias from pediatric or adult patients that are representative of the major lineage subtypes and the most common oncoproteins found in *KMT2A*r leukemias (**Fig. 1a**, Supplementary Information). To identify sites that are specifically bound by the oncoprotein, we performed AutoCUT&RUN assays using antibodies targeting the N-terminal domain of KMT2A, which is present in the oncoprotein as well as the remaining wild-type copy of KMT2A, and antibodies targeting the C-terminal domain of KMT2A, which is specific to the wild-type protein (**Fig. 1b**). Then, we called oncoprotein target sites according to the ratio of the KMT2A N-terminal signal over the C-terminal signal and determined the statistical significance of this ratio by considering the number of reads that fell in each interval (**Fig. 1c**). We set a threshold that produced 3 false-positive sites in the CD34+, oncoprotein-negative control sample, and identified 156 oncoprotein target sites in the SEM cell line (**Fig. 1c**). By applying this threshold to all 36 *KMT2A*r leukemia samples, we identified 1692 oncoprotein-binding sites and 1092 direct oncoprotein-target genes (Supplementary Information).

To characterize the degree of inter-tumoral heterogeneity in the oncoprotein-binding sites amongst *KMT2A*r leukemias, we assigned an "oncoprotein score" to all 1692 intervals that incorporates the KMT2A N over C terminal signal as well as the statistical significance of this value. We performed dimensionality reduction using principal component analysis and found the first four components captured greater than 50% of the variance in oncoprotein scores (Extended Data Fig. 1a). When organized in UMAP space according to the first four components, the *KMT2A*r leukemia samples grouped into four clusters (**Fig. 1d,** Extended Data Fig. 1b). This analysis indicates the lineage identity and translocation-partner genes both contribute to the genome-wide oncoprotein scores. For example, Cluster 2 is comprised exclusively of *KMT2A*r AML samples, and Clusters 3 and 4 are almost all ALLs (**Fig. 1d**). Cluster 4 is unique in that it contains only leukemias bearing the *KMT2A::AF4* translocation (**Fig. 1e**). The *AF4* locus is the most common *KMT2A* translocation partner, and this rearrangement is found almost exclusively in B-ALLs at diagnosis[21]. Our results show the KMT2A oncoproteins bind to distinct sites in AMLs versus ALLs, and in a subset of the ALL samples, the KMT2A::AF4 oncoproteins occupy the chromatin at regions that distinguish them from other *KMT2A*r leukemias.

The three B-ALL samples that went on to lineage-switch are interspersed amongst the other B-ALL samples we profiled. Cluster 4 includes a B-ALL sample and the paired AML relapse sample (**Fig. 1d, labeled** α). Cluster 3 contains two B-ALLs that went on to lineage switch. The first bears a *KMT2A::ENL* rearrangement and lineage switched during treatment with a Menin-inhibitor, and the paired AML sample is in Cluster 1 (**Fig. 1d, labeled** β). The last lineage-switching sample we profiled bears a *KMT2A::EPS15* rearrangement (**Fig. 1d, labeled** γ). For this patient, a post-lineage-switching sample was not available, but we were able to profile a second sample collected one week prior to lineage switching that also grouped in Cluster 3. We conclude the genome-wide localization of the oncoprotein does not obviously distinguish the B-ALL samples that went on to lineage switch from the other B-ALL samples we profiled.

In addition to the lineage identity and fusion partner genes, we detect significant heterogeneity in the oncoprotein levels within our collection of samples. For example, Cluster 1 includes the CD34+ oncoprotein-negative control sample as well as 8 AML samples, 2 infant ALLs, 1 mixed phenotype acute leukemia (MPAL), and an AML that resulted from a lineage-switching event that occurred during treatment with the menin inhibitor. The oncoprotein signal is barely detectable in these Cluster 1 samples (**Fig. 1f**). The average oncoprotein scores of Clusters 2 and 3 are significantly higher than those found in Cluster 1, and the oncoprotein scores are highest for samples in Cluster 4 (**Fig. 1f,g**). By performing oncogene specific qPCR on samples that share the same minimal *KMT2A::AF4*, *KMT2A::ENL* and *KMT2A::AFDN* exon junctions, we find the differences in oncoprotein scores we observed using CUT&RUN generally reflect differences in the expression levels of the oncogene (**Fig. 1h**, Extended Data Fig. 1c-e). Only two samples that both bear the *KMT2A::AFDN* rearrangement did not fit this trend (marked by a star and asterisk in **Fig. 1h**). The ML-2 cell line (star in **Fig. 1h**) lacks the wild-type copy of *KMT2A*, suggesting wild-type KMT2A may be required for efficient oncoprotein loading. The wild-type KMT2A C-terminal signal was detectable in the second sample (asterisk in **Fig. 1h**), but it is possible the oncoprotein loading efficiency is reduced in this sample through an alternative mechanism. We conclude the average oncoprotein score provides a semi-quantitative metric that is indicative of differences in the oncogene expression levels between samples. Among leukemias bearing the most common *KMT2A* translocations, the gene expression levels of the oncoprotein are highly heterogeneous.

KMT2A oncoproteins are thought to promote transcriptional activation of their target genes[1,2]. To test this model, we examined whether differences in oncoprotein levels at specific target loci are correlated with target gene expression. *MBNL1* is the most frequently called oncoprotein target gene across the collection of leukemias we profiled (24/36 samples, Supplementary Information). We detected differences in oncoprotein binding on the *MBNL1* locus that are representative of the average oncoprotein scores

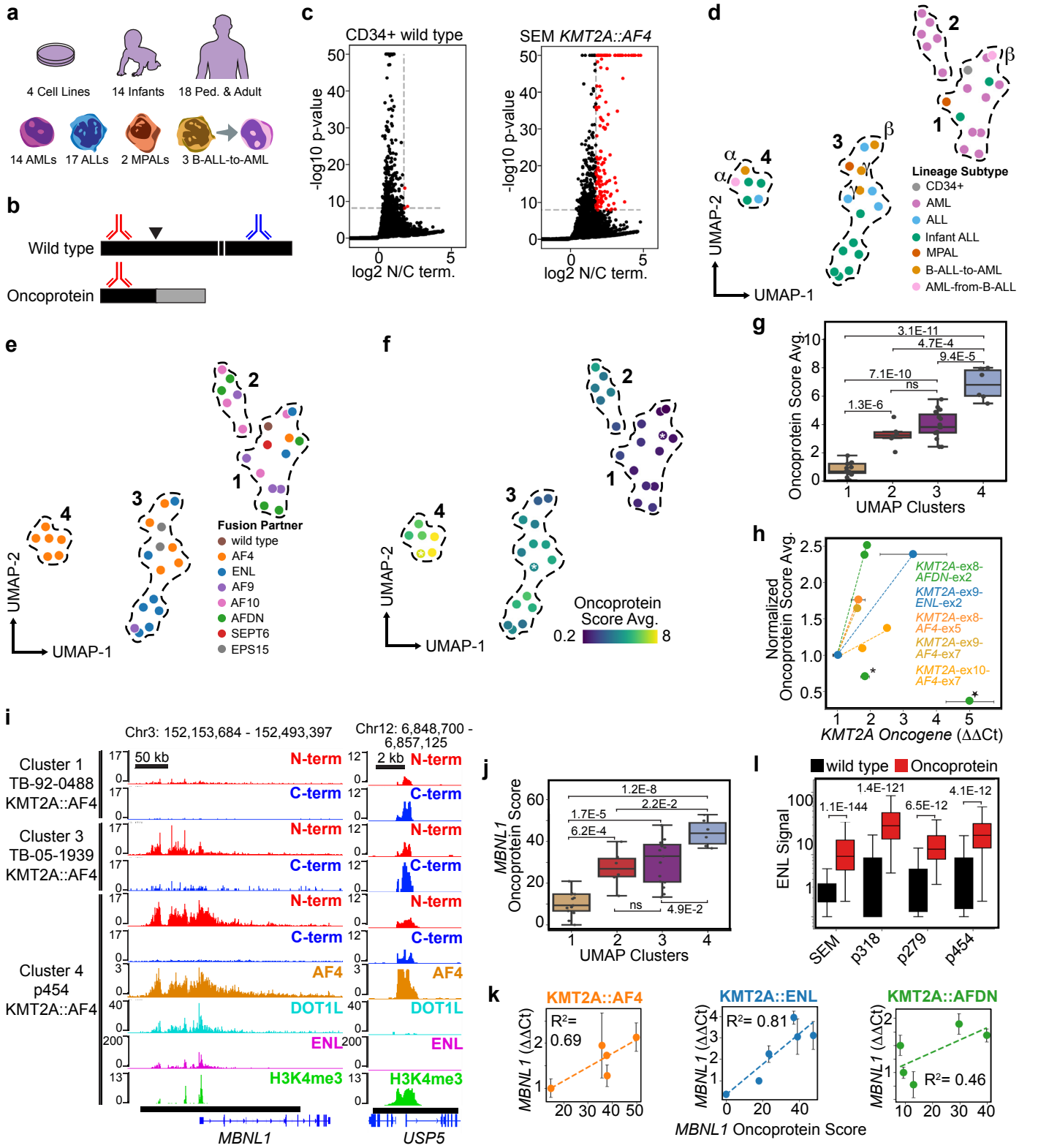


Fig 1: Genomic enrichment of KMT2A oncoproteins is highly heterogeneous. a, We performed AutoCUT&RUN on a collection of *KMT2A*r cell lines and patient samples including a range of patient ages and lineage subtypes. **b,** Comparison of CUT&RUN profiles using antibodies targeting the KMT2A N-terminal (red) and C-terminal (blue) regions can discriminate the oncoprotein-binding sites from wild-type KMT2A binding sites. **c,** Applying the same magnitude and significance thresholds to the KMT2A N versus C terminal profiles identifies many more oncoprotein-target sites (red dots) in the SEM *KMT2A*r cell line than the CD34+ control sample. **d,** Principal component analysis and UMAP embedding splits the *KMT2A*r samples into four clusters that are outlined with dotted lines; samples are colored according to the lineage subtype; patient-matched lineage-switching samples are indicated by the Greek letters α, B, γ. **e,** The UMAP embedding from (d) but colored according to the KMT2A-oncoprotein-fusion partner. **f,** Same as (d) colored according to the average oncoprotein scores (N/C terminal magnitude x significance); samples marked by an white asterisk correspond to panel (i)**. g,** The UMAP clusters capture difference oncoprotein levels. *p* values were computed using a two-tailed independent samples t-test;

Fig 1 continued: Cluster 1 n = 13, Cluster 2 n = 6, Cluster 3 n = 15, Cluster 4 n = 5. Here and throughout boxplot center lines = median, box limits = first and third quartiles, whiskers = 1.5 times the interquartile range (IQR); **h,** The average oncoprotein scores scale with the relative oncogene expression. Samples are grouped and colored according to the minimal *KMT2A*-fusion-partner exon junctions. Dotted lines indicate the regression. Two *KMT2A::AFDN* samples were identified as outliers: ML-2 is indicated by a star and A70498 by an asterisk. **i,** Genome Browser tracks show the difference in KMT2A N-terminal (red) and C-terminal (blue) signal in three *KMT2A::AF4* bearing leukemias, indicated by the white asterisks in (d), over the most common oncoprotein-target locus *MBNL1* (bottom black bar = oncoprotein-target peak). The *USP5* promoter is an example of a wild-type KMT2A target site (bottom black bar = wild-type KMT2A peak). AF4 (brown) DOT1L (cyan) ENL (magenta) and H3K4me3 (green) signal is also shown for the p454 sample. **j,** Oncoprotein scores over *MBNL1* (quantified over the bottom-black bar in i) are representative of the average oncoprotein scores in each UMAP cluster; *n* values are the same as (g). **k,** The *MBNL1* oncoprotein scores are correlated with the relative expression of *MBNL1* as measured by qPCR; dotted lines indicate the regression and R^2 measures the fit to the data. **l,** ENL is significantly enriched over the oncoprotein-target sites. *p* values were computed using a two-tailed independent samples t-test; *n* values are listed as wild-type site #, oncoprotein site #: SEM = 15611, 491; p318 = 22014, 246; p279 = 25672, 272; p454 = 18509, 438.

in samples from each of the four UMAP clusters (**Fig. 1i,j**). In *KMT2A::AF4* and *KMT2A::ENL* samples we profiled by qPCR, the *MBNL1* oncoprotein scores show a strong positive correlation with the relative expression levels of *MBNL1* (**Fig. 1k**). A positive but less pronounced correlation is observed in the *KMT2A::AFDN* samples (**Fig. 1k**). *MEIS1, JMJD1C* and *MEF2C* are identified as oncoprotein target genes in 21/36, 22/36, and 21/36 of the samples, respectively (Supplementary Information). We observe a strong positive correlation between *MEIS1* oncoprotein scores and *MESI1* gene expression, similar to the correlation seen with *MBNL1* (Extended Data Fig. 2a-c). However, *KMT2A::AF4* and *KMT2A::AFDN* oncoprotein scores do not show a significant correlation with the expression of *JMJD1C* or *MEF2C* (Extended Data Fig. 2d-i). Our results are consistent with the model that KMT2A oncoproteins act as transcriptional activators, with some genes, like *MBNL1* and *MEIS1* being more sensitive to the KMT2A oncoprotein dosage than others like *JMJD1C* and *MEF2C*, which may be more susceptible to regulation by other factors.

