Characterizing variation in GBAF/cBAF activity in acute myeloid leukemia through analysis of the epigenome and transcriptome

Maria Stephenson

Abstract

BRG1/BRM-Associated Factor (BAF) complex dysregulation is heavily implicated in cancer (1; 2). A prime example is synovial sarcoma, which is driven by changes in the relative activity of BAF complex isoforms, due to the degradation of cBAF (3). There is evidence that the relative activities of cBAF and GBAF differ across synovial sarcoma tumors, with one group appearing to have higher relative cBAF and lower GBAF activity, while the other has higher relative GBAF and lower cBAF. Recently BRD9 degraders have shown promise in inhibiting the growth of acute myeloid leukemia (AML) cell lines, however sensitivity to the degraders remains variable. As BRD9 is a subunit protein of GBAF, a potential explanation for this variation in sensitivity is that, like in synovial sarcoma, the relative activities of cBAF and GBAF vary across tumors, with some exhibiting relatively higher cBAF and lower GBAF activity, resulting in their resistance to BRD9 degradation. To investigate if such variation in BAF complex activity is apparent in AML tumors, we analyzed the epigenomes and transcriptomes of AML patient samples to determine if they reflected variation in relative cBAF and GBAF activity. We found that the epigenomes of some patient samples were indicative of higher relative GBAF activity, as a greater proportion of their H3K27ac peaks were promoter-proximal, indicating a higher ratio of GBAF to cBAF binding sites. However, further analyses supported such samples having high cBAF activity as well, as their peaks had similar enrichment in cBAF-associated motifs as those from samples indicative of higher relative cBAF activity, and they also had higher RNA expression of cBAF subunit ARID1A. Additionally RNA expression of BRD9, SMARCB1 and ARID1A was found to be positively correlated across AML patient samples and cell lines. Together, this evidence supports the hypothesis that cBAF and GBAF activity varies in AML, however, unlike in synovial sarcoma, their activity does not appear to be negatively correlated, and may instead be positively correlated.

Introduction

The term epigenome refers to chemical modifications to DNA and histone proteins, excluding alterations to the DNA sequence (4). Histone proteins can be modified in several different ways, including acetylation, methylation, ubiquitination, and phosphorylation (4). Depending on the type of modification, and the amino acid and histone protein being modified, these epigenetic “marks” can have positive or negative effects on chromatin accessibility and transcription. For example, tri-methylation of lysine-9 in histone H3 (H3K27me3) is associated with transcriptional inhibition at heterochromatic regions (5; 6), and is thus reffered to as a “repressive” mark. In contrast, acetylation of lysine-27 in histone H3 is viewed as an “activating” mark, as it is observed at the promoters of actively transcribed genes and active enhancers (7).

Histone modifications can also affect the binding of BRG1/BRM-Associated Factor (BAF) complexes (8). These complexes are a family of ATP-dependent chromatin remodellers, that play an essential role in regulating gene expression and differentiation (9). They use energy produced from ATP hydrolysis to reposition and evict nucleosomes, increasing DNA accessibility (10). Several BAF complex isoforms are expressed in human cells. Three of the major isoforms include canonical BAF (cBAF), polybromo-associated BAF (PBAF), and GLTSCR1/like-containing BAF (GBAF), which is also reffered to as non-canonical BAF (ncBAF). These isoforms are each composed of shared and distinct subunits. For example, cBAF, PBAF and GBAF all share core subunits SMARCC1 (BAF155) and SMARCD1 (BAF60A), however subunit ARID1A is unique to cBAF,

BRD9 is unique to GBAF, and SMARCB1 is in both cBAF and PBAF but not GBAF (8). The isoforms also tend to localize to different genomic regions. For example, while cBAF and GBAF both localize to regions marked by H3K27ac, GBAF tends to localize to promoters while cBAF tends to localize to distal enhancers (11). Additionally, GBAF tends to localize to CTCF sequence motifs, while cBAF binding sites are enriched for FOS/JUN, AP-1, SPDEF and ETS motifs, amongst others (12).

Given their key role in regulating gene expression, it is no surprise that BAF complex dysregulation is highly pervasive in cancer. Approximately 20% of human tumors contain genomic alterations in BAF complex subunits (1; 2). A prominent example of oncogenesis driven by BAF complex dysregulation is synovial sarcoma. Synovial sarcoma is a rare soft tissue sarcoma that results from a t(X;18) chromosomal translocation leading to the fusion of genes SS18 and SSX (13; 14; 15; 16; 17; 3). SS18 is a subunit of GBAF and cBAF, and its substitution with the fusion protein SS18-SSX results in degradation of cBAF, increasing the relative abundance of GBAF and PBAF in the cell (3). In contrast, while mutations in BAF complex subunits are infrequent in acute myeloid leukemia (AML), the expression of certain subunits has been found to be dysregulated (18; 19; 20). For example, GBAF subunit BRD9 is frequently upregulated in AML (19), and cBAF subunits SMARCB1 and ARID1A are downregulated (18; 20). This suggests that AML, like synovial sarcoma, could be driven by changes in the relative activity of BAF complex isoforms.

Supportive of this hypothesis, BRD9 degraders, which have demonstrated efficacy in treating synovial sarcoma cells (21), have also shown to be effective at inhibiting the growth of AML cell lines (22). Sensitivity to the degraders, however, is variable. For example, QA-68, a promising new BRD9 degrader, achieved an IC50 in the range of 1-10nM in AML cell lines MV4;11 and SKM-1, and an IC50 in the range of 10–100 nM in Kasumi-1-luc+ cells, however, was ineffective at inhibiting the growth of both HEL and KO52 cells (22). While the degrader proved ineffective at degrading BRD9 in the HEL cell line, likely resulting in its ineffectiveness at inhibiting HEL cell growth, the degrader was able to effectively degrade BRD9 in KO52 cells. A possible explanation for KO52's tolerance to BRD9 degradation, is that these cells have relatively lower GBAF activity and higher cBAF activity compared to the sensitive cell lines, meaning they are less dependent on GBAF activity to maintain their proliferative state. The hypothesis that some leukemias have relatively lower GBAF and higher cBAF activity aligns with the lab's current research into synovial sarcoma, where we have identified two distinct groups of synovial sarcoma tumors, one that appears to have relatively higher GBAF and lower cBAF activity, and the other relatively higher cBAF and lower GBAF activity.

The purpose of this study was to examine the epigenomes and transcriptomes of AML patient tumors and see if they reflect variation in cBAF and GBAF activity. At the epigenomic level, this was done by identifying groups of patients with distinct patterns of H3K27ac (which marks regions of cBAF and GBAF binding (11)) and comparing group regions for differences in motif enrichment and proportion of promoter proximal/distal regions. At the transcriptomic level, the expression of cBAF and GBAF specific subunits was compared between groups. These analyses were also applied to the samples individually, in order to characterize within group variance.

Materials and Methods

IHEC ChIP-Seq

ChIP-Seq data for H3K27ac was obtained from the International Human Epigenome Consortium (IHEC) BLUEPRINT and CEEHRC projects. Only samples from patients with AML were used. The data was processed using the IHEC Integrative Analysis ChIP Pipeline (available at: github.com/IHEC/integrative_analysis_chip/).

IHEC RNA-Seq

RNA-Seq data was obtained from the International Human Epigenome Consortium (IHEC) BLUEPRINT and CEEHRC projects. Only samples from patients whose ChIP-Seq data was also used

were considered. Samples were processed using the IHEC grape-nf pipeline (available at: https://github.com/IHEC/grape-nf).

TCGA RNA-Seq

Gene expression counts were obtained from the TCGA-LAML project using the R package TCGAbiolinks (23; 24; 25).

DepMap RNA-Seq data visualization

Cell line gene expression from the Expression Public 22Q4 dataset was visualized using the DepMap Portal Data Explorer (available at: https://depmap.org/portal/interactive/).

