KMT2A oncoproteins induce epigenetic resistance to targeted therapies.

Derek H. Janssens[1], Melodie Duran[1], Dominik J. Otto[1,2], Danielle Kirkey[3,4], Joanna S. Yi[5], Soheil Meshinchi[3,4], Jay F. Sarthy[6], Kami Ahmad[1+], Steven Henikoff[1,7+].

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

[+]Corresponding author. Email: kahmad@fredhutch.org; steveh@fredhutch.org.

Abstract

Chromosomal translocations involving the *Lysine-Methyl-Tansferase-2A* (*KMT2A*) locus generate potent oncogenes that cause highly aggressive acute leukemias[1]. *KMT2A* and the most frequent translocation partners encode proteins that interact with DNA to regulate developmental gene expression[2]. KMT2A-oncogenic fusion proteins (oncoproteins) contribute to the epigenetic mechanisms that allow *KMT2A*-rearranged leukemias to evade targeted therapies. By profiling the oncoprotein-target sites of 34 *KMT2A*-rearranged leukemia samples, we find that the genomic enrichment of oncoprotein binding is highly variable between samples. At high levels of expression, the oncoproteins preferentially activate either the lymphoid or myeloid lineage program depending on the fusion partner. These fusion-partner-dependent binding sites correspond to the frequencies of each mutation in acute lymphoid leukemia versus acute myeloid leukemia. By profiling a sample that underwent a lymphoid-to-myeloid lineage switching event in response to lymphoid-directed treatment, we find the global oncoprotein levels are reduced and the oncoprotein-target gene network changes. At lower levels of expression, the oncoprotein shifts to a non-canonical regulatory program that favors the myeloid lineage, and in a subset of resistant patients, the Menin inhibitor Revumenib induces a similar response. The dynamic shifts in KMT2A oncoproteins we describe likely contribute to epigenetic resistance of *KMT2A*-rearranged leukemias 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, a subset of *KMT2A*r leukemias 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[9-11]. 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[12,13]. In phase I clinical trials of the Menin-inhibitor Revumenib a reduction in the oncoprotein target gene expression was correlated with disease remission[5]. Mutations in *Menin* that block the interaction with Revumenib act as a genetic mechanism of resistance, however, some *KMT2A*r AMLs relapse in the absence of *Menin* mutations[14], suggesting that an epigenetic resistance mechanism may also exist. Single cell RNA sequencing and chromatin profiling studies have identified considerable intra-tumoral heterogeneity in *KMT2A*r leukemia samples[8,15], but why these leukemias 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[16], 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 lineage-instructive role to a non-instructive role. The same non-canonical AML gene network that is activated upon lineage switching is also upregulated in response to Revumenib, and this may act as a shared mechanism of epigenetic resistance that allows *KMT2A*r leukemias to evade targeted therapies.

Results

Heterogeneous KMT2A oncoprotein levels

We used AutoCUT&RUN[17] to profile KMT2A-oncoprotein binding sites in 34 *KMT2A*r leukemia samples, including 4 cells lines, 12 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 oncopro-

tein, 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 7 false-positive sites in the CD34+, oncoprotein-negative control sample, and identified 181 oncoprotein targets in the SEM cell line (**Fig. 1c**). By applying this threshold to all 34 *KMT2A*r leukemia samples, we identified 1598 oncoprotein-binding sites and 869 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 1598 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**). Cluster 1 includes the CD34+ oncoprotein-negative control sample as well as 8 AML samples, 2 infant ALLs and 1 MPAL sample, and the oncoprotein signal is barely detectable in these samples (**Fig. 1d**). The average oncoprotein scores of clusters 2 and 3 are significantly higher than those found in cluster 1, and the oncoprotein scores were highest for samples within cluster 4 (**Fig. 1d, Extended Data Fig. 1c**). The *GNAQ* locus is frequently called as an oncoprotein target gene in leukemias bearing the *KMT2A-AFF1* translocations, but the magnitude of the KMT2A N and C terminal enrichment over the *GNAQ* locus is visibly different between the samples in clusters 1, 3 and 4 (**Fig. 1e**). The differences we observe in oncoprotein signals over the *GNAQ* locus are representative of differences in the average oncoprotein scores between clusters (**Fig. 1d,e, Extended Data Fig. 1c**). These differences in oncoprotein scores suggest that among leukemias bearing the most common *KMT2A* translocations the levels of the oncoprotein are extremely heterogeneous.