Previously, we found the oncoprotein cofactors DOT1L and ENL are strongly enriched at KMT2A-oncoprotein target sites[19]. To test our modified approach for calling KMT2A-oncoprotein-binding sites, we applied AutoCUT&Tag to profile DOT1L and ENL and AutoCUT&RUN to profile the AF4 fusion partner in the SEM *KMT2A*r cell line and three additional patient samples bearing *KMT2A::AF4* translocations. AF4 shows a very similar pattern of enrichment to the KMT2A N terminus over regions called as oncoprotein-target sites (**Fig. 1i**, see *MBNL1*), but is also enriched over many of the wild-type KMT2A binding sites (**Fig. 1i**, see *USP5*). AF4 does not effectively distinguish oncoprotein-target sites from wild-type sites in patient samples (Extended Data Fig. 2j). DOT1L is significantly more enriched at the oncoprotein-target sites than the wild-type sites in three out of four samples (**Fig. 1i**, Extended Data Fig. 2k). ENL is significantly more enriched at oncoprotein-target sites in all of the samples we profiled (**Fig. 1i,l**). Wild-type KMT2A catalyzes Histone-H3-Lysine-4 trimethylation (H3K4me3) at gene promoters, and AutoCUT&Tag profiles of H3K4me3 do not effectively distinguish wild-type and oncoprotein-target sites, serving as a negative control (**Fig. 1i**, Extended Data Fig. 2l). We conclude, our modified approach for comparing the enrichment of the KMT2A N and C terminus is effective for calling oncoprotein-target sites. Co-occupancy with oncoprotein cofactors DOT1L and ENL can validate these calls, whereas co-occupancy with AF4 cannot.

Lineage-specific and fusion-partner-dependent binding sites

By arranging samples according to the first two principal components, we find the KMT2A-fusion partners determine a large proportion of the variance in the oncoprotein-binding sites. Specifically, Principal Component 1 (PC1) primarily captures the variance in the binding sites of the KMT2A::AF4 and KMT2A::ENL fusion proteins and is tightly associated with the average oncoprotein scores in the ALL samples (**Fig. 2a,** Extended Data Fig. 3a). On the other hand, Principal Component 2 (PC2) captures the variance in the binding sites of the KMT2A::AF10 and KMT2A::AFDN fusion proteins in the AML samples (**Fig. 2a,** Extended Data Fig. 3b). PC1 and PC2 are both associated with the magnitude of the oncoprotein scores in the KMT2A::AF9 samples, but to an intermediate degree (**Fig. 2a,** Extended Data Fig. 3a,b). When grouped according to the oncoprotein-fusion partners, the slope of the best-fit lines for PC1 versus PC2 matches the prevalence of each translocation in *KMT2A*r ALLs versus AMLs (**Fig. 2a,b**)[21]. This strongly suggests the *KMT2A*-fusion partner influences the genome-wide binding sites of the oncoprotein as well as the lineage specificities of the leukemias produced by the associated translocations.

Next, we identified oncoprotein-binding sites that contribute to the variance in PC1 and PC2 by rank ordering the sites according the loading values (**Fig. 2c,d**). The oncoprotein target genes *GNAQ*, *TAPT1* and *FLT3* are among the top contributors to PC1 (**Fig. 2c,** Supplementary Information), whereas *SKAP2*, *SENP6*, *ZNF521* and *HOXA9* are among the top contributors to PC2 (**Fig. 2d,** Supplementary Information). The group of "highly sensitive" genes that a previous study found are rapidly downregulated in response to auxin-mediated KMT2A::AF9 degradation includes genes that contribute to both PC1 and PC2 (*MEF2C*, *MEIS1* and *SOCS2*) as well as PC1-specific genes (*CPEB2*) or PC2-specific genes (*SKIDA1*, *PBX3*, *HOXA9;* Supplementary Information)[22]. The oncoprotein-target genes in PC1 are enriched for Gene Ontology (GO) terms that include the regulation of hematopoiesis, myeloid cell differentiation, leukocyte differentiation, and immune responses (Extended Data Fig. 3c). The oncoprotein-target genes in PC2 primarily regulate developmental patterning (Extended Data Fig. 3d). To determine how the PC1-ALL-gene program and the PC2-AML-gene program are related to normal hematopoiesis, we plotted the average z-scores on a previously generated UMAP of single-cell RNA-sequencing data from lineage depleted bone marrow[23] (**Fig. 2e-g**). We

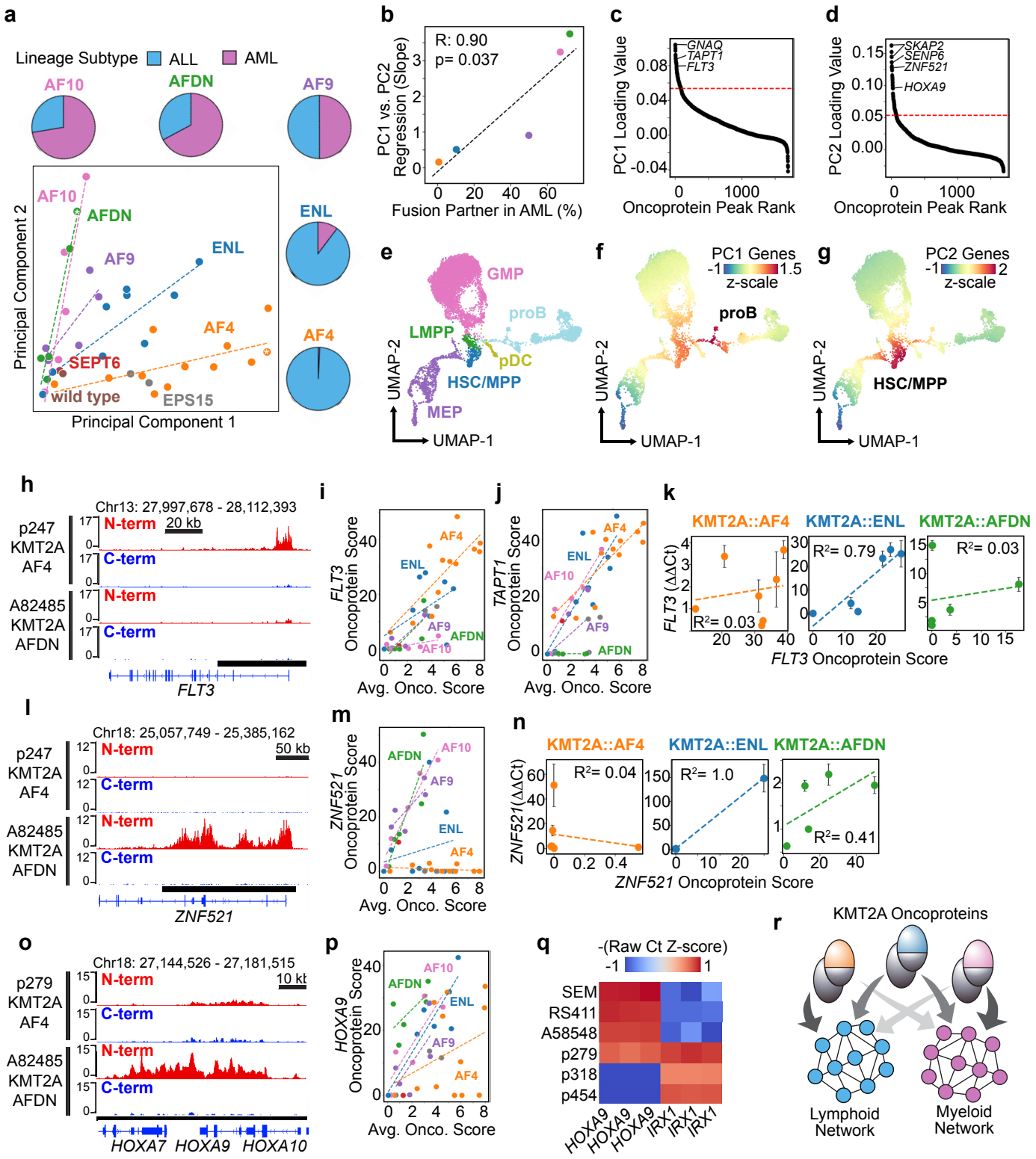


Fig 2: At elevated levels the KMT2A oncoproteins are instructive of lineage-specific genomic binding sites. a, Principal Components 1 and 2 organize *KMT2A*r leukemia samples according to the oncoprotein-fusion partners. Dotted regression lines are included for samples bearing related translocations. Pie charts indicate the frequency of each KMT2A-fusion partner found in ALL versus AML[18]; samples marked by a white asterisk correspond to panels (h) and (l). **b,** The slope of the KMT2A-oncoprotein regression lines from (a) are highly correlated with the frequency that each fusion partner is found in AML versus ALL. **c,** The top 5% of sites (indicated by the red line) that contribute most to Principal Component 1 were identified by rank ordering the 1oading values. **d,** same as (c) but for Principal Component 2. **e**, UMAP from Otto et al. 2024 (ref. 22) of single-cell RNA-seq data collected from lineage-depleted bone marrow; HSC/MPP = Hematopoietic Stem Cell/Multipotent Progenitor, MEP = Megakaryocyte Erythroid Progenitor, pDC = pre-Dendritic Cell, proB = pro-B Cell, LMPP = Lymphoid-Myeloid-Primed Progenitor, GMP = Granulocyte Monocyte Progenitor. **f,** UMAP embedding from (e) colored by the average z-score of genes from PC1. **g,** UMAP embedding from (e) colored by the average z-score of genes from PC2. **h,** Genome Browser tracks show the KMT2A::AF4 oncoprotein binds the *FLT3* locus while the KMT2A::AFDN oncoprotein does

Fig 2 continued: not (bottom black bar = oncoprotein-target peak). **i,** Scatterplot showing the *FLT3*-oncoprotein scores (quantified over the bottom black bar in h) relative to the average oncoprotein score in each sample. Samples are colored by the KMT2A-fusion partner and dotted-lines indicate the regression. **j,** Same as (i) but showing *TAPT1* oncoprotein scores. **k,** The KMT2A::ENL oncoprotein scores on *FLT3* are correlated with the relative expression of *FLT3* as measured by qPCR, while KMT2A::AF4 and KMT2A::AFDN levels are not correlated with expression of *FLT3*; dotted-lines indicate the regression and R^2 measures the fit to the data. **l,** Same as (h) showing the *ZNF521* locus. **m,** Same as (i) but showing *ZNF521* oncoprotein scores. **i,** The KMT2A::AFDN oncoprotein scores on *ZNF521* are correlated with the relative expression of *ZNF521* as measured by qPCR. Only one KMT2A::ENL sample expresses *ZNF521*. **o,** Same as (h) showing the *HOXA9* locus, and a different *KMT2A::AF4* rearranged sample in which *HOXA9* is not bound. **p,** Same as (i) but showing *HOXA9* oncoprotein scores, **q,** *HOXA9* and *IRX1* show mutually exclusive expression in *KMT2A::AF4* samples except for the p279 samples that went on to lineage switch into p318. The Ct values from three qPCR replicates are shown. **r,** The KMT2A-fusion partners direct the genomic occupancy of the oncoproteins to activate either lymphoid or myeloid lineage programs.

find the PC1-ALL-gene program is strongly enriched for genes normally expressed in pro-B cells, whereas the PC2-AML-gene program is strongly enriched for hematopoietic stem cell and multipotent progenitor specific genes (**Fig. 2e-g**).