Hierarchical clustering

Hierarchical clustering was performed using the R package pvclust (26). Spearman distance was used as the distance metric. The clustering matrix was made by binning the human genome (hg38) into 500bp bins and calculating the percentage overlap between MACS2 peaks in each sample, with the genome-wide bins. Various clustering approaches were performed using different subsets of the matrix. Approaches included filtering the matrix for bins marked by a peak in at least 5%, 10%, 20% or 30% of samples, but not all, and filtering for the top 1% most variable bins. An additional clustering approach used a matrix created by calculating percentage overlap between peaks in each sample, with promoter and AML enhancer regions. Promoter regions were obtained from the Ensembl gene annotation hg38v104 (27), and enhancers were obtained from EnhancerAtlas 2.0 (28), using UCSC liftOver (29) to convert to hg38.

Motif enrichment analysis

Motif enrichment analysis was performed using HOMER (v4.11) (30) findMotifsGenome.pl (with genome = hg38 and size = 200).

Distance from transcription start site eCDF

Distances from regions of interest to the nearest transcription start site were calculated using HOMER (v4.11) (30) annotatePeaks.pl (with genome = hg38). For eCDFs of group regions, eCDFs were calculated using the function from ggplot2 (31). When analysing samples separately, the eCDF was only calculated at numbers in the range $\{10^{0.1}, 10^{0.2}, 10^{0.3}, \ldots, 10^{5.9}, 10^{6}\}$ to ease computation when calculating means and standard deviations across groups of samples.

Results

Hierarchical clustering of AML patient samples

AML patient samples were clustered based on genome-wide H3K27ac using hierarchical clustering. Several approaches were employed to cluster the samples, including use of different linkage methods, and different genomic regions of interest. As for regions of interest, we attempted

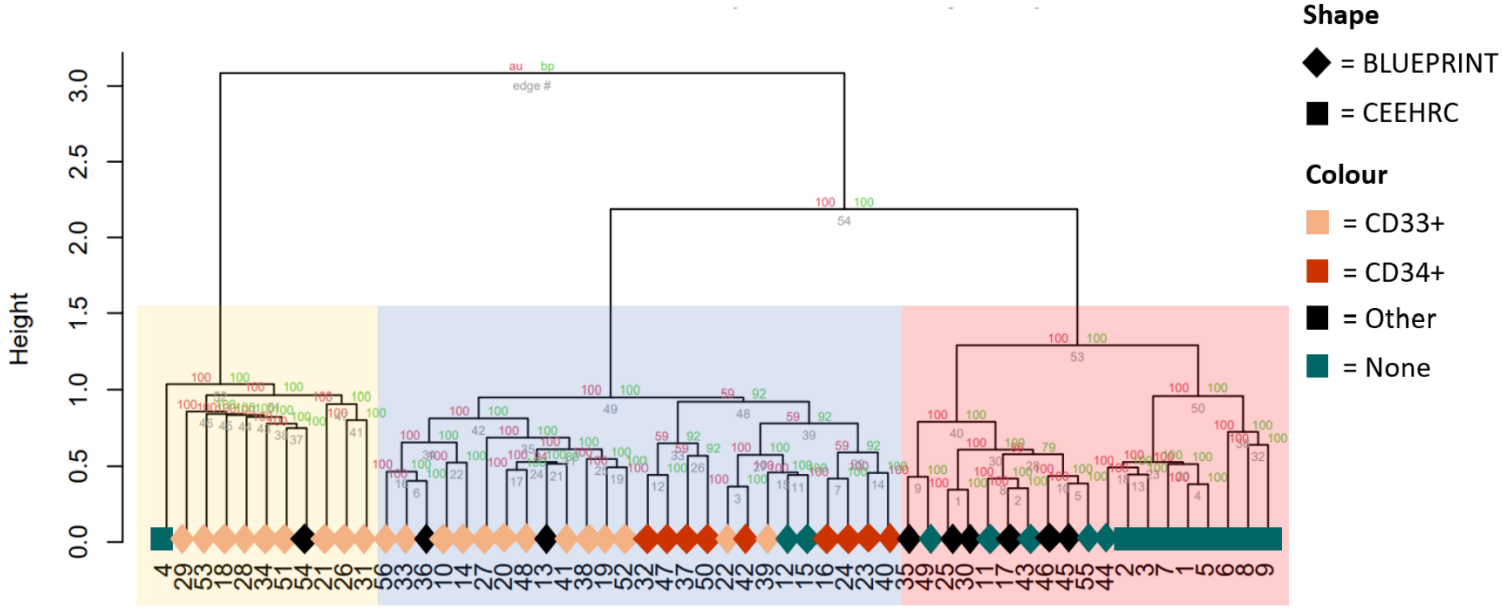


Figure 1: Hierarchical clustering of AML patient samples. Dendrogram displaying hierarchical clustering of AML patient samples based on H3K27ac. Genome was binned into 500bp bins and the percentage overlap between MACS2 peaks in each sample and the bins was calculated. Only bins marked by a peak in at least 3 samples (N=1,025,961) were considered. Spearman distance was used as the distance metric, ward.D as the linkage, and the number of bootstrap replications was 30.

clustering considering only regions marked by a peak in a certain number of samples, only the top 1% most variable regions, and only promoter and AML blast enhancer regions (Supplemental Fig. 1-3). Interestingly, distinct clusters were only apparent when linkages ward.D and ward.D2 were used. However, this may be expected as ward linkage has been demonstrated to outperform other linkages when clustering large sequencing datasets (32). The clustering approach chosen for downstream analyses used ward.D linkage and only considered genomic regions that were marked by a peak in at least three samples (and not all). This technique was chosen because the resulting clustering had low bootstrap p-values and was also largely conserved across other techniques when ward.D or ward.D2 linkage was used (Supplemental Fig. 1-3).

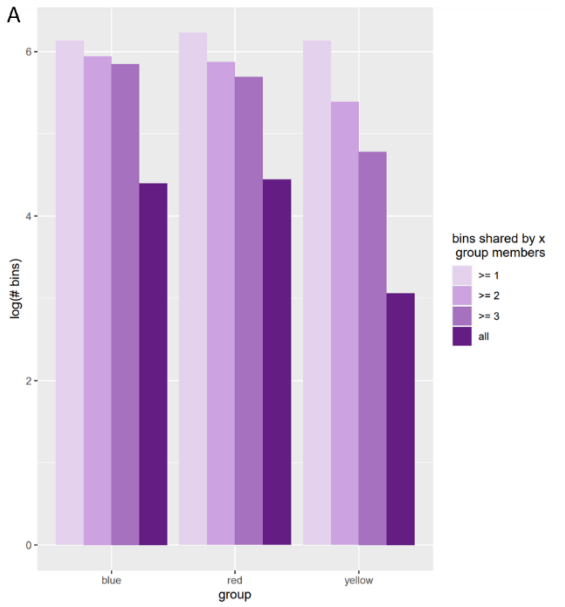
The clustering analysis revealed three distinct patient groups (Fig. 1) (referred to simply as yellow, blue and red). Samples tended to cluster with samples that had been sorted for similar markers (Fig. 1, Supplemental Table 1). The yellow group mostly contained samples sorted for CD33+. The blue was composed of two subgroups, one consisting mostly of samples sorted for CD34+, and the other mostly of samples sorted for CD33+. The red group also consisted of two subgroups: one that mostly contained samples that had not been sorted for cell markers, and the other that consisted of both unsorted samples, and samples that had been sorted for a variety of cell markers. Cell types have been shown to cluster based on H3K27ac peaks (33), so it is not surprising that the groups would be enriched for samples sorted with certain markers. As CD34+ is a hematopoietic stem cell marker (34), the CD33+ samples in the blue group may represent leukemias with more stem-cell dominant cellular hierarchies.