In addition to the oncoprotein levels, the lineage

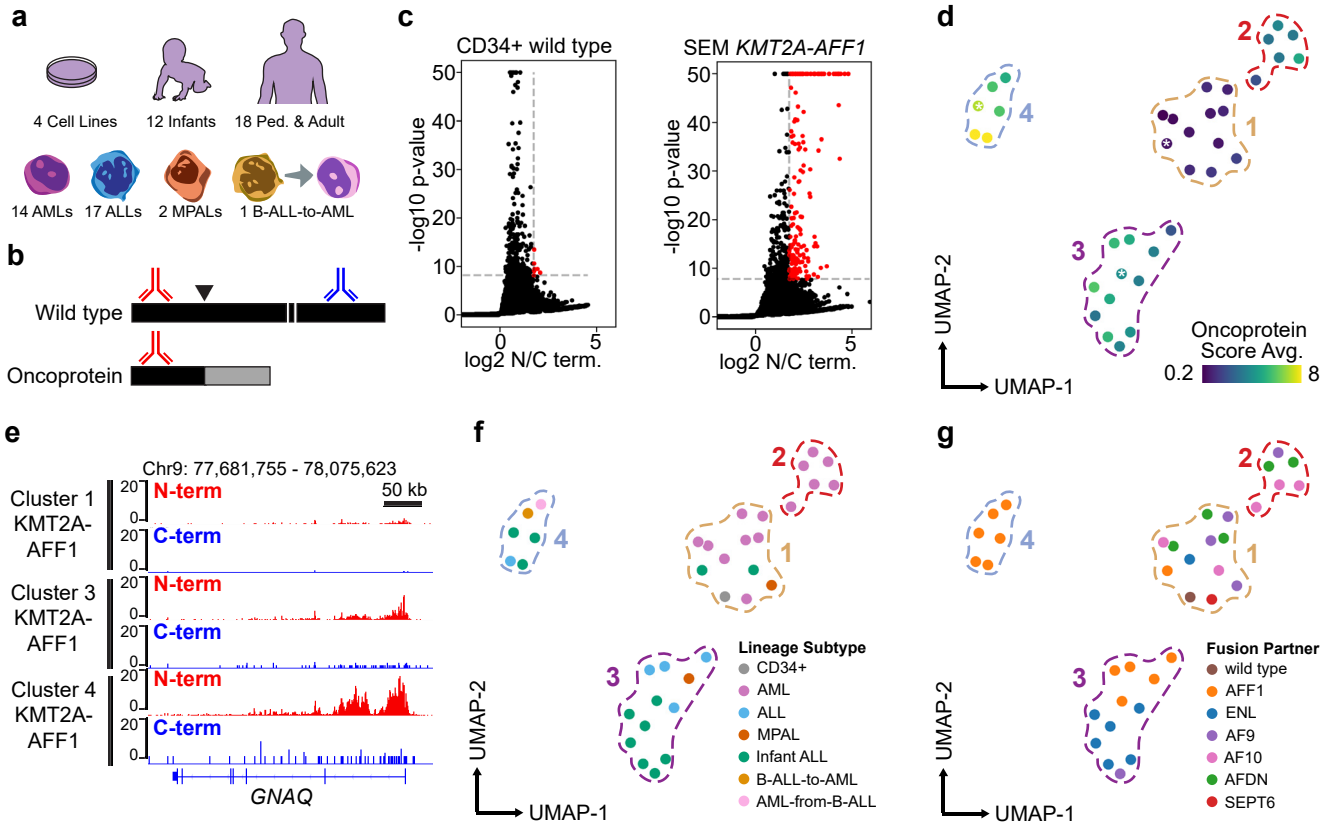


Fig 1: Genomic enrichment of KMT2A-oncoproteins is highly heterogeneous. a, We performed AutoCUT&RUN on a collection of *KMT2A*r cell lines and patient samples including a range of patient ages and lineage subtypes. **b,** Comparison of CUT&RUN profiles using antibodies targeting the KMT2A N terminal (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 colored dotted lines. Samples are colored according to the average oncoprotein scores (N/C terminal magnitude x significance)**. e,** Genome Browser tracks show the difference in KMT2A N terminal (red) and C terminal (blue) signal over the *GNAQ* locus in three *KMT2A-AFF1* bearing leukemias indicated by the white asterisks in (d). **f,** The UMAP embedding from (d) colored according to the lineage subtype. **g,** Same as (d) but colored according to the KMT2A oncoprotein fusion partner.