Our results suggest the *KMT2A*-fusion partner influences the lineage identity of the leukemia, which prompted us to examine this relationship more closely on specific oncoprotein target genes. The most frequent oncoprotein target genes *MBNL1*, *MEIS1*, *JMJD1C* and *MEF2C* all contribute to both PC1 and PC2, and the oncoprotein scores on these genes scale with the average oncoprotein score in each sample (Supplementary Information, Extended Data Fig. 3e-h). When the oncoproteins are highly abundant, PC1-specific oncoprotein-target genes, such as *FLT3* and *TAPT1*, are consistently bound by the B-ALL-enriched KMT2A::AF4 and KMT2A::ENL fusion proteins (**Fig. 2h-j,** Extended Data Fig. 3i). In comparison, only one KMT2A::AFDN sample and none of the KMT2A::AF10 samples had appreciable levels of the oncoprotein on the *FLT3* locus (**Fig. 2i**). Interestingly, the KMT2A::AF10 oncoprotein scores over the *TAPT1* locus scale with the overall oncoprotein levels in these samples, while the levels of KMT2A::AFDN on the *TAPT1* locus do not (**Fig. 2j**). This suggests *FLT3* is a lineage-specific oncoprotein-target gene whereas *TAPT1* is a fusion-partner-specific oncoprotein-target gene. By applying qPCR to a subset of these samples, we find the relative expression of *FLT3* is strongly correlated with the KMT2A::ENL oncoprotein scores over the *FLT3* locus but not the KMT2A::AF4 or KMT2A::AFDN oncoprotein scores (**Fig. 2k)**. This indicates the *FLT3* locus is sensitive to the dosage of KMT2A::ENL but may become saturated and reach its maximum expression level when occupied by even low levels of KMT2A::AF4.

We also Identified lineage-specific and fusion-partner-specific oncoprotein targets that contributed to PC2. The *ZNF521* gene is critical for *KMT2A*r AMLs[24,25], and the KMT2A::AF10, KMT2A::AFDN and KMT2A::AF9 oncoprotein scores on the *ZNF521* locus are positively correlated with the overall oncoprotein scores in the AML samples (**Fig. 2l,m**). In comparison, the *ZNF521* locus was not bound by the oncoproteins in any of the KMT2A::AF4 B-ALL samples, and was only bound in two of the KMT2A::ENL samples (**Fig. 2m**). As expected, the relative expression of *ZNF521* is positively correlated with KMT2A::AFDN, but not KMT2A::AF4 and only one of the KMT2A::ENL samples profiled by qPCR had detectable *ZNF521* expression (**Fig. 2n**). We conclude that in *KMT2A*r leukemias *ZNF521* is an AML-specific oncoprotein target gene.

The *HOXA9* locus also contributes to PC2, and the *HOXA9* oncoprotein scores scale with the overall oncoprotein levels in all the samples except those bearing *KMT2A::AF4* translocations (**Fig. 2d,o,p,** Supplementary Information). *HOXA9* was not called as an oncoprotein target in three out of six of the KMT2A::AF4 samples from UMAP Cluster 4, which had the highest overall oncoprotein levels (Extended Data Fig. 3j). Previous studies identified a subset of infant *KMT2A::AF4*-rearranged samples that have an exceptionally high risk of relapse and are characterized by low expression of genes in the *HOXA* cluster, and high expression of the homeobox transcription factor *IRX1*[8,26-28]. This gene expression pattern has also been reported in *KMT2A::AF4*-rearranged B-ALLs that went on to lineage switch[10]. We find that *IRX1* is a direct target of KMT2A::AF4 in the three samples from Cluster 4 in which *HOXA9* is not called as an oncoprotein-target gene (Extended Data Fig. 3k-m, Supplementary Information). When measured by qPCR, we find the expression of *HOXA9* and *IRX1* are generally mutually exclusive (**Fig. 2q**). The p279 sample is a B-ALL that went on to lineage switch in response to ALL-directed chemotherapy and is the only sample that expresses both *HOXA9* and *IRX1*. In p318, the AML that resulted from the lineage switch of p279, we find *HOXA9* is downregulated whereas the expression of *IRX1* is maintained (**Fig. 2q**). Our data suggests there is a dichotomy in the direct regulation of *HOXA9* and *IRX1* by the KMT2A::AF4 oncoproteins which may contribute to lineage-switching events. Together, we conclude that at high levels of expression the oncoproteins activate genes that influence the lymphoid versus myeloid identity of the leukemia (**Fig. 2r**).

Oncoprotein dynamics and lineage switching

To characterize the epigenetic mechanisms underlying the lineage plasticity of *KMT2A*r leukemias, we performed a comprehensive comparison of the oncoprotein-target genes in the patient-matched lineage-switching samples p279 and p318. We found that oncoprotein scores are generally reduced following relapse as AML (**Fig. 3a**), with a corresponding decrease in the mRNA expression of the oncogene (**Fig. 3b**). This finding is consistent with a previous report showing decreased oncogene expression during lineage switching in three out of four of the KMT2A::AF4-rearranged samples profiled by RNA-seq[10]. Next, we identified oncoprotein-target sites with the most dramatic differences between the p279 B-ALL sample and the p318 AML sample (Extended Data Fig. 4a). *BANK1*, *IGF2BP2*, *WDR66* and *DNTT* were among the B-ALL-specific oncoprotein-target genes (**Fig.**

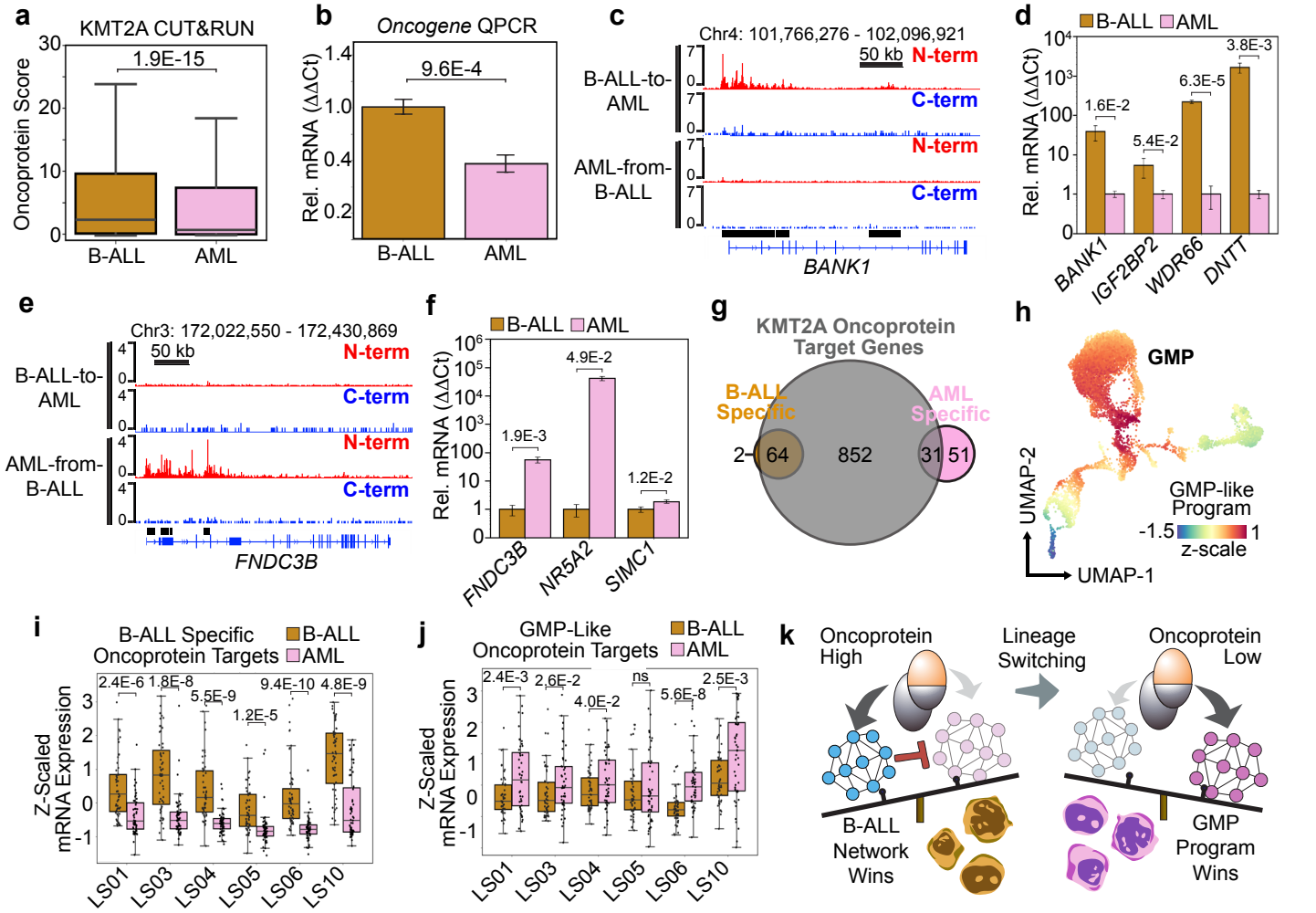


Fig 3: The oncoprotein binding sites change during B-ALL-to-AML lineage-switching events. a, Boxplot comparing the oncoprotein scores in a patient matched B-ALL and AML sample that lineage switched in response to B-ALL directed therapy. **b,** Oncogene expression is significantly reduced in the AML after lineage switching. Error bars = standard deviation, and the *p* value was computed using a two-tailed independent samples t-test; *n* = 3 qPCR replicates. **c,** Genome Browser tracks showing the oncoprotein levels are elevated over the *BANK1* locus in the B-ALL sample prior to lineage switching. Bottom black bars indicate B-ALL specific oncoprotein target regions. **d,** B-ALL specific oncoprotein target genes are expressed at significantly higher levels in the B-ALL sample (gold) prior to lineage switching than the AML sample (pink). Error bars = standard deviation; *p* values computed using the using a two-tailed independent samples t-test; *n* = 3 qPCR replicates. **e,** Genome Browser tracks showing that the oncoprotein levels are elevated over the *FNDC3B* locus in the AML sample after lineage switching. Bottom black bars indicate AML specific oncoprotein target regions. **f,** Same as (d) but for AML-specific oncoprotein target genes. **g,** A Venn Diagram comparing the lineage-switching B-ALL-specific and AML-specific oncoprotein-target genes to all other oncoprotein-target genes. The AML group includes 51 non-overlapping genes. **h,** The UMAP embedding of healthy lineage-depleted bone marrow colored by the average z-scores of the 51 non-overlapping AML-specific oncoprotein target genes from (g). **i,** Boxplot comparing the relative expression of the B-ALL-specific oncoprotein target genes in a cohort of 6 B-ALL (gold) and AML (pink) patient-matched samples that underwent a lineage-switching event as described in Tirtakusuma et al. 2022 (ref. 9). *p* values were calculated using a Wilcoxon sign-rank test with a one-sided hypothesis, n = 66 genes. **j,** Same as (i) but comparing the relative expression of the 51 non-overlapping AML-specific oncoprotein target genes we refer to as the GMP-like program, n = 51. **k,** At high levels the KMT2A oncoprotein reinforces the B-ALL identify by activating a lymphoid program. During lineage switching the oncoprotein levels are reduced, and the oncoprotein activates an GMP-like program.

3c, Extended Data Fig. 4a, Supplementary Information). As expected, these genes were expressed at significantly higher levels in the B-ALL sample prior to lineage-switching (**Fig. 3d**). Conversely, *FNDC3B*, *NR5A*, *and SIMC1* were among the genes identified as AML-specific targets (**Fig. 3e**, Extended Data Fig. 4a, Supplementary Information), and were dramatically upregulated in the AML sample (**Fig. 3f**). We conclude the expression levels and genome-wide localization of the KMT2A oncoproteins are dynamic and these changes are likely interdependent with the lineage identity of the leukemia.