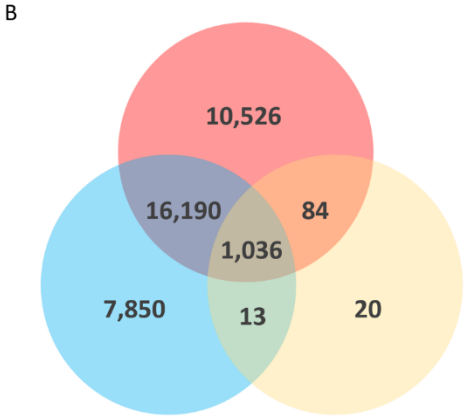
Analysis of H3K27ac peaks shared within groups

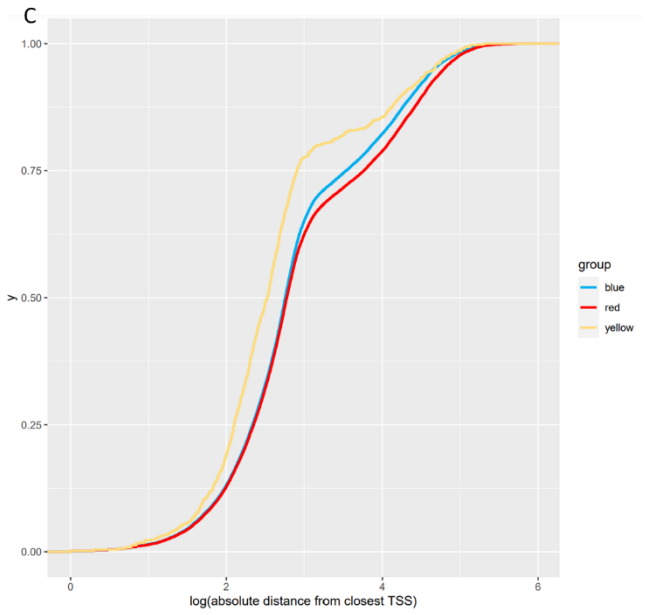
In order to characterize differences in the H3K27ac landscape across groups, we characterized regions marked by a peak amongst all group members (referred to as "shared" regions) for each group. A much smaller percentage of yellow regions were shared, compared to blue and red, indicating that the yellow group has a greater diversity of regulatory states (Fig. 2a). While the top enriched motifs for each group's shared regions greatly overlapped (Supplemental Fig. 4), grouping motifs by transcription factor family revealed that the yellow group contained less enriched Zf, bHLH, bZIP and IRF motifs (Fig. 2d) – the latter two being associated with cBAF binding (12). Additionally, a greater proportion of yellow shared regions were promoter-proximal than blue and red shared regions (Fig. 2c). As GBAF tends to localize to promoters (11), this, in addition to the motif analysis, indicates that the ratio of GBAF activity to cBAF activity may be higher in the yellow group than in the red and blue groups.

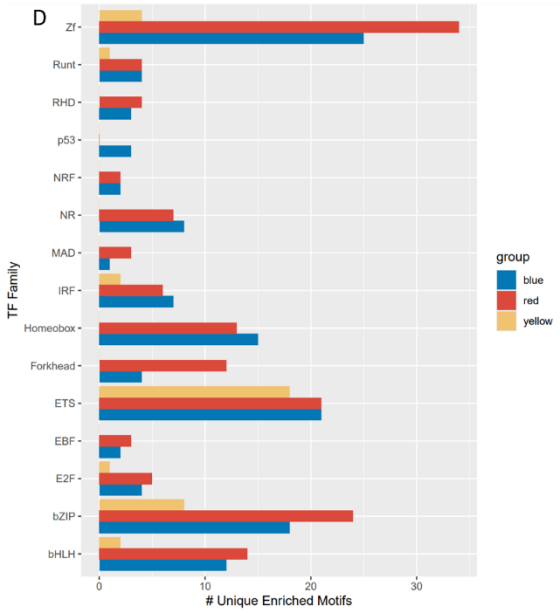
We then investigated how much overlap occurred between shared regions of the different groups. Yellow shared regions were found to be a subset of both red and blue shared regions (Fig. 2b), demonstrating that all groups contained shared regions reflective of GBAF activity, and that the blue and red just contained more shared cBAF-associated regions.

Additionally, RNA expression of GBAF-associated subunit BRD9, and cBAF-associated subunits ARID1A and SMARCB1 was compared across the groups (Fig. 2e). Interestingly, CEEHRC samples exhibited higher expression of both SMARCB1 and ARID1A than BLUEPRINT samples (Fig. 2e). Such differences could be due to differences in sample handling, library preparation or sequencing.


A
log(# bins)
group
blue
red
yellow
bins shared by x group members
>= 1
>= 2
>= 3
all


B
10,526
16,190
84
1,036
7,850
13
20


C
y
log(absolute distance from closest TSS)
group
blue
red
yellow


D
TF Family
Zf
Runt
RHD
p53
NRF
NR
MAD
IRF
Homeobox
Forkhead
ETS
EBF
E2F
bZIP
bHLH
Unique Enriched Motifs
group
blue
red
yellow

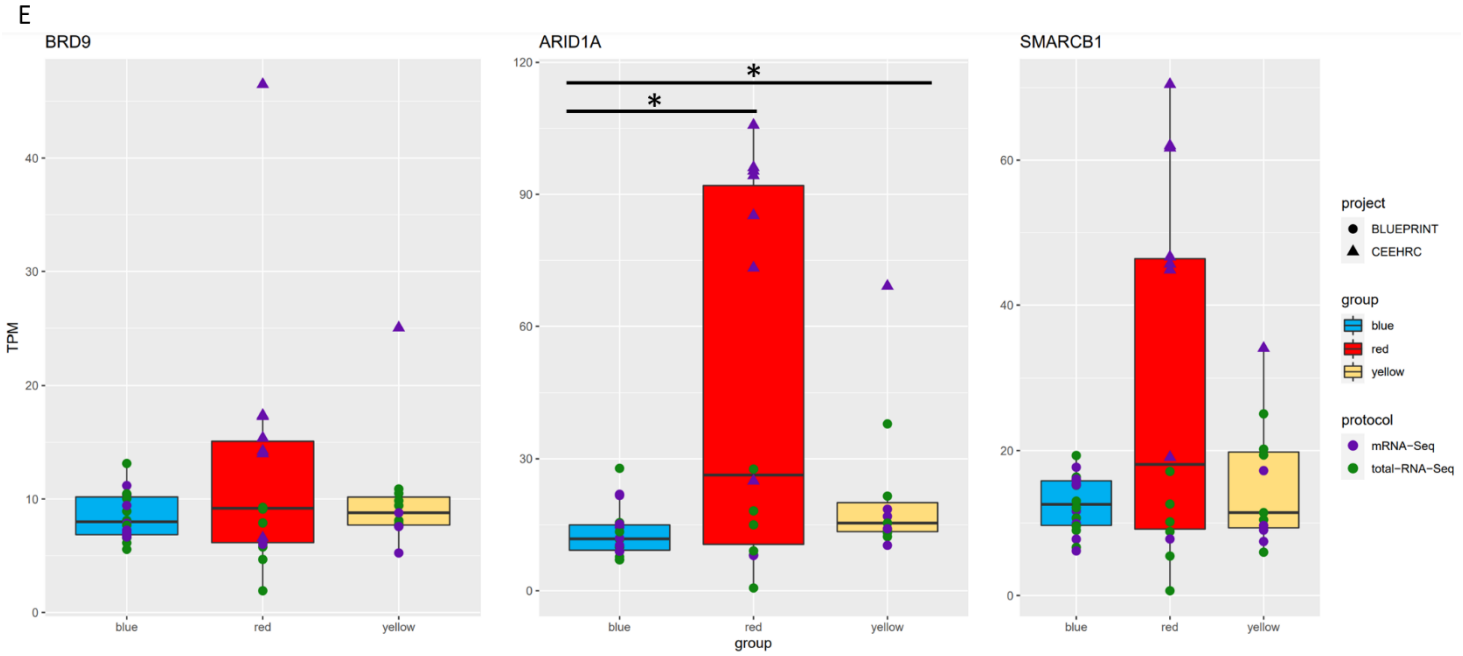

E
BRD9
ARID1A
SMARCB1
TPM
group
blue
red
yellow
project
BLUEPRINT
CEEHRC
protocol
mRNA-Seq
total-RNA-Seq

Figure 2: Characterization of regions marked by H3K27ac shared within groups. a) Box plot depicting number of 500bp bins marked by a peak in different proportions of group members. **b)** Venn diagram depicting overlap in shared regions between red, blue and yellow groups. **c)** eCDFs of the common log of shared regions' absolute distance to its nearest transcription start site (TSS) for each group (yellow distribution is significantly different than red and blue, $p < 2.16e\text{-}16$, KS test). **d)** Bar plot depicting the count of unique enriched motifs for the top 15 most variable (in terms of enriched motif count) transcription factor families across groups. **e)** Box plots depicting RNA expression (TPM) for BRD9, ARID1A and SMARCB1 across groups (* = $p < 0.05$, Wilcoxon test).