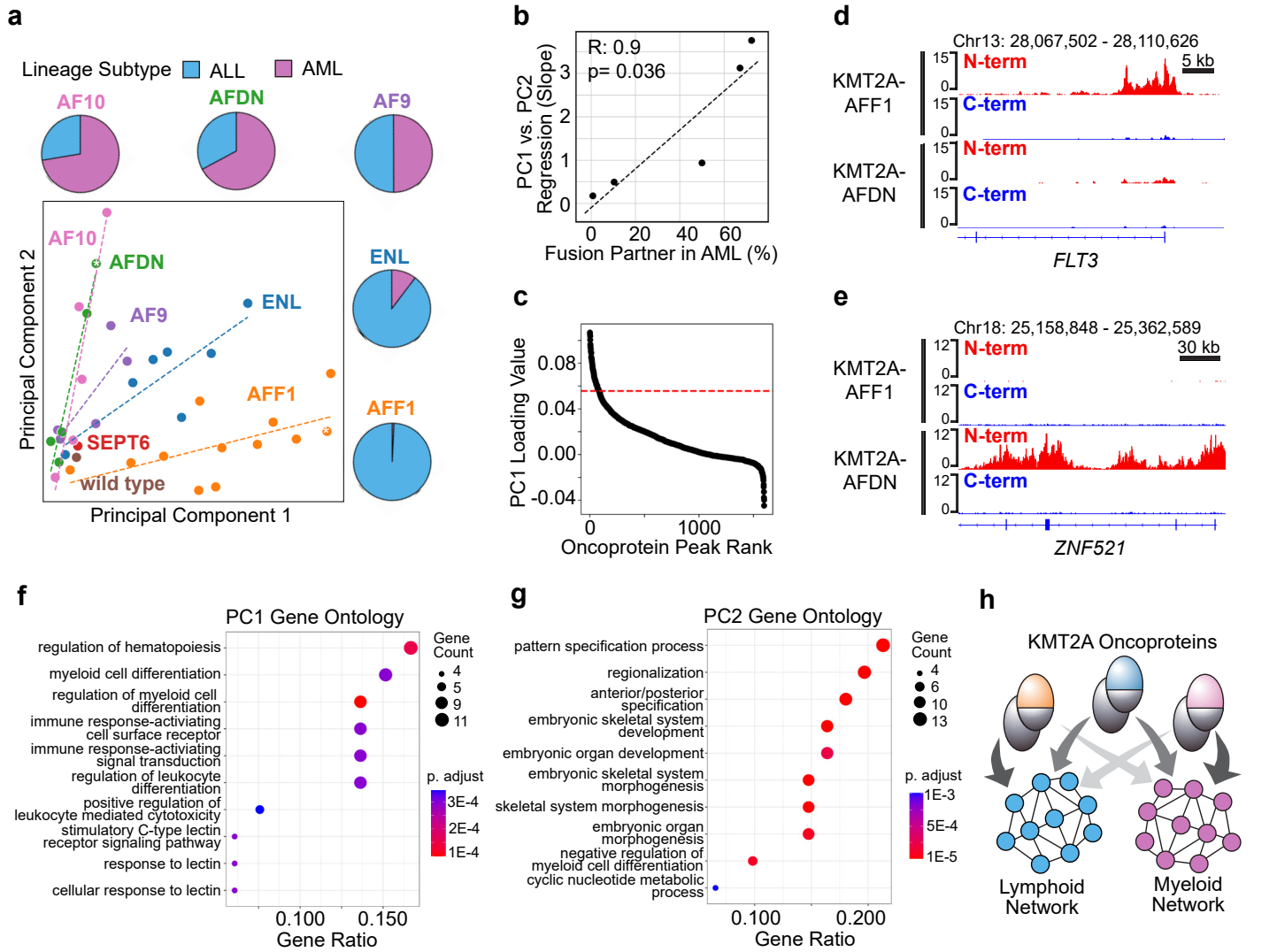


Fig 2: At elevated levels the KMT2A-oncoproteins are instructive of lineage-specific genomic binding sites. a, Principal Components 1 and 2 organize *KMT2A*r leukemia samples according to the oncoprotein-fusion partners. Dotted regression lines are included for samples bearing related translocations. Pie charts indicate the frequency of each KMT2A-fusion partner found in ALL versus AML[18]. **b,** Scatterplot showing that 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 loading values (indicated by the red line) were identified as the sites that contribute most to the variance in Principal Component 1. **d,** Genome Browser tracks show the difference in KMT2A N terminal (red) and C terminal (blue) signal over the *FLT3* locus between a *KMT2A-AFF1* bearing leukemia and a *KMT2A-AFDN* bearing leukemia indicated by white asterisks in (a). **e,** Same as (d) showing the *ZNF521* locus. **f,** Dot plot showing the significance of the top GO terms related to the oncoprotein target gene that contribute to Principal Component 1. **g,** Same as (f) showing the GO terms associated with Principal Component 2. **h,** The KMT2A-fusion partners direct the genomic occupancy of the oncoproteins to activate either lymphoid or myeloid lineage programs.

identity and translocation-partner genes also 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. 1f**). In addition to having the highest oncoprotein scores, cluster 4 is unique in that it contains only leukemias bearing the *KMT2A-AFF1* translocation (**Fig. 1g**). The *AFF1* locus is the most common *KMT2A* translocation partner, and this mutation is found almost exclusively in B-ALLs at diagnosis[18]. Interestingly, cluster 4 includes a B-ALL sample as well as a paired relapse sample that switched to an AML identity in response to treatment (**Fig. 1f**). Our results show the KMT2A oncoproteins bind to distinct sites in AMLs versus ALLs, and in a subset of the ALL samples, the KMT2A-AFF1 oncoproteins occupy the chromatin at exceptionally high levels at regions that distinguish them from other *KMT2A*r leukemias.

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-AFF1 and KMT2A-ENL fusion proteins and is tightly associated with the average oncoprotein scores in the ALL samples (**Fig. 2a and Extended Data Fig. 2a**). 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 and Extended Data Fig. 2b**). 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 and Extended Data Fig. 2a,b**). When grouped according to the oncoprotein-fusion partners, the slope of the best-fit lines for PC1 versus PC2 match the prevalence of each translocation in *KMT2A*r ALLs versus AMLs (**Fig. 2a,b**)[18]. This strongly suggests the *KMT2A*-fusion partner influences genome-wide binding sites of the oncoprotein as well as the lineage specificities of the associated translocations.