We then compared the collection of B-ALL-specific and AML-specific oncoprotein-target genes to samples that were not involved in a lineage-switching event. The majority of the B-ALL-specific oncoprotein-target sites (64/66) were also identified in other samples (**Fig. 3g,** Supplementary Information). *BANK1*, *IGF2BP2*, *WDR66* and *DNTT* all contributed to the PC1-ALL-gene program

(Supplementary Information), and expression of the B-ALL-specific oncoprotein-target genes is normally enriched in pro-B cells (Extended Data Fig. 4b). Interestingly, 51/82 of the AML-specific oncoprotein-target genes did not overlap with the oncoprotein-target genes in the other samples we profiled (**Fig. 3g,** Supplementary Information). This group of 51 oncoprotein-target genes, which we refer to as the "GMP-like program," is enriched for genes that are expressed in GMPs during normal hematopoiesis (**Fig. 3h**). By analyzing RNA-seq data previously collected from six *KMT2A::AF4* rearranged samples that underwent a B-ALL-to-AML lineage-switching event[10], we found the B-ALL-specific oncoprotein-target genes are expressed at significantly higher levels in all six patient-matched B-ALL samples (**Fig. 3i**). Conversely, the GMP-like program is upregulated in all six of the AML samples, although this difference was not significant in one of the samples (**Fig. 3j**). We conclude that the shift from direct activation of lymphoid-associated genes to activation of the GMP-like program is likely a recurring mechanism of oncoprotein-regulated lineage switching (**Fig. 3k**).

Menin-inhibitor-resistant ENL-marked epigenetic lesions

A subset of *KMT2A*r leukemias initially respond to the menin inhibitor revumenib by downregulating oncoprotein-target genes, but for unknown reasons are resistant to treatment, and rapidly relapse[16]. One of the *KMT2A::ENL*-rearranged samples we profiled underwent a lineage-switching event during treatment with the menin inhibitor. We compared the B-ALL and AML patient-matched samples (designated 148752 and 152985, respectively), and while 46 oncoprotein-binding sites were detectable in the B-ALL sample at diagnosis, treatment with the menin inhibitor reduced the genome-wide KMT2A N and C terminal signal to undetectable levels (**Fig. 4a,b,** Extended Data Fig. 5a, Supplementary Information). Similar to the p279/p318 samples, we observed a reduction in oncogene expression in the 152985 AML sample after lineage switching (**Fig. 4c**). However, while the oncoprotein became undetectable in the menin-inhibitor-treated AML, the oncogene was still expressed, but at lower levels, as confirmed by reverse transcription PCR using two different primer sets (**Fig. 4d**). This suggests the menin inhibitor effectively eliminated the B-ALL cells by targeting the oncoprotein, yet the leukemia persisted in the AML state.

To explore how the AML cells expanded despite the targeted loss of the oncoprotein, we profiled the genome-wide localization of the oncoprotein cofactors menin, DOT1L and ENL by AutoCUT&Tag. As expected, menin, DOT1L and ENL were all bound to oncoprotein-target sites in the B-ALL sample (**Fig. 4a,b,** Extended Data Fig. 5a), and both menin and DOT1L were reduced to undetectable levels in the menin-inhibitor-treated AML (**Fig. 4a,b,e**, Extended Data Fig. 5a,b). Strikingly, ENL remains bound at appreciable levels over numerous canonical oncoprotein-target genes like *MBNL1* and *RUNX2* as well as genes in the GMP-like program such as *KLF6* (**Fig. 4a,b,f,** Extended Data Fig. 5a).

Next, we investigated the shifts in ENL binding during lineage switching by quantifying the relative levels of ENL over the B-ALL-specific genes and GMP-like program. We used the set of all oncoprotein-target intervals identified in the non-lineage-switching samples as an internal control. The B-ALL-specific and GMP-like gene signatures were originally identified by comparing the oncoprotein-target sites in the p279 and p318 samples, and, as expected, ENL is significantly more enriched over the B-ALL-specific genes in the p279 B-ALL sample and is significantly more enriched over the GMP-like program in the p318 AML sample (**Fig. 4g**). Interestingly, prior to treatment with the menin inhibitor, ENL was significantly enriched over both the B-ALL-specific genes and GMP-like program in the 148752 B-ALL sample, and during treatment, in the 152985 AML sample, ENL was only significantly enriched over the GMP-like program (**Fig. 4g**). The expression of B-ALL-specific genes *BANK1* and *WDR66* was reduced by ~50 and ~6 fold, respectively, in the 152985 AML sample (Extended Data Fig. 5c). The expression of *IGF2BP2* increased slightly in the 152985 AML sample (Extended Data Fig. 5c), and *DNTT* was unchanged, but was expressed at vanishingly low levels to begin with. In comparison, the expression of genes in the GMP-like program was either unchanged, such as *FNDC3B*, or reduced but to a lesser extent (Extended Data Fig. 5d). Specifically, *SIMC1* was reduced ~2 fold, *KLF6* was reduced ~6 fold and *NR5A2* was not expressed (Extended Data Fig. 5d,e). Together, we conclude the KMT2A oncoprotein induces accumulation of menin, DOT1L and ENL over the oncoprotein target sites. Once this epigenetic lesion forms, ENL can remain bound over a subset of oncoprotein-target sites, including genes involved in the GMP-like program, even after the oncoprotein is disrupted by the menin inhibitor (**Fig. 4h**).

Lastly, we examined whether the lineage-switching mechanism of menin-inhibitor resistance we discovered is related to other resistance mechanisms that have been described in the literature. Mutations in the *KMT2C* gene have been shown to lead to menin-inhibitor resistance in genetic models[17]. We found that in the both of the AML samples that resulted from lineage-switching events the expression of *KMT2C* is downregulated, however this difference was only significant in the 152985 sample that lineage switched during treatment with the menin inhibitor (**Fig. 4i**). Next, we analyzed RNA-seq data from patient samples that were originally diagnosed as AML and developed resistance to that menin inhibitor in the absence of *MEN1* mutations[16]. As compared to the MV4;11 AML cell line and an AML patient-derived xenograft (PDX), which both demonstrated a prolonged response to menin inhibitors, the GMP-like program was significantly upregulated in the P9 AML patient sample that developed resistance (**Fig. 4j**). This indicates reductions in *KMT2C* expression and a shift toward the GMP-like program may be recurring mechanisms of epigenetic resistance to the menin inhibitors.

Discussion

The lineage identities and therapeutic responses of *KMT2A*r leukemias are highly heterogeneous[2,5,7,8]. To characterize the molecular basis of this heterogeneity we applied Automated CUT&RUN to profile the oncoprotein-binding sites across 36 *KMT2A*r leukemia samples that span the age of onset

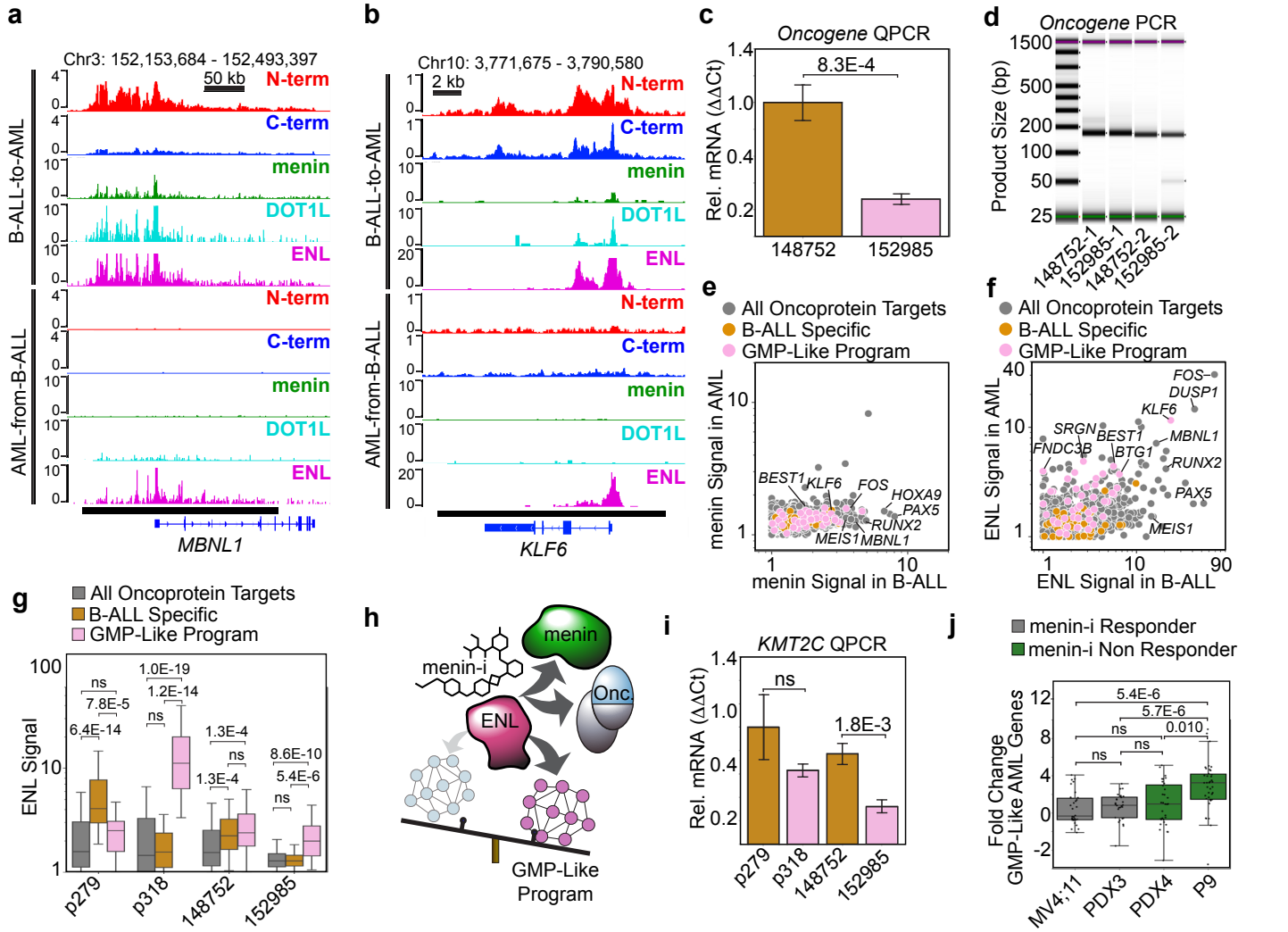


Fig 4: Lineage switching, epigenetic lesions and resistance to menin inhibitors. a, Genome Browser tracks showing the oncoprotein and the cofactors menin, DOT1L and ENL bound to the *MBNL1* locus in the B-ALL sample (top). In the patient-matched AML sample treated with the menin inhibitor only ENL remains bound (bottom). **b,** Same as (a) but showing the *KLF6* gene from the GMP-like program. **c,** Oncogene expression is significantly reduced in the AML after lineage switching. Error bars = standard deviation, and the *p* value was computed using the using a two-tailed independent samples t-test; *n* = 3 qPCR replicates. **d,** Real time PCR with two primer sets confirms the *KMT2A::ENL* oncogene is expressed in the AML treated with the menin inhibitor (expected size for 1 = 159 bp, and 2 = 150 bp). **e,** Scatterplot comparing the normalized menin levels in the B-ALL and the AML sample treated with the menin inhibitor over an internal control group of oncoprotein-target sites (grey) the B-ALL-specific program (gold) and the GMP-like program (pink). **f,** Same as (e) but showing the ENL signal persists over many of the oncoprotein-target genes in the AML sample. **g,** Box-plot comparing ENL signal over an internal control group of oncoprotein-target genes (grey, n = 1595) the B-ALL-specific program (gold, n = 52) and the GMP-like program (pink n = 45). *p* values were computed using the Mann-Whitney U Test with a Bonferroni correction for multiple hypothesis testing. **h,** The menin inhibitor disrupts KMT2A oncoproteins, but ENL accumulation persists on oncoprotein targets in the GMP-like program. **i,** Expression of the *KMT2C* gene is significantly reduced in the lineage-switching sample treated with the menin inhibitor. Error bars = standard deviation; *p* value computed using a two-tailed independent samples t-test; *n* = 3 qPCR replicates. **j,** The menin-inhibitor-resistant AML patient P9 sample has significantly higher expression of the GMP-like program than two menin-inhibitor responder samples and one additional non responder profiled in Perner et al. 2023. *p* values were computed using the two-sample t-test for independent samples. For comparisons of samples classified as non-responders to responders we used a one-sided test, otherwise we used a two-sided test. *n* is the number of GMP-like genes that had a significant fold change between the menin-inhibitor treated and untreated samples (padj. <0.05): MV4;11 = 30, PDX3 = 28, PDX4 = 32, P9 = 39.

and are representative of the diverse mutations and lineage subtypes that are observed in *KMT2A*r leukemias. We find that the oncoprotein levels and genome-wide binding sites are highly heterogeneous among patient samples and at high levels of expression the oncoprotein-fusion partner influences the oncoprotein-binding sites and is likely instructive of the lineage identity of the leukemia. By profiling patient samples that underwent B-ALL-to-AML lineage-switching events, we demonstrate that oncoprotein binding is dynamic. As compared to the patient-matched B-ALL samples, the oncogene expression and oncoprotein levels are generally reduced in the AML and the oncoprotein binding is shifted to a GMP-like program. To our knowledge, this is the first record of a B-ALL-to-AML lineage-switching event

that occurred during treatment with a menin inhibitor. We find a similar GMP-like program is activated in a patient sample that presented as AML at diagnosis and initially responded to menin inhibitors but rapidly relapsed in the absence of *MEN1* mutations[16]. We propose that KMT2A-oncoprotein dynamics and the induction of "epigenetic lesions," marked by ENL, play a critical role in the mechanisms that allow *KMT2A*r leukemias to evade targeted therapies.