ARID1A was found to be significantly differentially expressed between the blue group and both the red and yellow groups (Fig. 2e). However, when CEEHRC samples (which were found in both the red and yellow groups but not the blue) were excluded, the difference in expression was no longer significant, indicating that such difference may have just been due to batch effects.

Analysis of H3K27ac peaks in individual samples

In the blue, red and yellow groups, the number of shared regions is a small fraction of the number of regions marked by a peak in any one of its members (Fig. 2a), and thus is likely not reflective of its members' regulatory states. The yellow group was especially diverse, and if this diversity was enriched in regions of cBAF binding, this would have resulted in the shared regions being reflective of higher relative GBAF activity. If yellow group members truly clustered together due to having higher relative GBAF activity compared to red and blue members, we would expect that, for any yellow sample, a greater proportion of the clustering matrix bins marked by its peaks would be reflective of GBAF binding, or not indicative of cBAF binding, than for any red or blue sample. For this reason, we decided to repeat the same motif enrichment and TSS eCDF analyses on each sample individually, considering only clustering matrix bins marked by at least one of its peaks.

In this analysis, on average, yellow samples did not have a greater proportion of promoter-proximal regions than red or blue samples (Fig. 3a).

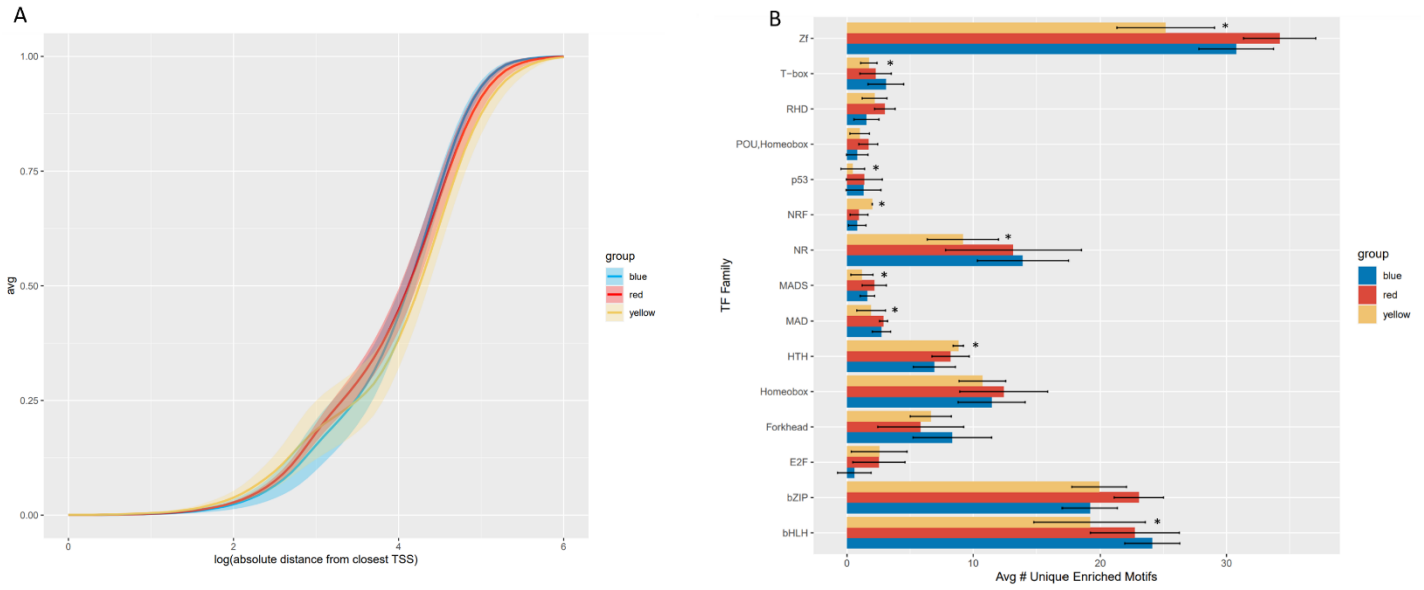


Figure 3: Characterization of regions marked by H3K27ac peaks in individual samples. a) Average of each group member's eCDF of the common log of marked regions' absolute distance to its nearest transcription start site (TSS), for blue, red and yellow groups. Ribbons represent standard deviation. **b)** Bar plot depicting the average count of unique enriched motifs for the top 15 most variable (in terms of mean enriched motif count) transcription families across blue, red, and yellow groups (* = $p < 0.05$, t-test, yellow vs blue and red groups).

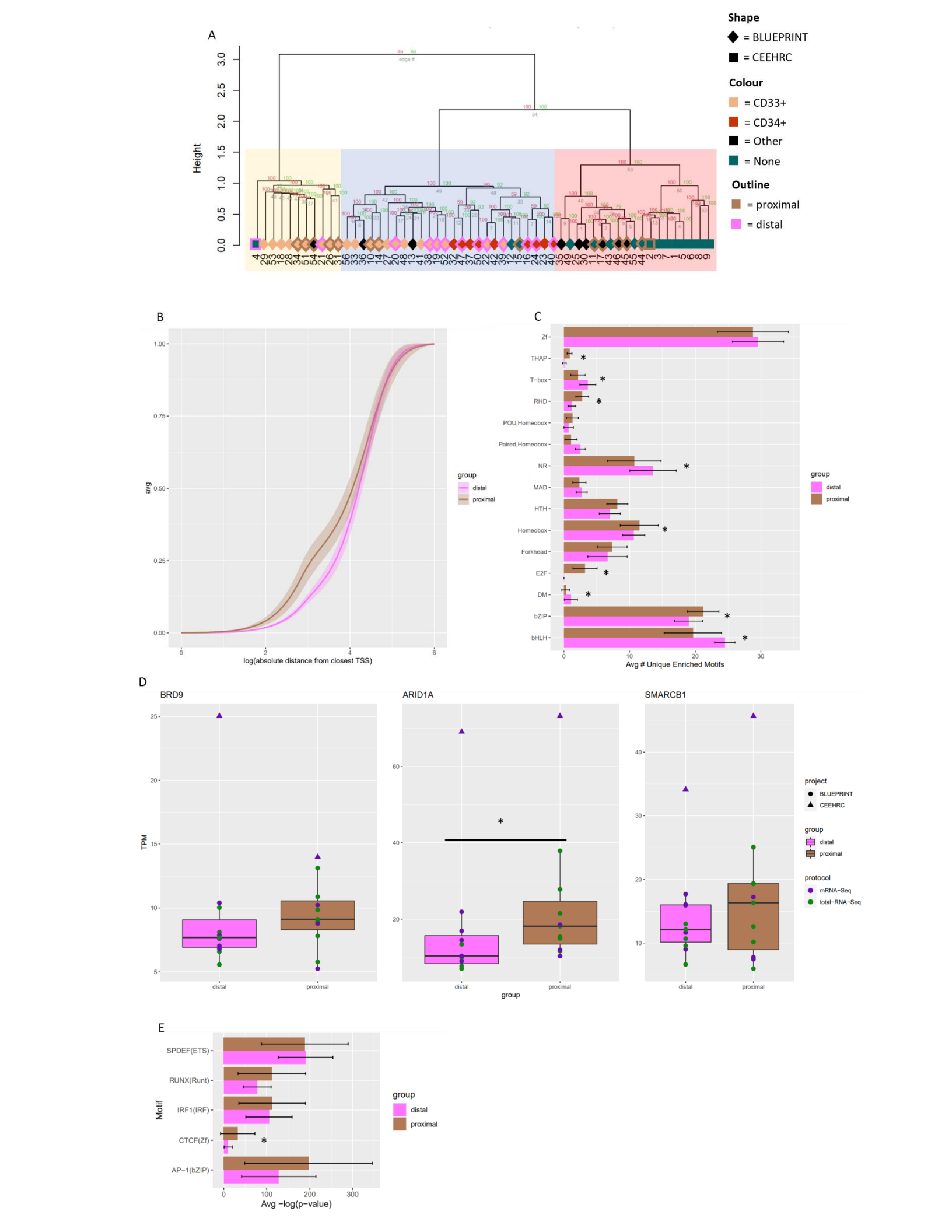

A
Height
Shape
= BLUEPRINT
= CEEHRC
Colour
= CD33+
= CD34+
= Other
= None
Outline
= proximal
= distal
B
avg
log(absolute distance from closest TSS)
group
distal
proximal
C
Zf
THAP
T−box
RHD
POU,Homeobox
Paired,Homeobox
NR
MAD
HTH
Homeobox
Forkhead
E2F
DM
bZIP
bHLH
Avg # Unique Enriched Motifs
group
distal
proximal
D
BRD9
ARID1A
SMARCB1
TPM
distal
proximal
group
project
BLUEPRINT
CEEHRC
group
distal
proximal
protocol
mRNA−Seq
total−RNA−Seq
E
Motif
SPDEF(ETS)
RUNX(Runt)
IRF1(IRF)
CTCF(Zf)
AP−1(bZIP)
Avg −log(p−value)
group
distal
proximal