Next, we identified oncoprotein-binding sites that contribute to the variance in PC1 and PC2 (**Fig. 2c**). *FLT3* encodes a cytokine receptor that is involved in hematopoiesis[19], and variations in oncoprotein binding over the *FLT3* locus contribute to PC1 (**Supplementary Information**). When the oncoproteins are highly abundant, *FLT3* is bound by the KMT2A-AFF1 fusion protein but not the KMT2A-AFDN fusion protein (**Fig. 2d**). Conversely, the *ZNF521* gene is critical for *KMT2A*r AMLs[20,21], and this locus contributes to PC2 (**Fig. 2e, Supplementary Information**). The oncoprotein-target genes in PC1 are enriched for Gene Ontology (GO) terms that include the regulation of hematopoiesis, myeloid differentiation, lymphoid differentiation, and immune responses (**Fig. 2f**). The oncoprotein-target genes in PC2 primarily regulate developmental patterning (**Fig. 2g**). PC1 and PC2 show an intermediate sensitivity to the levels of the KMT2A-AF9 fusion protein, and the group of "highly sensitive" genes that a previous study found are rapidly downregulated in response to auxin-mediated KMT2A-AF9 degradation include genes that contribute to PC1 (*MEF2C*, *CPEB2* and *MEIS1*) as well as PC2 (*SKIDA1*, *PBX3*, *HOXA9* and *SOCS2;* **Supplementary Information**)[22]. We conclude that at high levels, the oncoproteins activate genes that influence the lymphoid versus myeloid identities of the leukemias (**Fig. 2h**).

Oncoprotein dynamics and lineage switching

To characterize the epigenetic mechanisms underlying the lineage plasticity of *KMT2A*r leukemias, we compared the oncoprotein-target genes in a sample that underwent a B-ALL-to-AML lineage-switching event in response to ALL-directed chemotherapy. As compared to the patient-matched B-ALL sample, the oncoprotein scores are generally reduced following relapse as an AML (**Fig. 3a**), and this difference is apparent when comparing the oncoprotein levels over the *BCL7A, HOXA9* and *EBF1* loci (**Fig. 3b, Extended Data Fig. 3a,b**). We also identified sites that show an increase in oncoprotein binding when the leukemia transitions to a myeloid identity, including the promoter of *ZBTB16*, a myeloid transcription factor (**Fig. 3c,d**). By analyzing RNA sequencing data from a collection of *KMT2A*r samples that underwent a B-ALL-to-AML lineage-switching event[9], we find that the changes in gene expression are concordant with B-ALL-specific and AML-specific oncoprotein-target genes (**Fig. 3e,f**). This demonstrates that oncoprotein expression levels and binding sites are dynamic. At diagnosis the oncoprotein is expressed at relatively high levels and contributes to the ALL phenotype by directly activating the expression of