Several mechanisms have been proposed to explain why the translocation partner genes *AF4* and *ENL* are much more common in B-ALLs whereas *AF10* and *AFDN* are more prevalent in AMLs[1,29,30]. Our results provide further evidence that these lineage biases are likely due to oncoprotein-fusion-partner dependent interactions with the chromatin. We find the fusion-partner-dependent chromatin-binding events are more pronounced when the oncoprotein levels are elevated. For example, the KMT2A::AF10 oncoprotein consistently binds the *TAPT1* locus in AMLs while KMT2A::AFDN does not. In a subset of the infant B-ALLs we profiled, KMT2A::AF4 binds the *IRX1* locus at the expense of *HOXA9*, but KMT2A::ENL did not bind *IRX1* in any of the samples we profiled. This is consistent with a previous report that *IRX1* is expressed much more frequently in infant B-ALLs harboring the *KMT2A::AF4* rearrangement than the *KMT2A::ENL* rearrangement[8]. Several observations suggest the KMT2A-fusion partner acts interdependently with the lymphoid-versus-myeloid lineage context to direct the genome-wide localization of the oncoprotein. Our principal component analysis identified genes that are preferentially bound by KMT2A::AF4 and KMT2A::ENL in the lymphoid lineage (e.g. *FLT3*) and KMT2A::AF10 and KMTA::AFDN in the myeloid context (e.g. *ZNF521*). Four of the samples we profiled did not conform to the fusion-partner-specific lineage biases. In the *KMT2A::AF10*-rearranged infant B-ALL and the *KMT2A::ENL*-rearranged AML we observed very low levels of the oncoprotein (Cluster 1 of the UMAP **Fig. 1d-f**). Similarly, in the *KMT2A::AF4* and *KMT2A::ENL* lineage-switching samples, we observed the oncogene and oncoprotein were depleted in the AML sample after relapse. This suggests the AF4, ENL, AF10 and AFDN fusion oncoproteins require the appropriate lineage context to "thrive," and efficiently initiate and reinforce the corresponding lineage subtype of leukemia.

The *KMT2A::AF9* mutation is found in roughly equivalent numbers of B-ALLs and AMLs[21]. We see KMT2A::AF9 accumulates over the *FLT3* and *TAPT1* B-ALL-associated target loci in AMLs (**Fig. 2i,j**), but the levels are below the threshold to be called as oncoprotein-target genes (Supplementary Information). In an AML induction model, KMT2A::AF9 binding is required to activate genes that contribute to both the PC1-B-ALL program as well as the PC2-AML programs[22]. Normally, AF9 is found in both the DOTCOM and SEC transcriptional complexes[31,32], and perhaps this broad interaction potential also underlies the lineage promiscuity of the *KMT2A::AF9* translocations.

The p279/p318 sample we profiled before and after lineage-switching provides a unique paradigm to consider the lineage-specific roles of the KMT2A oncoproteins. In addition to changes in the expression levels, the oncoprotein also binds to different sites in the B-ALL and AML samples. At diagnosis the oncoprotein is directly bound to genes in the B-ALL lineage program, and it is possible the lower levels of expression in the AML facilitate the "non-canonical" binding pattern over the GMP-like program. This provides a compelling case for a lineage-interdependency model in which the oncogene expression levels and oncoprotein-binding site selection both contribute to the lineage biases of specific *KMT2A* translocations.

Lineage switching could result from the trans-differentiation of B-ALL cells to an AML identity, or from the selective outgrowth of a latent multipotent cell type that retains both B-ALL and AML potential. A strong shift in the oncoprotein levels might support a trans-differentiation event. However, multiple lines of evidence suggest a selective outgrowth model is more likely. First, our comparison of oncoprotein-binding sites did not readily distinguish the B-ALL lineage-switching samples from the other B-ALL samples we profiled. Second, whole-exome sequencing of samples before and after lineage switching revealed a paucity of shared mutations[10]. And third, studies using single-cell RNA-seq identified a rare population of myeloid-biased cells in lineage-switching samples[18]. The latent multipotent progenitor model predicts that while the KMT2A::AF4 oncoprotein was thriving in the p279 B-ALL cells, the ALL-directed chemotherapy was effectively able to eliminate them. Then, the p318 AML sample arose from the same latent population that gave rise to the p279 B-ALL cells.

Single-cell-genomic profiling of hematopoietic stem and progenitor cells indicates the lymphoid lineage arises from a restricted progenitor subtype, called a Lymphoid-Myeloid-Primed Progenitor (LMPP), that also generates GMPs[23,33,34]. Perhaps this LMPP population acts as the cell of origin for lineage-switching *KMT2A*r leukemias, and provides an opportunity for the oncoprotein to induce or reinforce the GMP-like program as a secondary relapse mechanism. Future studies are necessary to elucidate any hematopoietic stem and progenitor cell-type-specific activities of the oncoproteins and determine the cellular context(s) where the oncoprotein activates the GMP-like program.

Intensive genetic and biochemical characterization has demonstrated menin, DOT1L and ENL act as transcriptional cofactors of the oncoprotein[31,32,35-40]. Numerous compounds targeting each of these interactions have shown promise in preclinical models[14,15,22,41-45], and phase I clinical trials indicate menin inhibitors are effective for treating a subset of relapse refractory patients with *KMT2A*r AML[5]. Several examples of genetic mechanisms that lead to lineage switching or acquired resistance to menin inhibitors have been described [10,16,17]. Our results suggest KMT2A oncoproteins also promotes resistance to targeted therapies through epigenetic mechanisms. During lineage switching the oncoprotein activates a secondary GMP-like program, and ENL can persist on these genes even after the oncoprotein is effectively eliminated through treatment with a menin inhibitor. Although genetic alterations might have contributed to resistance in these samples, our results point to ENL as a linchpin of epigenetic resistance to menin inhibitors and warrant investigation of combination therapies.

Acknowledgements

We thank Christine Codomo, Terri Bryson and Jorja Henikoff for technical assistance, Matthew Fitzgibbon for bioinformatics support and Phil Corrin and Merari Santa-Carbajal and the rest of the Fred Hutch Genomics Shared Resources for performing AutoCUT&RUN as well as sequencing and data processing.

Author Contributions

D.H.J. and J.F.S. conceived the study. D.H.J., W.W. and Y.X. performed the experiments; D.H.J., M.D. and D.J.O. performed data analysis. D.K., C.G.M. and J.S.Y., S.M. provided critical materials; D.H.J. and M.D. wrote the manuscript; S.H., J.F.S. and K.A. reviewed and edited the manuscript, and all authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Howard Hughes Medical Institute (S.H.), a Postdoctoral Fellowship from the Hartwell Foundation (D.H.J.) and a grant from the National Institutes of Health: NHGRI R01 HG010492 (S.H.). C.G.M. is supported by NCI R35 CA297695 and P30 CA021765 and the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital.

Methods

Patients

Patient samples were obtained from member COG institutions, Texas Children's Hospital and the St. Jude Children's Research Hospital in accordance with the Declaration of Helsinki. Written consent was obtained from all patients to permit the use of their de-identified samples in medical research. In the event the patient was a minor, written consent was obtained from the parent/guardian. The studies were overseen by the institutional review boards at the Fred Hutch Cancer Center (IR protocol 9950), St. Jude Children's Research Hospital and Texas Children's Hospital. None of the patients received compensation for their inclusion in this study.

Sample Information

The primary CD34+ HSPC control sample, as well as the four *KMT2A*r cell lines, 1 primary-patient ALL, 5 primary-patient AMLs, and 2 primary-patient MPALs were profiled in a previous study[19] and the paired-end fastq files are available through the Gene Expression Omnibus under accession code GSE159608.

In this study we profiled an additional 8 primary *KMT2A*r AML samples from a biobank maintained by the Meshinchi laboratory at the Fred Hutchinson Cancer Research Center, 7 infant KMT2Ar ALL leukemias from the St. Jude's Children's Research Hospital, 2 pediatric ALLs and 6 infant ALLs from the Texas Children's Hospital, including the samples that underwent a lineage-switching event in response to B-ALL directed treatment and during treatment with the menin inhibitor, and finally two infant B-ALL samples collected from the same donor prior to a lineage-switching event. All sample transfers were performed in accordance with institutional regulatory practices. These samples were collected from the whole blood of patients with greater than 80% blast counts by performing a Ficoll separation to remove red blood cells and neutrophils and the leukemia enriched samples were then cryopreserved. The specific *KMT2A* chromosomal translocation in each sample was determined using either whole-genome sequencing, targeted capture sequencing, or using standard cytogenetic approaches as previously described[9,46].

Antibodies

For AutoCUT&RUN profiling of KMT2A oncoproteins, we used the strategy we previously described in Janssens et al. 2021. Each sample was profiled using two antibodies targeting the KMT2A N terminus, KMT2A N1 (mouse monoclonal anti KMT2A used at a 1:100 dilution; Millipore, clone N4.4, cat. no. 05-764; https://scicrunch.org/resolver/RRID:AB_309976) and KMT2A N2 (rabbit monoclonal anti-KMT2A used at a 1:100 dilution; Cell Signaling Technology, clone D2M7U, cat. no. 14689; https://scicrunch.org/resolver/RRID:AB_2688009) as well as two antibodies targeting the KMT2A C terminus, KMT2A C1 (mouse monoclonal anti KMT2A used at a 1:100 dilution; Santa Cruz Biotechnology, clone H-10, cat. no. sc-374392; https://scicrunch.org/resolver/RRID:AB_10988264), and KMT2A C2 (mouse monoclonal anti KMT2A used at a 1:100 dilution; Millipore, clone 9-12, cat. no. 05-765; https://scicrunch.org/resolver/RRID:AB_309977). To profile the occupancy of the oncoprotein fusion partner AF4, we used two antibodies: (1) mouse monoclonal anti AF4 used at a 1:50 dilution; MyBioSource, cat. no MBS190886, (2) rabbit polyclonal anti AF4 used at a 1:50 dilution; Thermo Fischer Scientific, cat no. PA5-77068; RRID:AB_2720795. Reads from CUT&RUN profiles of the same samples using the two AF4 antibodies were the pooled for downstream analysis. For all mouse primary antibodies, we also included a subsequent incubation with the secondary rabbit anti-mouse IgG antibody (used at a 1:100 dilution; Abcam, cat. no. ab46540; https://scicrunch.org/resolver/RRID:AB_2614925); this secondary serves as an adaptor to ensure efficient binding of pA-MNase. This rabbit anti-mouse IgG antibody was also run in the absence of a primary antibody as an IgG negative control for each sample.