Figure 4: Characterization of regions marked by H3K27ac in predicted GBAF-low and GBAF-high groups. a) Dendrogram displaying hierarchical clustering of AML patient samples based on H3K27ac, clustered in the same way as in Fig. 1. Proximal and distal groups are identified by colour of shape outline. **b)** Average of each group member's eCDF of the common log of marked regions' absolute distance to the nearest transcription start site (TSS), for proximal and distal groups. Ribbons represent standard deviation. **c)** Bar plot depicting the average count of unique enriched motifs for the top 15 most variable (in terms of mean enriched motif count) transcription families across proximal and distal groups. (* = $p < 0.05$, t-test) **d)** Box plots depicting RNA expression of BRD9, ARID1A and SMARCB1 in proximal and distal groups. (* = $p < 0.05$, Wilcoxon test) **e)** Bar plot depicting the average -log(p-value) for cBAF and GBAF associated motifs (* = $p < 0.05$, Wilcoxon test)

Additionally, while the average yellow sample had a lower count of enriched Zf, T-box, p53, NR, NRF, MAD, MADS, HTH and bHLH motifs, it did not have significantly less enriched bZIP motifs compared to both blue and red samples (Fig. 3b). This indicates that yellow group members on average do not have a higher ratio of GBAF to cBAF activity compared to blue and red, and that there is greater diversity amongst predicted cBAF binding sites in the yellow group than in the blue and red.

Analysis of predicted GBAF-high and GBAF-low groups

The average TSS eCDFs for the yellow and blue groups had high standard deviation around 100-1000bp from the nearest TSS (Fig. 3a), indicating that the proportion of promoter-proximal regions varies greatly amongst members in these groups, and thus that relative GBAF activity could vary amongst group members as well. To identify individual samples within groups that could be reflective of higher or lower relative GBAF activity, we characterized samples whose TSS eCDF at 1000bp (a cut-off for considering regions promoter-proximal or distal (35)) were in the top or bottom quartiles (these groups were referred to as "proximal" and "distal" respectively). Samples belonging to the proximal group were found in each of the blue, red and yellow groups, though they were more concentrated in the yellow and red (Fig. 4a). In contrast, distal group samples were strongly concentrated in the blue group, with all but two distal samples belonging to the blue cluster (Fig. 4a).

Despite the proximal group being predicted to be GBAF high and cBAF low (as a greater proportion of its H3K27ac peaks are promoter-proximal), the proximal and distal groups had a significantly higher count of enriched bZIP motifs, which are associated with cBAF binding (12) (Fig. 4b). Additionally, in terms of specific cBAF and GBAF associated motifs, GBAF-associated motif CTCF was found to be significantly more enriched in the proximal group, and the mean enrichment p-value for cBAF-associated motifs AP-1 and RUNX, were lower in the proximal group as well, though not significantly (Fig. 4e). Furthermore, the proximal group showed greater mean expression of BRD9, SMARCB1 and ARID1A than the distal group (Fig. 4d). While the differences in expression of BRD9 and SMARCB1 were not significant, ARID1A was significantly more expressed in the proximal group. Together, this indicates that though locations of H3K27ac peaks in the proximal group are indicative of higher relative GBAF activity, in contrast to the hypothesis, the motif enrichment and RNA expression analyses do not support this group having lower cBAF activity relative to the distal group, which was predicted to be GBAF low and cBAF high. Rather, these analyses indicate that the proximal group has the same or even higher cBAF activity.

Analysis of RNA expression of cBAF and GBAF associated subunits across AML patient samples and cell lines

RNA expression of BRD9, ARID1A and SMARCB1 in the TCGA-LAML dataset and AML cell lines was also analyzed. Interestingly, expression of ARID1A and SMARCB1 had strong positive correlation with the expression of BRD9 in the TCGA-LAML dataset (Fig. 5a,b), and had moderate positive correlation across AML cell lines (Fig. 5c,d).