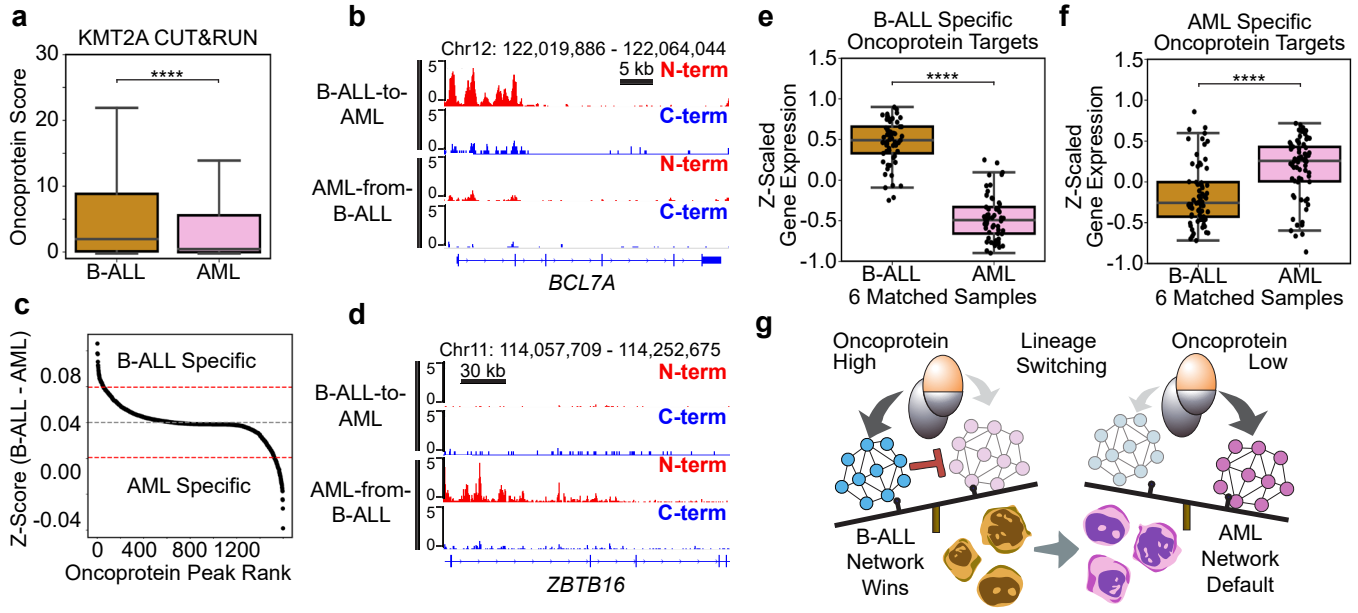


Fig 3: The oncoprotein binding sites change during B-ALL-to-AML lineage-switching events. a, Boxplot comparing the global oncoprotein scores in a patient matched B-ALL and AML sample before and after lineage switching. **b,** Genome Browser tracks showing the oncoprotein levels are elevated over the *BCL7A* locus in the B-ALL sample prior to lineage switching. **c,** 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 and AML specific oncoprotein target sites, respectively. **d,** Genome Browser tracks showing that the oncoprotein levels are elevated over the *ZBTB16* locus in the AML sample after lineage switching. **e,** Boxplot comparing the relative expression of the B-ALL-specific oncoprotein target genes from (c) in a cohort of 6 samples that underwent a BALL-to-AML lineage switching event[9]. **f,** Same as (e) but comparing the AML specific oncoprotein target genes from (c). **g,** At high levels the KMT2A oncoprotein instructs the ALL identify by activating a lymphoid program. During lineage switching the oncoprotein levels are reduced, and the oncoprotein activates an AML program by default.

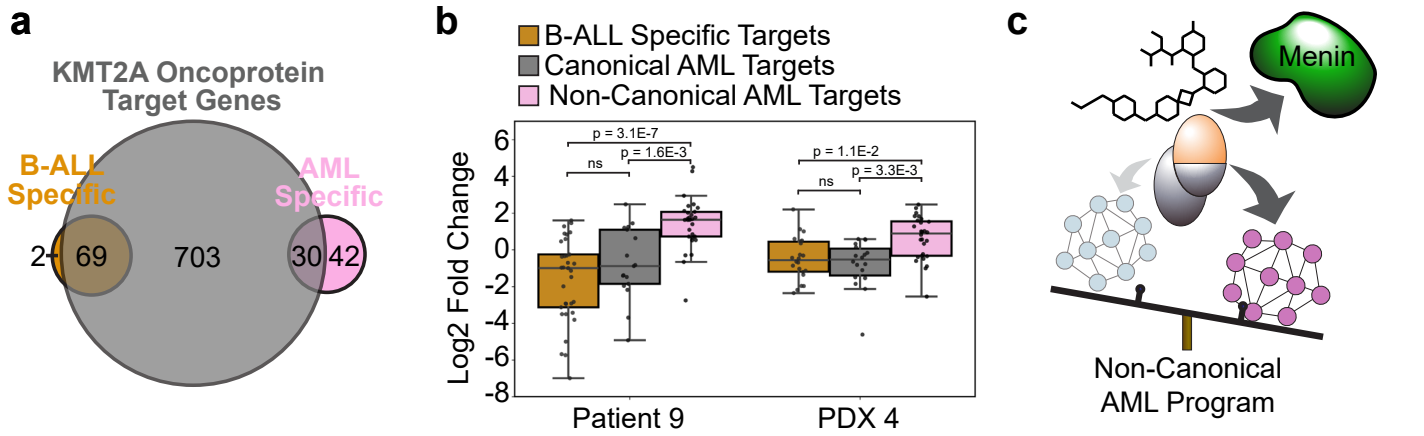


Fig 4: *KMT2A*r leukemias activate a non-canonical AML program in response to Revumenib. a, A Venn Diagram comparing the lineage-switching B-ALL-specific and AML-specific oncoprotein target genes to all other oncoproteins target genes. The AML group includes 42 non-overlapping genes that we designate as a non-canonical AML program. **b,** In two Revumenib resistant patient samples (Patient 9 and PDX 4) [14], the non-canonical AML program (pink boxes) is upregulated in response to Renumenib and this is significantly different than the canonical oncoprotein target genes. **c,** Activation of the non-canonical AML program may promote epigenetic resistance to targeted therapies.

lymphoid-associated genes. In response to therapeutic intervention, the levels of the oncoprotein are reduced and this allows the leukemia to shift to a myeloid lineage (**Fig. 3g**).

A general epigenetic resistance program

Next, we compared the oncoprotein-target genes in the B-ALL-to-AML lineage-switching sample to the 33 other samples we profiled. We find that 69/71 B-ALL enriched oncoprotein-target sites were also identified in at least one other samples (**Fig. 4a, Supplementary Information**). In contrast, 42/72 AML-specific oncoprotein target genes are specific to the AML sample after lineage switching (**Fig. 4a, Supplementary Information**). By reanalyzing the RNA-seq data[9], we find these 42 genes are consistently upregulated during lineage switching across the cohort of 6 other samples, whereas the 30 genes that overlap with other samples are not upregulated during lineage switching (**Extended Data Fig. 4a,b**). We conclude that during lineage switching, the KMT2A-AFF1 oncoprotein activates a collection of target genes that are distinct from those found in other *KMT2A*r AMLs at diagnosis, and we define this group of 42 genes as a "non-canonical" AML program that promotes lineage switching. Many of the genes in this non-canonical AML are normally expressed in granulocytes, more specifically neutrophils[23,24], including *Serglycin* and *Lysosomal-Trafficking-Regulator* which are involved in the synthesis and secretion of hematopoietic granules[25,26].