For AutoCUT&Tag profiling of the KMT2A-oncoprotien transcriptional cofactors we used Rabbit anti menin at a 1:50 dilution (Bethyl Cat# A300-105A); rabbit anti Dot1L at a 1:50 dilution (Cell Signaling Tech Cat# 90878S); and rabbit anti ENL at a 1:50 dilution (Cell Signaling Tech Cat# 14893S). For AutoCUT&Tag profiling of H3K4me3 we used Rabbit anti H3K4me3 at a 1:50 dilution (Active Motif Cat# 39159). To increase the number of pA-Tn5 molecules tethered to each antibody target site, all CUT&Tag reactions

included the secondary antibody Guinea Pig anti-Rabbit IgG (1:100, antibodies-online Cat# ABIN101961).

DNA sequencing and data processing

AutoCUT&RUN and AutoCUT&Tag sample processing was performed by the Fred Hutch Cancer Center Genomics Shared Resources Facility according to previously published protocols available on the protocols.io website (dx.doi.org/10.17504/protocols.io.ufeetje and dx.doi.org/10.17504/protocols.io.bgztjx6n). We used the Agilent 4200 Bioanalyzer to assess the AutoCUT&RUN libraries. Up to 96 samples were pooled at equimolar concentrations and sequenced on a NextSeq 2000 instrument with a P2-100 flow cell by the Fred Hutchinson Cancer Research Center Genomics Shared Resources. This yielded 5-10 million 2 x 50 bp paired end reads per sample. To remove adapter sequences, we preprocessed the reads using cutadapt version 2.9 with parameters -j 8 --nextseq-trim 20 -m 20 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -Z. Paired-end reads were then aligned to the UCSC hg38 human genome build using Bowtie2 version 2.4.4 with parameters --very-sensitive-local --soft-clipped-unmapped-tlen --dovetail --no-mixed --no-discordant -q --phred33 -I 10 -X 1000.

To generate the KMT2A N-terminal and C-terminal bed files for each sample, we first removed duplicate reads from individual replicates and then pooled and sorted the remaining reads from the two KMT2A N-terminal samples into a single file, and the two KMT2A C-terminal samples into a second file. These combined KMT2A N-terminal and KMT2A C-terminal bed files were then used to generate coverage normalized bedgraph and bigwig files.

Identification of KMT2A target sites

We began by assembling a list of all KMT2A target sites and called peaks on the KMT2A N and C terminal files. We called peaks with SEACR[47] version 1.3 using the “stringent”, “non” normalized settings with a false-discovery rate (FDR) threshold of 0.01. We then assembled a master KMT2A peak list by concatenating and merging all the individual KMT2A N and C terminal peak sets and removing non-specific background peaks and repetitive elements. Specifically, we combined the IgG negative control reads from samples processed on the same 96 well plate, and called peaks using the “stringent”, “non” normalized settings with a variable FDR threshold ranging from 0.001 to 0.0001 to avoid calling more than 2500 IgG peaks for any group of samples. We then removed any of the KMT2A peaks that overlapped with the peaks called on the IgG sample set.

Comparison of KMT2A oncoprotein target sites between samples

To call KMT2A oncoprotein target sites, we first quantified the number of base pairs from the N-terminal and C-terminal signal of each sample that overlap with all the combined KMT2A peaks genomic interval. We then used these values to assign an “oncoprotein score” to each interval. Specifically, a Bayesian model was implemented, assuming two Multinomial distributions for the C-terminal and N-terminal reads. Each distribution was modeled with a uniform prior, indicating no preconceived knowledge of the distribution of reads into the peaks. Under this model, the fraction of C-terminal reads assigned to each peak (denoted as p_C) and the fraction of N-terminal reads assigned to each peak (denoted as p_N) were derived from the posterior distribution, assuming a Dirichlet distribution based on the count of reads (either N-terminal or C-terminal) in each interval plus a vector of ones. This implies the ratio between the two, r_{NC}, can be obtained as $\log_2(p_N/p_C)$; this value is referred to as the N/C score.

To estimate the mean and standard deviation of the N/C score for each interval, Monte Carlo sampling was performed. A total of 1000 samples (N = 1000) were drawn from this posterior distribution and the mean (μ_i) and standard deviation (s_i) were calculated for each interval. These statistics provided the necessary information to compute a z-score and p-value for the N/C score of each genomic interval. The resulting p-values were then corrected for multiple testing using the Benjamini/Hochberg correction method, using the fdrcorrection function from the statsmodels Python package. This procedure offers a rigorous approach to statistically measure the differential distribution of N-terminal and C-terminal CUT&RUN reads across the genomic intervals of interest. For each sample, intervals with an N/C score greater than 1.75 and a p-value less than 0.00001 were called as oncoprotein target sites and these sites were combined to obtain a master list of 1598 oncoprotein target sites across all samples. We then calculated the “oncoprotein score” for all 1598 intervals in each sample by multiplying the N/C score by the $-\log_{10}$ of the p-value plus a pseudo count of 1E-10. The average oncoprotein score was then calculated by taking the average of these values for all 1598 sites.

Comparisons to RNA-seq

To examine the relative RNA expression in a cohort of six lineage-switching samples, the previously determined RPKM values for all of the available B-ALL-specific and AML-specific oncoprotein-target genes in the LS01, LS03, LS04, LS05, LS06 and LS10 primary and relapsed patient samples were pulled from Supplementary Table 2 of Tirtakusuma et al. 2022. We then transformed these RPKM values according to the z-scale of each gene across all 12 samples, and compared the differences in z-scaled expression for B-ALL-specific and AML-specific oncoprotein-target genes between primary (B-ALL) and relapsed (AML) patient samples.

To determine how the B-ALL-specific and AML-specific oncoprotein-target genes change in response to revumenib we pulled all of the log2FoldChange values that were available in the “suppTab.9_RNAseq_DEGs_non-genet” of Perner et al. 2023. We then directly compared the log2FoldChange values from the PDX 4 sample and Patient 9 sample (referred to as patient 15 in suppTab.9) between the canonical B-ALL-specific, canonical AML-specific and non-canonical AML-specific genes identified in the present study.

Statistics

Comparisons between the average oncoprotein scores for the four clusters of KMT2Ar leukemia samples, were performed using the using two-tailed independent samples t-tests. To account for multiple comparisons and control the family-wise error rate, a Bonferroni correction was applied. This correction involved dividing the conventional alpha level (0.05) by the number of comparisons being made. Only p-values that met the corrected threshold for significance were considered statistically significant.

To compare the RNA-Seq gene expression values we assumed a non-normal distribution and employed the Mann-Whitney-Wilcoxon test, a non-parametric equivalent to the independent samples t-test. This test was performed as a two-sided test to detect differences in both directions. Given multiple pairwise comparisons, we adjusted for the increased risk of Type I error by applying the Bonferroni correction. Specifically, the significance level for each test was set by dividing the conventional alpha level (0.05) by the number of pairwise comparisons

All statistical analyses were conducted using the add_stat_annotation function in Python, which automates the application of the specified statistical test and adjustment across the defined box pairs.

Data Availability

The primary CUT&RUN and CUT&Tag sequencing data that was generated for this study has been deposited as paired-end fastq files into the Gene Expression Omnibus and is available under the accession code GSE252378.

References

1 Hu, D. & Shilatifard, A. Epigenetics of hematopoiesis and hematological malignancies. *Genes Dev* **30**, 2021-2041 (2016). https://doi.org:10.1101/gad.284109.116

2 Winters, A. C. & Bernt, K. M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4 (2017). https://doi.org:10.3389/fped.2017.00004

3 Balgobind, B. V. *et al.* Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. *Blood* **114**, 2489-2496 (2009). https://doi.org:10.1182/blood-2009-04-215152

4 Szczepanski, T., Harrison, C. J. & van Dongen, J. J. Genetic aberrations in paediatric acute leukaemias and implications for management of patients. *Lancet Oncol* **11**, 880-889 (2010). https://doi.org:10.1016/S1470-2045(09)70369-9

5 Issa, G. C. *et al.* The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia. *Nature* **615**, 920-924 (2023). https://doi.org:10.1038/s41586-023-05812-3

6 van der Sluis, I. M. *et al.* Blinatumomab Added to Chemotherapy in Infant Lymphoblastic Leukemia. *N Engl J Med* **388**, 1572-1581 (2023). https://doi.org:10.1056/NEJMoa2214171

7 Lambo, S. *et al.* A longitudinal single-cell atlas of treatment response in pediatric AML. *Cancer Cell* (2023). https://doi.org:10.1016/j.ccell.2023.10.008

8 Isobe, T. *et al.* Multi-omics analysis defines highly refractory RAS burdened immature subgroup of infant acute lymphoblastic leukemia. *Nat Commun* **13**, 4501 (2022). https://doi.org:10.1038/s41467-022-32266-4

9 Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* **47**, 330-337 (2015). https://doi.org:10.1038/ng.3230

10 Tirtakusuma, R. *et al.* Epigenetic regulator genes direct lineage switching in MLL/AF4 leukemia. *Blood* **140**, 1875-1890 (2022). https://doi.org:10.1182/blood.2021015036

11 Gardner, R. *et al.* Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy. *Blood* **127**, 2406-2410 (2016). https://doi.org:10.1182/blood-2015-08-665547

12 Khabirova, E. *et al.* Single-cell transcriptomics reveals a distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic leukemia. *Nat Med* **28**, 743-751 (2022). https://doi.org:10.1038/s41591-022-01720-7

13 Liao, W., Kohler, M. E., Fry, T. & Ernst, P. Does lineage plasticity enable escape from CAR-T cell therapy? Lessons from MLL-r leukemia. *Exp Hematol* **100**, 1-11 (2021). https://doi.org:10.1016/j.exphem.2021.07.002

14 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 e611 (2019). https://doi.org:10.1016/j.ccell.2019.11.001

15 Klossowski, S. *et al.* Menin inhibitor MI-3454 induces remission in MLL1-rearranged and NPM1-mutated models of leukemia. *J Clin Invest* **130**, 981-997 (2020). https://doi.org:10.1172/JCI129126

16 Perner, F. *et al.* MEN1 mutations mediate clinical resistance to menin inhibition. *Nature* **615**, 913-919 (2023). https://doi.org:10.1038/s41586-023-05755-9

17 Soto-Feliciano, Y. M. *et al.* A Molecular Switch between Mammalian MLL Complexes Dictates Response to Menin-MLL Inhibition. *Cancer Discov* **13**, 146-169 (2023). https://doi.org:10.1158/2159-8290.CD-22-0416

18 Chen, C. *et al.* Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell-like blasts in KMT2A-rearranged leukemia. *Blood* **139**, 2198-2211 (2022). https://doi.org:10.1182/blood.2021013442

19 Janssens, D. H. *et al.* Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat Genet* **53**, 1586-1596 (2021). https://doi.org:10.1038/s41588-021-00941-9

20 Janssens, D. H. *et al.* Automated in situ chromatin profiling efficiently resolves cell types and gene

regulatory programs. *Epigenetics Chromatin* **11**, 74 (2018). https://doi.org:10.1186/s13072-018-0243-8

21 Meyer, C. *et al.* The KMT2A recombinome of acute leukemias in 2023. *Leukemia* **37**, 988-1005 (2023). https://doi.org:10.1038/s41375-023-01877-1

22 Olsen, S. N. *et al.* MLL::AF9 degradation induces rapid changes in transcriptional elongation and subsequent loss of an active chromatin landscape. *Mol Cell* **82**, 1140-1155 e1111 (2022). https://doi.org:10.1016/j.molcel.2022.02.013

23 Otto, D. J., Jordan, C., Dury, B., Dien, C. & Setty, M. Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. *Nat Methods* (2024). https://doi.org:10.1038/s41592-024-02302-w

24 Germano, G. *et al.* ZNF521 sustains the differentiation block in MLL-rearranged acute myeloid leukemia. *Oncotarget* **8**, 26129-26141 (2017). https://doi.org:10.18632/oncotarget.15387

25 Chiarella, E. *et al.* ZNF521 Enhances MLL-AF9-Dependent Hematopoietic Stem Cell Transformation in Acute Myeloid Leukemias by Altering the Gene Expression Landscape. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms221910814

26 Kang, H. *et al.* Gene expression profiles predictive of outcome and age in infant acute lymphoblastic leukemia: a Children's Oncology Group study. *Blood* **119**, 1872-1881 (2012). https://doi.org:10.1182/blood-2011-10-382861