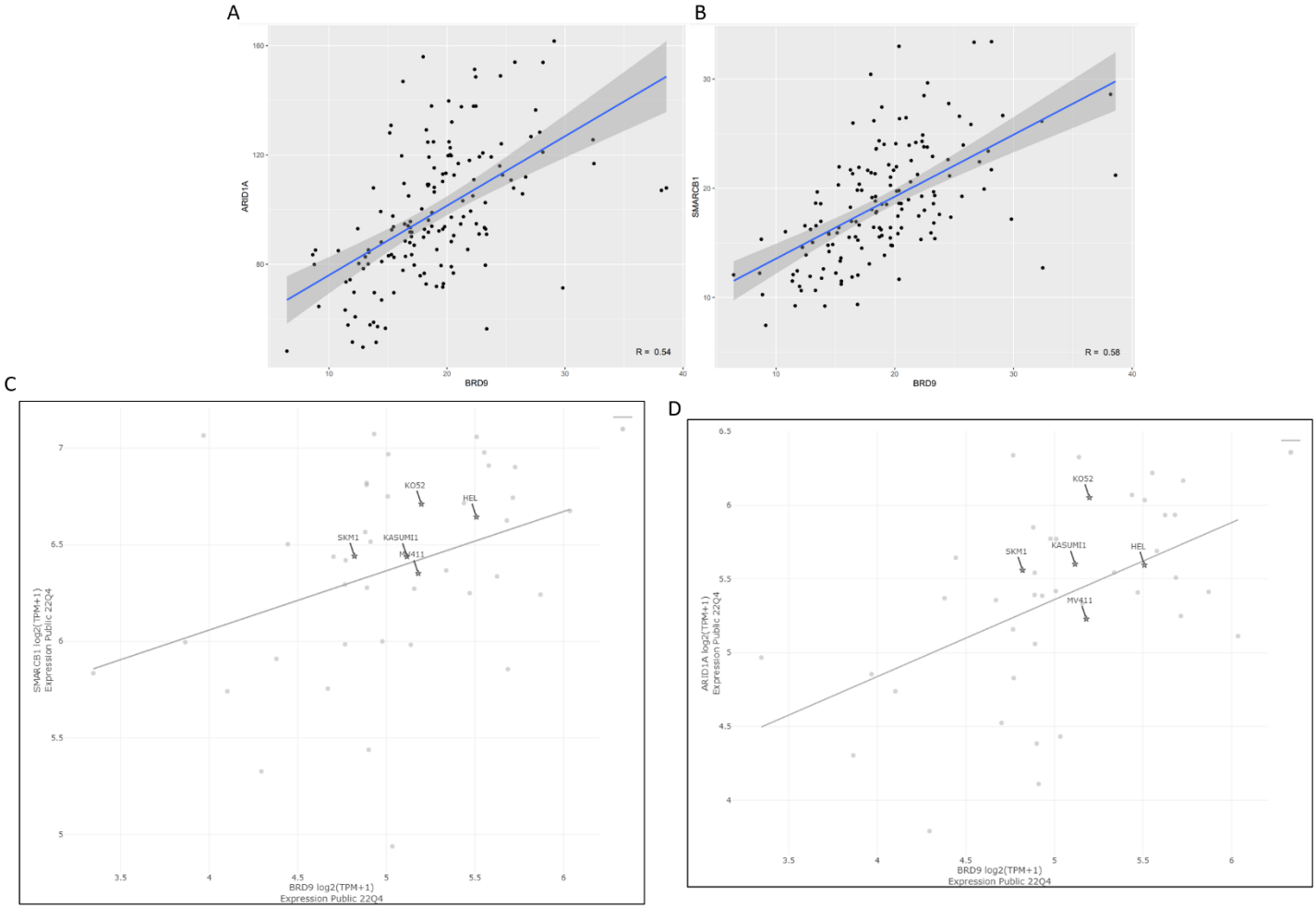


Figure 5: Analysis of cBAF and GBAF subunit RNA expression across AML patient samples and cell lines. a) Scatterplot depicting ARID1A and BRD9 RNA expression across TCGA-LAML samples ($R = 0.54$, $p < 0.05$) **b)** Scatterplot depicting SMARCB1 and BRD9 RNA expression across TCGA-LAML samples ($R = 0.58$, $p < 0.05$) **c)** Scatterplot depicting SMARCB1 and BRD9 RNA expression across AML cell lines ($R = 0.316$, $p < 0.05$) **d)** Scatterplot depicting ARID1A and BRD9 RNA expression across AML cell lines ($R = 0.466$, $p < 0.05$)

This further supports the idea that cBAF and GBAF activity may be positively correlated in leukemia.

Additionally, BRD9-resistant AML cell line KO52, which was originally hypothesized to have higher cBAF activity and lower GBAF activity, was observed to have higher expression of SMARCB1 and ARID1A compared to sensitive cell lines MV4;11, Kasumi-1 and SKM-1, in line with the hypothesis. However, it also exhibited higher or equal expression of BRD9 (Fig. 5c,d), indicating that it may have higher GBAF activity as well.

Discussion

In this study, we characterized the epigenomes and transcriptomes of AML patient samples, to determine if they were reflective of differences in the relative activity of BAF complex isoforms cBAF and GBAF. Clustering of the samples based on H3K27ac revealed 3 distinct patient groups (red, blue, and yellow). Samples tended to cluster with those that had been sorted for the same cell markers, and this was reflected across the groups, with the yellow group containing mostly CD33+ samples, the blue group containing roughly half CD34+ and half CD33+ samples, and the red group containing mostly samples that had not been sorted or had been sorted for various cell markers (Fig. 1, Supplemental Table 1). When characterizing shared regions (marked by a H3K27ac peak in all group members) for each group, we found that the yellow

shared regions were reflective of having a higher ratio of GBAF activity to cBAF activity compared to blue and red (Fig. 2). Analyses on individual samples revealed that, while the epigenomes of some samples, based on the proportion of promoter-proximal H3K27ac peaks, were indicative of higher relative GBAF activity these samples were not confined to one group. Instead, samples predicted to be "GBAF high", referred to as the proximal group, were found largely in the red and yellow groups. Samples predicted to be "GBAF low", referred to as the distal group, were found largely in the blue group (Fig. 4a). As the blue group was predicted to contain leukemias with more stem-cell dominant cellular hierarchies, this indicates that lower GBAF activity might be associated with the stem cell phenotype.

Interestingly, samples predicted to be GBAF high had similar enrichment of cBAF-associated motifs, as those predicted to be GBAF low and cBAF high (Fig 4e). Additionally, GBAF high samples not only exhibited higher mean expression of BRD9 compared to GBAF low samples, but also higher mean expression of cBAF subunits SMARCB1 and ARID1A, with ARID1A being significantly over expressed (Fig. 4d). Together this indicates that leukemias with higher GBAF activity may also exhibit higher or equivalent cBAF activity compared to leukemias with lower GBAF activity. This was further supported by the fact that RNA expression of ARID1A, SMARCB1 and BRD9 is strongly positively correlated in TCGA-LAML patient samples (Fig. 5a,b), and moderately positively correlated across AML cell lines (Fig. 5c,d).

Regarding sensitivity to BRD9 degraders, the idea of GBAF and cBAF activity being positively correlated in leukemia could still explain variable sensitivity to BRD9 degradation. BRD9 degrader-resistant KO52 was observed to have higher expression of cBAF subunits ARID1A and SMARCB1 than sensitive cell lines MV4;11, Kasumi-1 and SKM-1, and similar or higher expression of GBAF subunit BRD9, indicating that it might have relatively higher GBAF and cBAF activity (Fig. 5c,d). Its resistance could be explained by the fact that even if BRD9 is degraded, the cell line would still have high cBAF activity, and as there is significant overlap between cBAF and GBAF binding sites (12), this would perhaps allow KO52 cells to tolerate BRD9 degradation.

There are several limitations to this study. Firstly, activity of GBAF and cBAF was inferred through comparison of differences in the amount of predicted cBAF and GBAF binding sites, however a cell having a greater number of cBAF/GBAF binding sites does not necessarily mean it has a greater amount of cBAF/GBAF binding activity. Additionally, RNA expression does not necessarily correlate with protein abundance, so even though ARID1A was found to have significantly greater expression in the proximal group compared to the distal, this does not mean this difference exists at the protein level, and even if it did, does not mean the proximal samples necessarily have more cBAF, as protein expression of the other subunits may not differ. Similarly, though the RNA expression of BRD9, SMARCB1 and ARID1A was strongly positively correlated, that does not mean that this correlation exists at the protein level either.

In summary this study supports the hypothesis that cBAF and GBAF activity vary across AML tumors. Samples whose epigenome and transcriptome were reflective of higher GBAF activity, were also reflective of higher or equivalent cBAF activity, indicating that GBAF and cBAF activity is likely not negatively correlated in AML, and could potentially be positively correlated. Future directions could include quantifying abundances of cBAF and GBAF-specific components at the protein level in samples or cell lines predicted to be GBAF high and GBAF low to ascertain if such samples contain higher/lower amounts of GBAF and cBAF at the protein level. Additionally, as based on KO52, higher cBAF activity and GBAF activity could correlate with resistance to BRD9 degradation, further characterization of the proximal group (indicative of higher GBAF and potentially higher cBAF activity) could potentially aid in the creation of a classifier to predict BRD9 resistance. If certain enhancers are active in proximal group but not in the distal group (indicative of lower GBAF and potentially lower cBAF activity), measuring eRNA expression at such enhancers could allow for

classifying samples as GBAF and cBAF high or low, and thus BRD9 resistant or tolerant. And as proximal and distal group members mostly cluster apart based on H3K27ac (Fig. 4a), it is likely that their enhancer landscapes are different.

References

1. *SWI/SNF complex in cancer.* **Lu, C., & Allis, C. D.** s.l. : Nat Genet, 2017, Vols. 49(2), 178–179. https://doi.org/10.1038/ng.3779.

2. *Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy.* **Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R.** s.l. : Nature Genet, 2013, Vols. 45(6), 592–601. https://doi.org/10.1038/ng.2628.

3. *A Role for SMARCB1 in Synovial Sarcomagenesis Reveals That SS18-SSX Induces Canonical BAF Destruction. .* **Li, J., Mulvihill, T. S., Li, L., Barrott, J. J., Nelson, M. L., Wagner, L., Lock, I. C., Pozner, A., Lambert, S. L., Ozenberger, B. B., Ward, M. B., Grossmann, A. H., Liu, T., Banito, A., Cairns, B. R., & Jones, K. B.** s.l. : Cancer discovery, 2021, Vols. 11(10), 2620–2637. https://doi.org/10.1158/2159-8290.CD-20-1219.

4. *The mammalian epigenome.* **Bernstein BE, Meissner A, Lander ES.** s.l. : Cell, 2007, Vols. 128(4), 669–681. https://doi.org/10.1016/j.cell.2007.01.033.

5. **Saksouk, N., Simboeck, E. & Déjardin, J.** *Constitutive heterochromatin formation and transcription in mammals.* s.l. : Epigenetics & Chromati, 2015. 8, 3.https://doi.org/10.1186/1756-8935-8-3.

6. **Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D.** *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* s.l. : Genes Dev, 2002. 16(22):2893–905. 10.1101/gad.1035902.

7. *Modification of enhancer chromatin: what, how, and why?* **Calo, E., & Wysocka, J.** s.l. : Mol Cell, 2013, Vols. 49(5), 825–837. https://doi.org/10.1016/j.molcel.2013.01.038.