A subset of *KMT2A*r leukemias initially respond to Revumenib by downregulating oncoprotein-target genes, but for unknown reasons are resistant to treatment, and rapidly relapse[14]. To examine whether this cohort of resistant patients might activate a similar non-canonical AML program in response to Revumenib, we examined the RNA-sequencing data that was previously collected from two Revumenib-resistant AML patient samples[14]. We find the expression of the B-ALL-specific oncoprotein target genes, as well as the 30 canonical AML-specific oncoprotein target genes were generally reduced following treatment with Revumenib (**Fig. 4b**). Strikingly, we find the expression of the non-canonical AML oncoprotein-target genes was upregulated in response to Revumenib, and this change in gene expression is significantly different than the canonical oncoprotein target genes (**Fig. 4b**). This suggests the oncoprotein activates a similar non-canonical AML program in the Revumenib-resistant patients to the one that is activated in the lineage-switching patients (**Fig. 4c**). We propose that a dynamic shift in the binding sites of the oncoprotein activates an alternate leukemia program and this epigenetic mechanism promotes resistance to targeted therapies.

Discussion

The lineage identities and therapeutic responses of *KMT2A*r leukemias are highly heterogeneous[1,5,7,8]. To characterize the molecular basis of this heterogeneity we applied Automated CUT&RUN to profile the oncoprotein binding sites across 34 *KMT2A*r leukemia samples that span the age of onset 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 the oncoprotein-fusion partner influences the oncoprotein-binding sites and is likely instructive of the lineage identity of the leukemia. By profiling a patient sample that underwent a B-ALL-to-AML lineage-switching event in response to ALL-directed chemotherapy, we demonstrate that oncoprotein binding is dynamic. Reductions in the oncoprotein levels allow the oncoprotein to active a non-canonical AML program that promotes lineage switching. We find a similar non-canonical AML program is activated in samples from patients who initially responded to Revumenib with reductions in canonical oncoprotein target genes but rapidly relapsed. We propose that KMT2A-oncoprotein dynamics play a critical role in the epigenetic mechanisms that allow *KMT2A*r leukemias to evade targeted therapies.

Several mechanisms have been proposed to explain why the translocation partner genes *AFF1* and *ENL* are much more common in B-ALLs whereas the *AF10* and

AFDN are more prevalent in AMLs[2,27]. 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, indicating that when the oncoproteins are expressed at high levels they become instructive of lineage identity and dominate the original identity by ectopically activating the expression of lineage-specific genes. At low levels the oncoproteins are "non instructive" with regards to lineage identity, and simply tweak the pre-existing lineage programs. This instructive model is supported by overexpression experiments in mice[28]. We propose that in addition to the cell of origin, the specific translocation mutations, and the levels of oncoprotein expression likely determine the lineage identity of the leukemia.

*KMT2A*r leukemias have long been associated with an unusual degree of lineage plasticity, and occasionally transition from a B-ALL-to-AML identity in response to therapeutic pressures[29]. The ability of the oncoproteins to transition from a lineage-instructive role to a non-instructive role depending on the expression levels likely underlies the lineage plasticity associated with *KMT2A*r leukemias. An attractive model is that the lineage-switching *KMT2A*r B-ALL leukemia sample we profiled arose from the myeloid lineage, but because the KMT2A-AFF1 fusion protein was expressed at high enough levels, it directed the leukemia to assume a B-ALL identity. Then, in response to ALL-directed treatment, the oncoprotein levels were reduced and it transitioned to a non-instructive role, allowing the leukemia to revert to its original myeloid state. In support of this model, we find the non-canonical AML program includes numerous genes enriched in neutrophils. A common Lymphoid-Myeloid-Primed Progenitor gives rise to both neutrophil-monocyte progenitors as well as the B-cell lineage[30,31]. This lineage relationship could lower the epigenetic barriers for the KMT2A-AFF1 oncoproteins to direct a neutrophil progenitor to assume a B-ALL identity.

*KMT2A*r leukemias bearing *KMT2A-AF10* translocations have also been shown to undergo B-ALL-to-AML lineage-switching events to evade targeted therapies[10], but in contrast to KMT2A-AFF1, the KMT2A-AF10 oncoproteins preferentially activate myeloid lineage programs. We speculate that in these cases, the *KMT2A-AF10* mutation may have arisen in the lymphoid lineage, and in response to treatment, the expression of the KMT2A-AF10 fusion protein increases to instruct a transition to an AML identity. We propose that alterations in the oncoprotein levels result in transitions between an instructive and a non-instructive role of the oncoprotein in determining the lineage identity of the leukemia and depending on the context this can result in B-ALL-to-AML lineage-switching events.