27 Stam, R. W. *et al.* Gene expression profiling-based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants. *Blood* **115**, 2835-2844 (2010). https://doi.org:10.1182/blood-2009-07-233049

28 Trentin, L. *et al.* Two independent gene signatures in pediatric t(4;11) acute lymphoblastic leukemia patients. *Eur J Haematol* **83**, 406-419 (2009). https://doi.org:10.1111/j.1600-0609.2009.01305.x

29 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). https://doi.org:10.1038/onc.2016.488

30 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). https://doi.org:10.1016/j.ccell.2016.10.008

31 Luo, Z. *et al.* The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. *Mol Cell Biol* **32**, 2608-2617 (2012). https://doi.org:10.1128/MCB.00182-12

32 Mohan, M. *et al.* Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev* **24**, 574-589 (2010). https://doi.org:10.1101/gad.1898410

33 Adolfsson, J. *et al.* Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295-306 (2005). https://doi.org:10.1016/j.cell.2005.02.013

34 Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol* **21**, 2-8 (2019). https://doi.org:10.1038/s41556-018-0227-8

35 Wan, L. *et al.* ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* **543**, 265-269 (2017). https://doi.org:10.1038/nature21687

36 Bernt, K. M. *et al.* MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66-78 (2011). https://doi.org:10.1016/j.ccr.2011.06.010

37 Bernt, K. M. & Armstrong, S. A. Targeting epigenetic programs in MLL-rearranged leukemias. *Hematology Am Soc Hematol Educ Program* **2011**, 354-360 (2011). https://doi.org:10.1182/asheducation-2011.1.354

38 Chen, C. W. *et al.* DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nat Med* **21**, 335-343 (2015). https://doi.org:10.1038/nm.3832

39 Yokoyama, A. & Cleary, M. L. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* **14**, 36-46 (2008). https://doi.org:10.1016/j.ccr.2008.05.003

40 Erb, M. A. *et al.* Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **543**, 270-274 (2017). https://doi.org:10.1038/nature21688

41 Garnar-Wortzel, L. *et al.* Chemical Inhibition of ENL/AF9 YEATS Domains in Acute Leukemia. *ACS Cent Sci* **7**, 815-830 (2021). https://doi.org:10.1021/acscentsci.0c01550

42 Liu, Y. *et al.* Small-Molecule Inhibition of the Acyl-Lysine Reader ENL as a Strategy against Acute Myeloid Leukemia. *Cancer Discov* **12**, 2684-2709 (2022). https://doi.org:10.1158/2159-8290.CD-21-1307

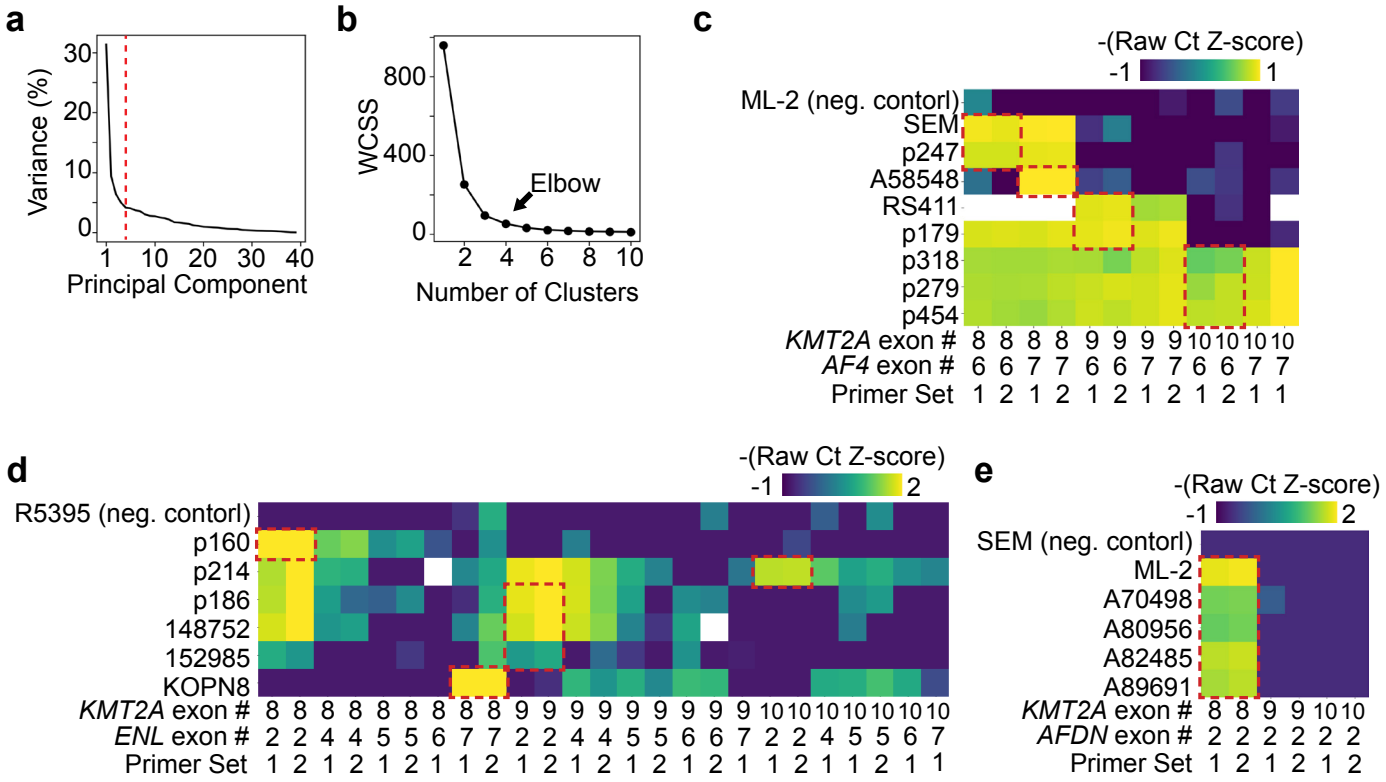
43 Ma, X. R. *et al.* Discovery of Selective Small-Molecule Inhibitors for the ENL YEATS Domain. *J Med Chem* **64**, 10997-11013 (2021). https://doi.org:10.1021/acs.jmedchem.1c00367

44 Godfrey, L. *et al.* DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat Commun* **10**, 2803 (2019). https://doi.org:10.1038/s41467-019-10844-3

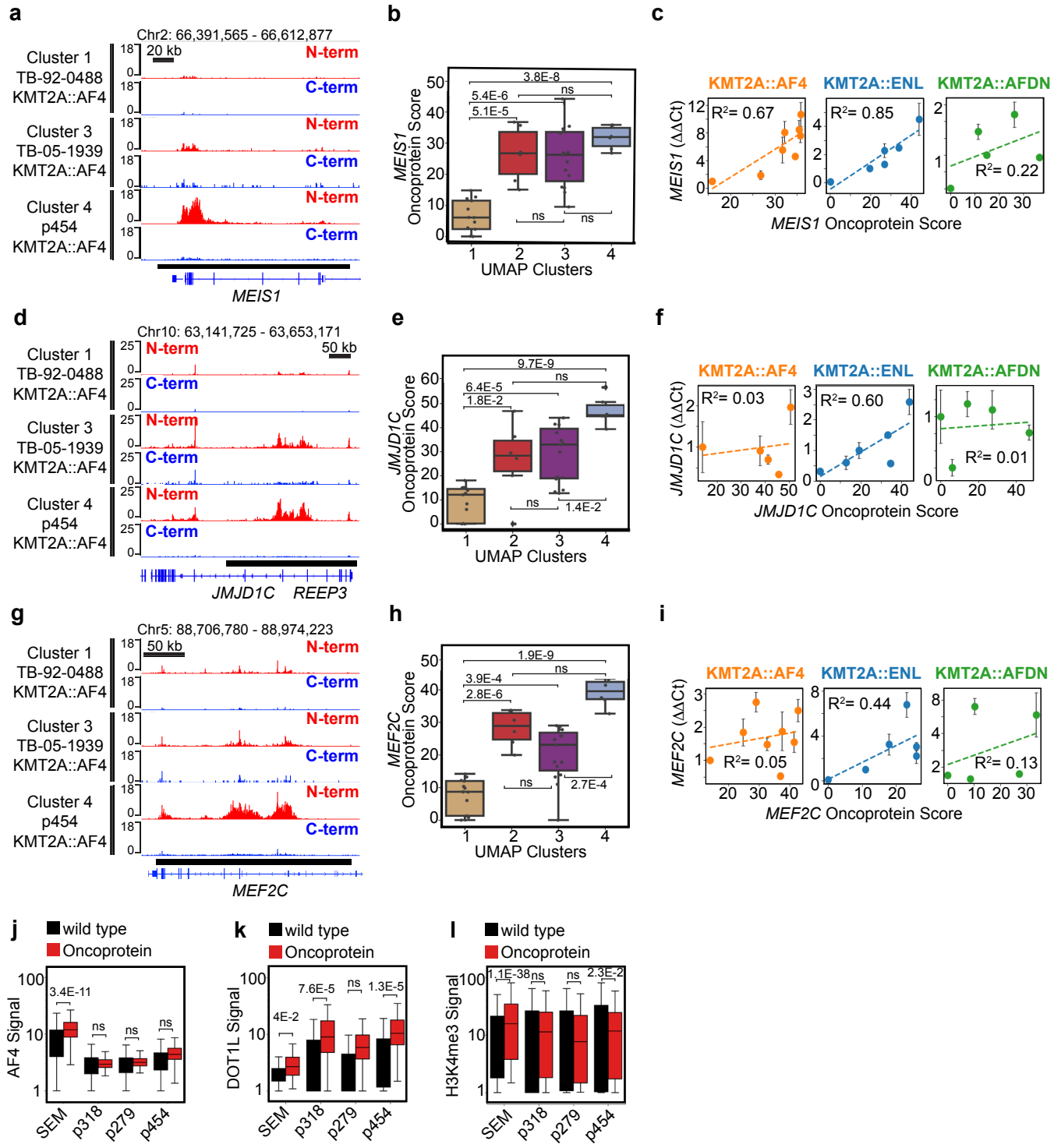
45 Stein, E. M. *et al.* The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood* **131**, 2661-2669 (2018). https://doi.org:10.1182/blood-2017-12-818948

46 Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103-112 (2018). https://doi.org:10.1038/nm.4439