8. *Unwinding chromatin at the right places: how BAF is targeted to specific genomic locations during development.* **Ho, P. J., Lloyd, S. M., & Bao, X.** s.l. : Development, 2019, Vol. 146(19). https://doi.org/10.1242/dev.178780.

9. *The BAF complex in development and disease.* **Alfert, A., Moreno, N. & Kerl, K.** s.l. : Epigenetics & Chromatin, 2019, Vols. 12, 19. https://doi.org/10.1186/s13072-019-0264-y.

10. *Mechanism(s) of SWI/SNF-induced nucleosome mobilization.* **Liu, N., Balliano, A., & Hayes, J. J.** s.l. : Chembiochem, 2011, Vols. 12(2), 196–204. https://doi.org/10.1002/cbic.201000455.

11. *A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells.* **Gatchalian, J., Malik, S., Ho, J., Lee, D. S., Kelso, T. W. R., Shokhirev, M. N., Dixon, J. R., & Hargreaves, D. C.** s.l. : Nature commun, 2018, Vols. 9(1), 5139. https://doi.org/10.1038/s41467-018-07528-9.

12. *A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation.* **Michel, B. C., D'Avino, A. R., Cassel, S. H., Mashtalir, N., McKenzie, Z. M., McBride, M. J., Valencia, A. M., Zhou, Q., Bocker, M., Soares, L. M. M., Pan, J., Remillard, D. I., Lareau, C. A., Zullow, H. J., Fortoul, N., Gray, N. S., Bradner, J. E., Chan,.** s.l. : Nat Cell Biol, 2018, Vols. 20(12), 1410–1420. https://doi.org/10.1038/s41556-018-0221-1.

13. *Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcom.* **Crew, A. J., Clark, J., Fisher, C., Gill, S., Grimer, R., Chand, A., Shipley, J., Gusterson, B. A., & Cooper, C. S.** https://doi.org/10.1002/j.1460-2075.1995.tb07228.x, 1995, The EMBO journal, pp. 14(10), 2333–2340.

14. *Nuclear localization of SYT, SSX and the synovial sarcoma-associated SYT-SSX fusion proteins.* **dos Santos, N. R., de Bruijn, D. R.,**

Balemans, M., Janssen, B., Gärtner, F., Lopes, J. M., de Leeuw, B., & Geurts van Kessel, A. s.l. : Human molecular genetics, 1997, Vols. 6(9), 1549–1558. https://doi.org/10.1093/hmg/6.9.1549.

15. *Functional domains of the SYT and SYT-SSX synovial sarcoma translocation proteins and co-localization with the SNF protein BRM in the nucleus.* **Thaete, C., Brett, D., Monaghan, P., Whitehouse, S., Rennie, G., Rayner, E., Cooper, C. S., & Goodwin, G.** s.l. : Human molecular genetics, 1999, Vols. 8(4), 585–591. https://doi.org/10.1093/hmg/8.4.585.

16. *SYT associates with human SNF/SWI complexes and the C-terminal region of its fusion partner SSX1 targets histones.* **Kato, H., Tjernberg, A., Zhang, W., Krutchinsky, A. N., An, W., Takeuchi, T., Ohtsuki, Y., Sugano, S., de Bruijn, D. R., Chait, B. T., & Roeder, R. G.** s.l. : The Journal of biological chemistry, 2002, Vols. 277(7), 5498–5505. https://doi.org/10.1074/jbc.M108702200.

17. *SS18 together with animal-specific factors defines human BAF-type SWI/SNF complexes.* **Middeljans, E., Wan, X., Jansen, P. W., Sharma, V., Stunnenberg, H. G., & Logie, C.** s.l. : PloS one, 2012, Vol. 7(3). https://doi.org/10.1371/journal.pone.0033834.

18. *Therapeutic Opportunities of Targeting Canonical and Noncanonical PcG/TrxG Functions in Acute Myeloid Leukemia.* **Zeisig, B. B., & So, C.** s.l. : Annual review of genomics and human genetics, 2021, Vols. 22, 103–125. https://doi.org/10.1146/annurev-genom-111120-102443.

19. *BRD9 binds cell type-specific chromatin regions regulating leukemic cell survival via STAT5 inhibition.* **Del Gaudio, N., Di Costanzo, A., Liu, N. Q., Conte, L., Migliaccio, A., Vermeulen, M., Martens, J., Stunnenberg, H. G., Nebbioso, A., & Altucci, L.** s.l. : Cell death & disease, 2019, Vols. 10(5), 338. https://doi.org/10.1038/s41419-019-1570-9.

20. *ARID1A has prognostic value in acute myeloid leukemia and promotes cell proliferation via TGF-β1/SMAD3 signaling.* **Ren, T., Wang, J., Tang, W., Chen, D., Wang, S., Zhang, X., & Yang, D.** s.l. : Clinical and experimental medicine, 2022. https://doi.org/10.1007/s10238-022-00863-8.

21. *Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma.* **Brien, G. L., Remillard, D., Shi, J., Hemming, M. L., Chabon, J., Wynne, K., Dillon, E. T., Cagney, G., Van Mierlo, G., Baltissen, M. P., Vermeulen, M., Qi, J., Fröhling, S., Gray, N. S., Bradner, J. E., Vakoc, C. R., & Armstrong, S. A.** s.l. : eLife, 2018. https://doi.org/10.7554/eLife.41305.

22. *BRD9 degraders as chemosensitizers in acute leukemia and multiple myeloma.* **Weisberg E, Chowdhury B, Meng C, Case AE, Ni W, Garg S, Sattler M, Azab AK, Sun J, Muz B, Sanchez D, Toure A, Stone RM, Galinsky I, Winer E, Gleim S, Gkountela S, Kedves A, Harrington E, Abrams T, Zoller T, Vaupel A, Manley P, Faller M, Chung B, Chen X, B.** s.l. : Blood Cancer J, 2022, Vol. 12(7):110. doi: 10.1038/s41408-022-00704-7.

23. *TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data.* **Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot T, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H.** s.l. : Nucleic Acids Research, 2015. doi: 10.1093/nar/gkv1507, http://doi.org/10.1093/nar/gkv1507.

24. *TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages.* **Silva, C T, Colaprico, Antonio, Olsen, Catharina, D'Angelo, Fulvio, Bontempi, Gianluca, Ceccarelli, Michele, Noushmehr, Houtan.** s.l. : F1000Research, 2016, Vol. 5.

25. *New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx.* **Mounir, Mohamed, Lucchetta, Marta, Silva, C T, Olsen, Catharina, Bontempi, Gianluca, Chen, Xi, Noushmehr, Houtan, Colaprico, Antonio, Papaleo, Elena.** s.l. : PLoS computational biology, 2019, Vol. 15(3).

26. *Pvclust: an R package for assessing the uncertainty in hierarchical clustering.* **Suzuki R, Shimodaira H.** s.l. : Bioinformatics, 2006, Vols. 22(12):1540-2. 10.1093/bioinformatics/btl117.

27. *Ensembl 2022.* **Fiona Cunningham, James E. Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, Irina M. Armean , Olanrewaju Austine-Orimoloye, Andrey G. Azov, If Barnes, Ruth Bennett, Andrew Berry, Jyothish Bhai, Alexandra Bignell, Konstantinos Billis , Sanjay B.** s.l. : Nucleic Acids Res, 2022, Vols. 50(1):D988-D995. doi:10.1093/nar/gkab1049.

28. *EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species.* **Tianshun Gao, Jiang Qian.** s.l. : Nucleic Acids Research, 2020, Vols. D1(D58-D64). https://doi.org/10.1093/nar/gkz980.

29. *The UCSC Genome Browser Database: update 2006.* **Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower.** s.l. : Nucleic Acids Res, 2006, Vols. 34(Database issue):D590-8. doi: 10.1093/nar/gkj144.

30. *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* **Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.** s.l. : Mol Cell, 2010, Vols. 28;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004.

31. *ggplot2: Elegant Graphics for Data Analysis.* **H, Wickham.** s.l. : Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4.