Pharmacological agents, such as Revumenib, that target the chromatin scaffold protein Menin, are an effective treatment option for a subset of *KMT2A*r leukemia patients[5]. Phase I clinical trials demonstrated that at least two distinct mechanisms can cause patients to become resistant to Revumenib[14]. One mechanism is clearly genetic, and results from mutations in the binding-pocket of Menin that disrupt Revumenib binding[14]. The second mechanism is less clear. In these patients the leukemias rapidly relapse despite a persistent downregulation of canonical oncoprotein target genes[14]. We propose this second resistance mechanism is likely epigenetic, and in two representative samples from these patients the same non-canonical AML program that we identified in *KMT2A*r lineage-switching samples is activated in response to Revumenib. Identifying the contexts in which targeted therapies are most likely to have a positive outcome is a critical next step for personalized medicine, and profiling the oncoprotein binding sites and chromatin landscapes of patient samples could inform the management of *KMT2A*r leukemias in the clinic.

Acknowledgements

We thank Christine Codomo, Doris Xu, Terri Bryson and Jorja Henikoff for technical assistance, Charles Mullighan for providing critical materials, and Matthew Fitzgibbon for bioinformatics support and the Fred Hutch Genomics Shared Resource 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. performed the experiments; D.H.J., M.D. and D.J.O. performed data analysis. D.K., J.S.Y., and 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.).

Methods

Patients

Patient samples were obtained from member COG institutions and the St. Jude's Children's 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's Children's Hospital and the Texas Children's Hospital. None of the patients received compensation for their inclusion in this study.

Sample Information

The primary CD34+ HSPC control sample obtained from the FHCRC/UW Hematopoietic Disease Repository, 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[16] 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 and 2 pediatric ALLs and 5 infant ALLs (including one that transitioned to an AML in response to treatment) from the Texas Children's Hospital. 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[32,33].

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, KMT2 -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). 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.

AutoCUT&RUN DNA sequencing and data processing

AutoCUT&RUN sample processing was performed by the Fred Hutch Cancer Center Genomics Shared Resources Facility according to a previously published protocol available on the protocols.io website (https://doi.org/10.17504/protocols.io.ufeetje). 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[34] 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 refered 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 per-

formed. A total of 1000 samples ($N = 1000$) were drawn from this posterior distribution and the mean (μ_i) and standard deviation (σ_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 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 GSEXXXXXX.

References

1 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

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

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 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

10 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

11 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

12 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

13 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

14 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

15 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

16 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

17 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

18 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

19 Dolence, J. J., Gwin, K. A., Shapiro, M. B. & Medina, K. L. Flt3 signaling regulates the proliferation, survival, and maintenance of multipotent hematopoietic progenitors that generate B cell precursors. *Exp Hematol* **42**, 380-393 e383 (2014). https://doi.org:10.1016/j.exphem.2014.01.001

20 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

21 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

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 Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep* **26**, 1627-1640 e1627 (2019). https://doi.org:10.1016/j.celrep.2019.01.041

24 Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366** (2019). https://doi.org:10.1126/science.aax9198

25 Niemann, C. U. *et al.* Neutrophil elastase depends on serglycin proteoglycan for localization in granules. *Blood* **109**, 4478-4486 (2007). https://doi.org:10.1182/blood-2006-02-001719

26 Gil-Krzewska, A. *et al.* An actin cytoskeletal barrier inhibits lytic granule release from natural killer cells in patients with Chediak-Higashi syndrome. *J Allergy Clin Immunol* **142**, 914-927 e916 (2018). https://doi.org:10.1016/j.jaci.2017.10.040

27 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

28 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

29 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

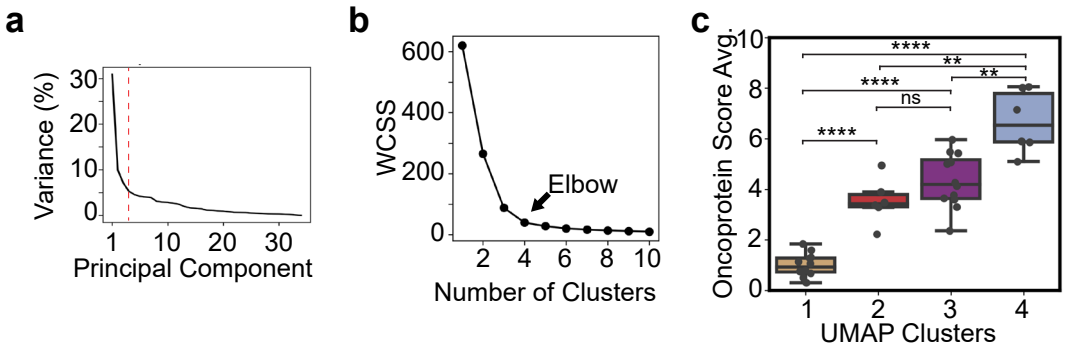
30 Adolfsson, J. *et al.* Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295-306 (2005). https://doi.org:10.1016/j.cell.2005.02.013