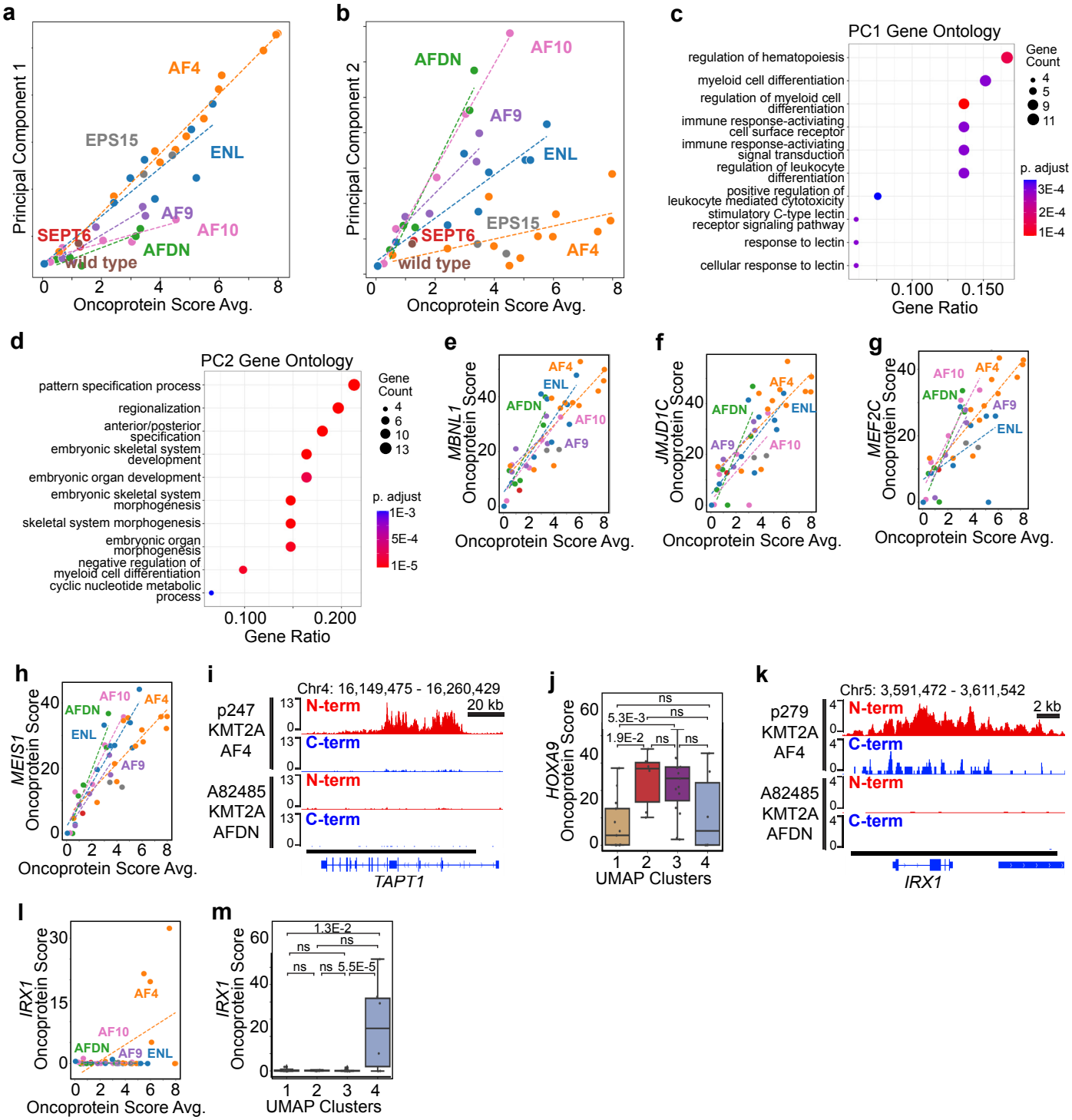
47 Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42



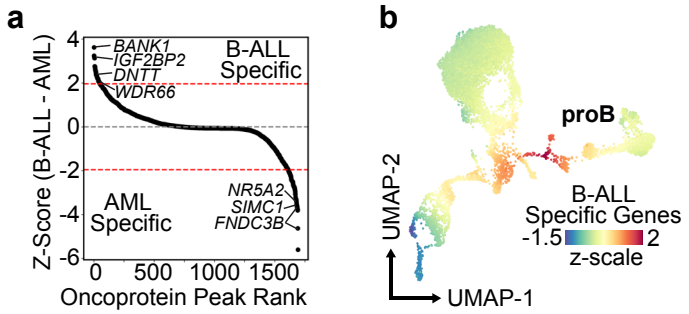
Extended Data Fig 1: Comparisons of KMT2A-oncoprotein scores and identification of minimal KMT2A-fusion-partner exon junctions. a, Rank order plot of the variance in oncoprotein scores detected by principal component analysis. The first 4 components (red line) capture >50% of the total variance. **b,** Elbow plot of the K-means Within-Cluster Sum of Squares (WCSS) scores suggests the *KMT2A*r leukemias form 4 clusters in UMAP space. **c,** Heatmap of the z-scaled Raw ct values for qPCR reactions targeting the exons flanking the most common breakpoints in the *KMT2A::AF4* translocations. Red boxes indicate the minimal KMT2A-fusion exon junctions. These primers were used to compare the relative oncogene expression between samples with the same minimal exon junctions in Fig.1 h. **d,** Same as (c) for *KMT2A::ENL* exon junctions. **e,** Same as (c) for *KMT2A::AFDN* exon junctions.



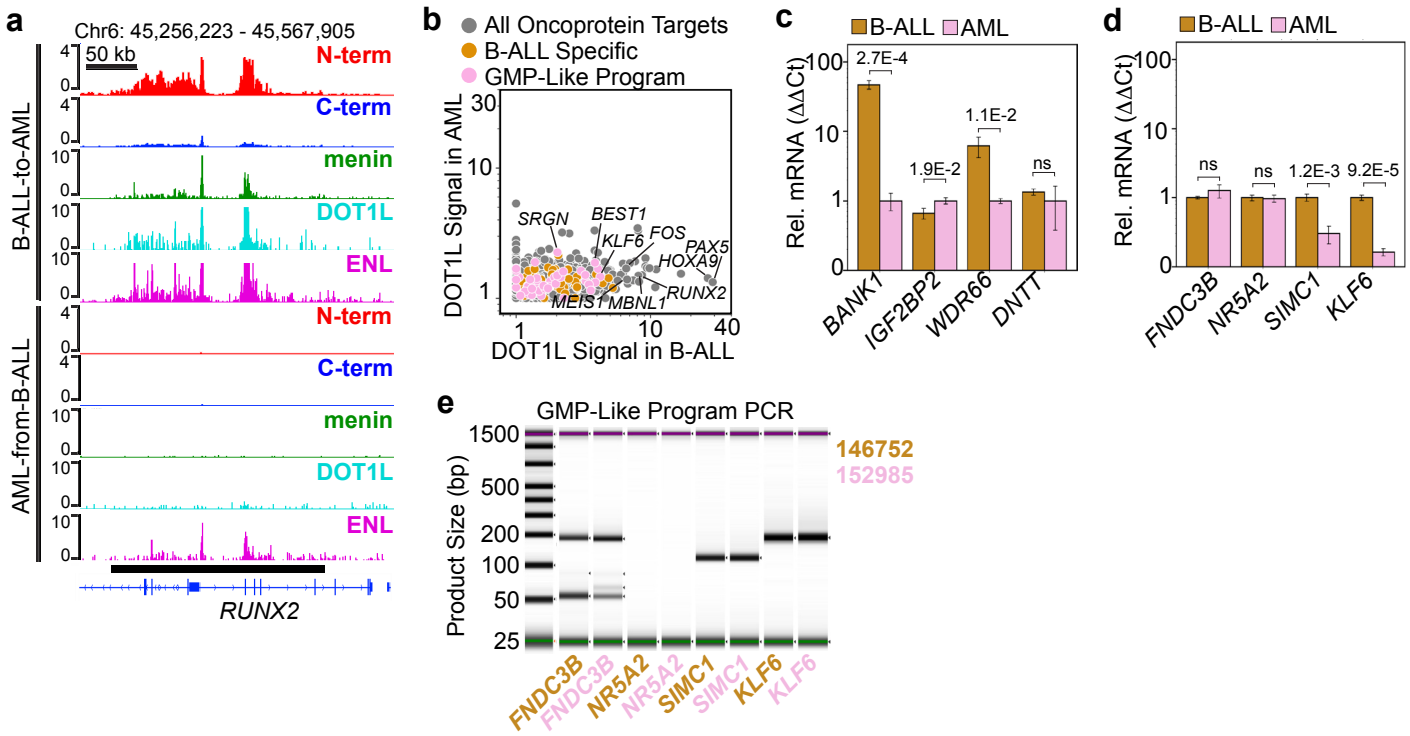
Extended Data Fig 2: KMT2A oncoproteins have heterogeneous expression levels that result in differential oncoprotein-target gene regulation. a, Genome Browser tracks of the *MEIS1* locus show the difference in KMT2A N-terminal (red) and C-terminal (blue) signal in three *KMT2A::AF4* rearranged leukemias, indicated by the white asterisks in (Fig. 1d); bottom black bar = oncoprotein-target peak. **b,** Boxplot comparison of the oncoprotein scores over *MEIS1* split into the four UMAP clusters from Fig. 1d; *p* values were computed using a two-tailed independent samples t-test; Cluster 1 n = 13, Cluster 2 n = 6, Cluster 3 n = 15, Cluster 4 n = 5. **c,** Scatterplots showing the correlation between the gene-specific oncoprotein scores and the expression of *MEIS1*; dotted lines indicate the regression and R^2 measures the fit to the data. **d,** Same as (a) showing the *JMJD1C* locus. **e,** Same as (b) showing *JMJD1C* oncoprotein scores. **f,** Same as (c) showing *JMJD1C* oncoprotein scores and expression. **g,** Same as (a) showing the *MEF2C* locus. **h,** Same as (b) showing *MEF2C* oncoprotein scores. **i,** Same as (c) showing *MEF2C* oncoprotein scores and expression. **j,** Boxplot showing the relative enrichment of AF4 over wild-type and oncoprotein-target sites in four *KMT2A::AF4*-rearranged samples. *p* values were computed using a two-tailed independent samples t-test; *n* values are listed as wild-type site #, oncoprotein site #: SEM = 15611, 491; p318 = 22014, 246; p279 = 25672, 272; p454 = 18509, 438. **k,** Same as (j) but comparing the DOT1L signal intensity. **l,** Same as (j) but comparing the H3K4me3 signal intensity.



Extended Data Fig 3: Principal Components 1 and 2 capture lineage-specific and fusion-partner-dependent oncoprotein-target genes. a, Scatterplot showing the average oncoprotein scores for AF4 and ENL fusion oncoproteins are more strongly associated with Principal Component 1. **b,** The average oncoprotein scores for AFDN and AF10 fusion oncoproteins are more strongly associated with Principal Component 2. **c,** Dot plot showing the significance of the top GO terms related to the oncoprotein-target genes that contribute to Principal Component 1. **d,** Same as (c) showing the GO terms associated with Principal Component 2. **e,** Scatterplot showing the *MBNL1* oncoprotein scores relative to the average oncoprotein score in each sample. Samples are colored by the KMT2A-fusion partner and dotted lines indicate the regression. **f,** Same as (e), but showing the *JMJD1C* oncoprotein scores. **g,** Same as (e), but showing the *MEF2C* oncoprotein scores. **h,** Same as (e), but showing the *MEIS1* oncoprotein scores. **i,** Genome Browser tracks show the KMT2A::AF4 oncoprotein binds the *TAPT1* locus while the KMT2A::AFDN oncoprotein does not (bottom black bar = oncoprotein-target peak). **j,** Boxplot comparison of the oncoprotein scores over *HOXA9* split into the four UMAP clusters from Fig. 1d; *p* values were computed using a two-tailed independent samples t-test; Cluster 1 n = 13, Cluster 2 n = 6, Cluster 3 n = 15, Cluster 4 n = 5. **k,** Same as (i) showing the *IRX1* locus, and the *KMT2A::AF4*-rearranged sample from Fig. 2o in which *HOXA9* is not bound. **l,** Same as (e), but showing the *IRX1* oncoprotein scores. **m,** Same as (j), but showing the *IRX1* oncoprotein scores.



Extended Data Fig 4: Identification of B-ALL and AML-specific oncoprotein-target genes during lineage switching. a, Rank order plot of the difference in oncoprotein scores between the B-ALL and AML patient matched sample. A threshold of the top 5% and bottom 5% of sites was used to call B-ALL-specific and AML-specific oncoprotein-target sites, respectively. **b,** The UMAP embedding of healthy lineage-depleted bone marrow colored by the average z-scores of the 66 BALL-specific oncoprotein target genes from Fig. 3g.



Extended Data Fig 5: Loss of oncoprotein binding and altered target-gene expression in the menin-inhibitor-treated lineage-switching sample. a, Genome Browser tracks showing the oncoprotein and the cofactors menin, DOT1L and ENL bound to the *RUNX2* locus in the B-ALL sample (top). In the patient-matched AML sample treated with the menin inhibitor only ENL remains bound (bottom). **b,** Scatterplot comparing the normalized DOT1L levels in the B-ALL and the AML sample treated with the menin inhibitor over an internal control group of oncoprotein-target sites (grey) the B-ALL-specific program (gold) and the GMP-like program (pink). **c,** The B-ALL-specific oncoprotein target genes *BANK1* and *WDR66* are expressed at significantly higher levels in the B-ALL sample (gold) prior to lineage switching than the AML sample (pink). Error bars = standard deviation; *p* value computed using the using a two-tailed independent samples t-test; $n = 3$ qPCR replicates. **d,** Expression of the GMP-like program *SIMC1* and *KLF6* genes is reduced in the menin-inhibitor-treated AML sample. **e,** Real time PCR indicates *FNDC3B*, *SIMC1* and *KLF6* are expressed in the B-ALL and AML menin-inhibitor-treated lineage-switching samples and that *NR5A2* is not expressed.