31 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

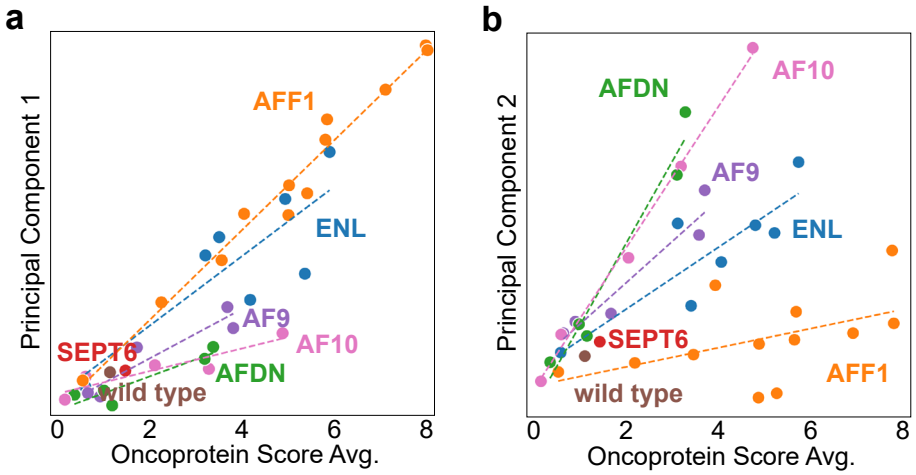
32 Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* **47**, 330-337 (2015). https://doi.org:10.1038/ng.3230

33 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

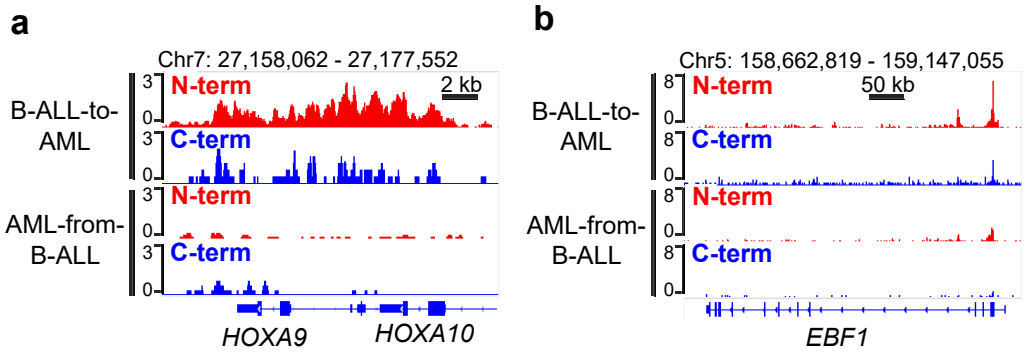
34 Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019). https://doi.org:10.1186/s13072-019-0287-4



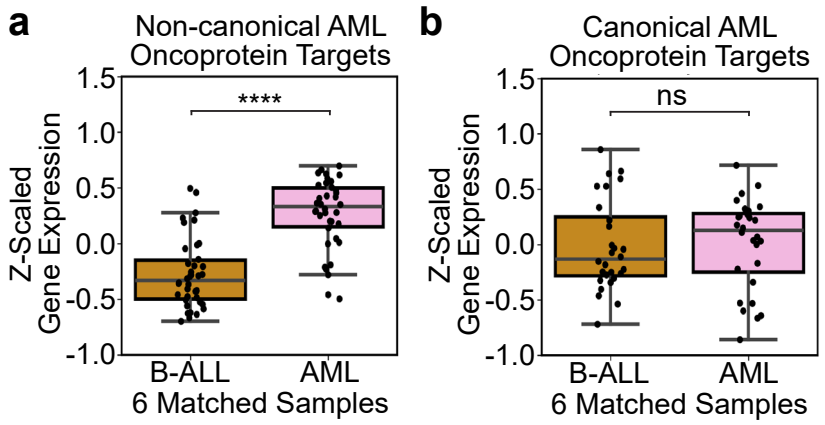
Extended Data Figure 1: Dimensionality reduction and clustering of the KMT2A-oncoprotein scores. a, Rank order plot of the variance in oncoprotein scores detected by principle component analysis. The first 4 components (red line) capture >50% of the total variance. **b,** Elbow plot of the K-means Within-Cluster Sum of Squares (WCSS) scores suggests the *KMT2A*r leukemias form 4 clusters in UMAP space. **c,** Boxplot comparing the average oncoprotein scores of the 4 clusters of *KMT2A*r leukemia samples.



Extended Data Figure 2: Principal Components 1 and 2 capture difference in the magnitude of oncoprotein fusion partners. a, Scatterplot showing the average oncoprotein scores for AFF1 and ENL oncoproteins are more strongly associated with Principal Component 1. **b,** The average oncoprotein scores for AFDN and AF10 oncoproteins are more strongly associated with Principal Component 2.



Extended Data Figure 3: The oncoprotein levels at *HOXA9* and *EBF1* are reduced during lineage switching. a, Genome Browser tracks showing the oncoprotein levels are elevated over the *HOXA9* locus in the B-ALL sample prior to lineage switching. **b,** Same as (a), but showing the *EBF1* locus.



Extended Data Figure 4: The non-canonical AML oncoprotein targets are consistently upregulated during lineage switching. **a,** Boxplot comparing the relative expression of the 42 non-canonical AML-oncoprotein target genes in a cohort of 6 samples that underwent a BALL-to-AML lineage-switching event. **b,** Same as (a) but comparing the expression of the 30 canonical AML-oncoprotein target genes.