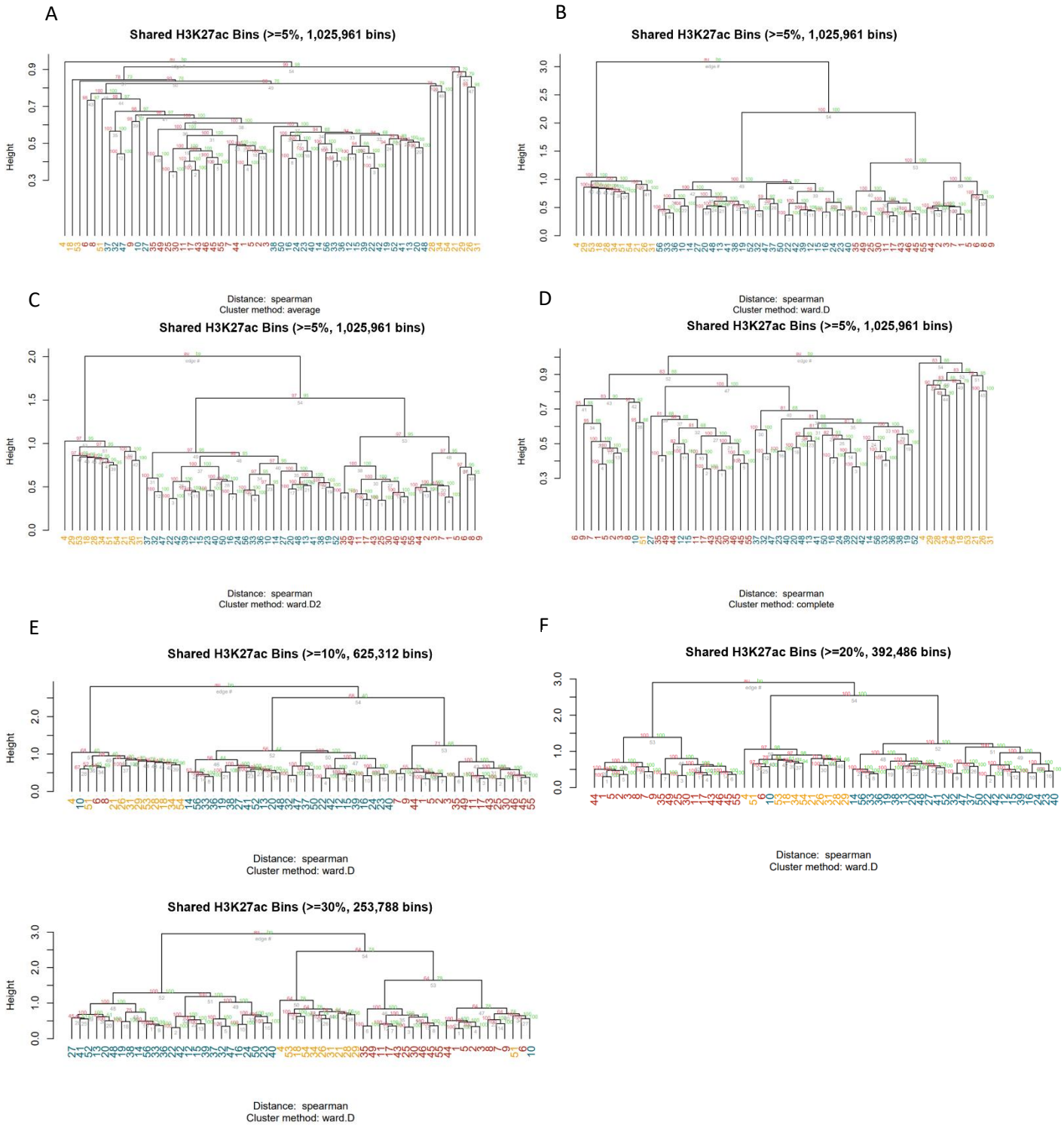
32. *An Analysis of Different Distance-Linkage Methods for Clustering Gene Expression Data and Observing Pleiotropy: Empirical Study.* **Choudhury J, Ashraf FB.** s.l. : JMIR Bioinform Biotech, 2022, Vol. 3(1). doi: 10.2196/30890.

33. *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* **Menno P. Creyghton, Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, Michael A. Lodato, Garrett M. Frampton, Phillip A. Sharp, Laurie A. Boyer, Richard A. Young, and Rudolf Jaenisch.** s.l. : PNAS, 2010. https://doi.org/10.1073/pnas.1016071107.

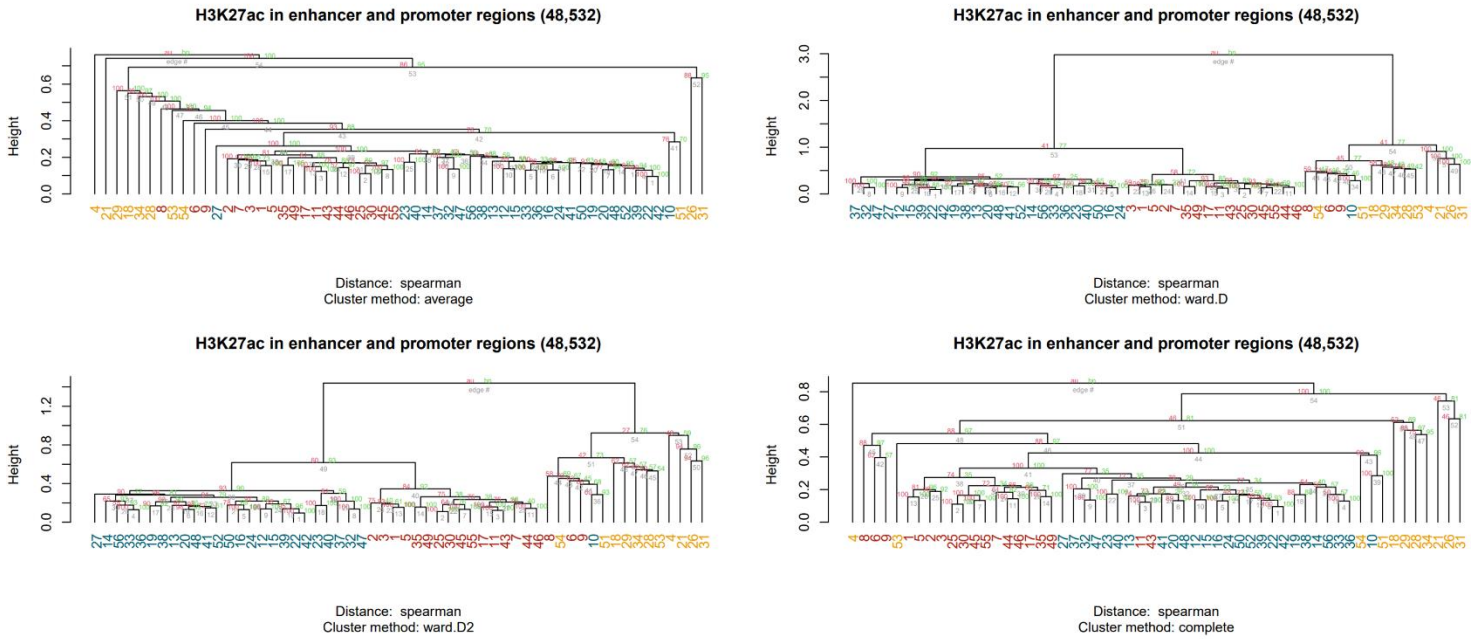
34. *Not just a marker: CD34 on human hematopoietic stem/progenitor cells dominates vascular selectin binding along with CD44.* **Dina B. AbuSamra, Fajr A. Aleisa, Asma S. Al-Amoodi, Heba M. Jalal Ahmed, Chee Jia Chin, Ayman F. Abuelela, Ptissam Bergam, Rachid Sougrat, Jasmeen S. Merzaban.** s.l. : Blood Adv, 2017. https://doi.org/10.1182/bloodadvances.2017004317.

35. *Genome-wide characterization of mammalian promoters with distal enhancer functions.* **Dao, L., Galindo-Albarrán, A., Castro-Mondragon, J.** s.l. : Nat Genet, 2017, Vols. 49, 1073–1081. https://doi.org/10.1038/ng.3884.

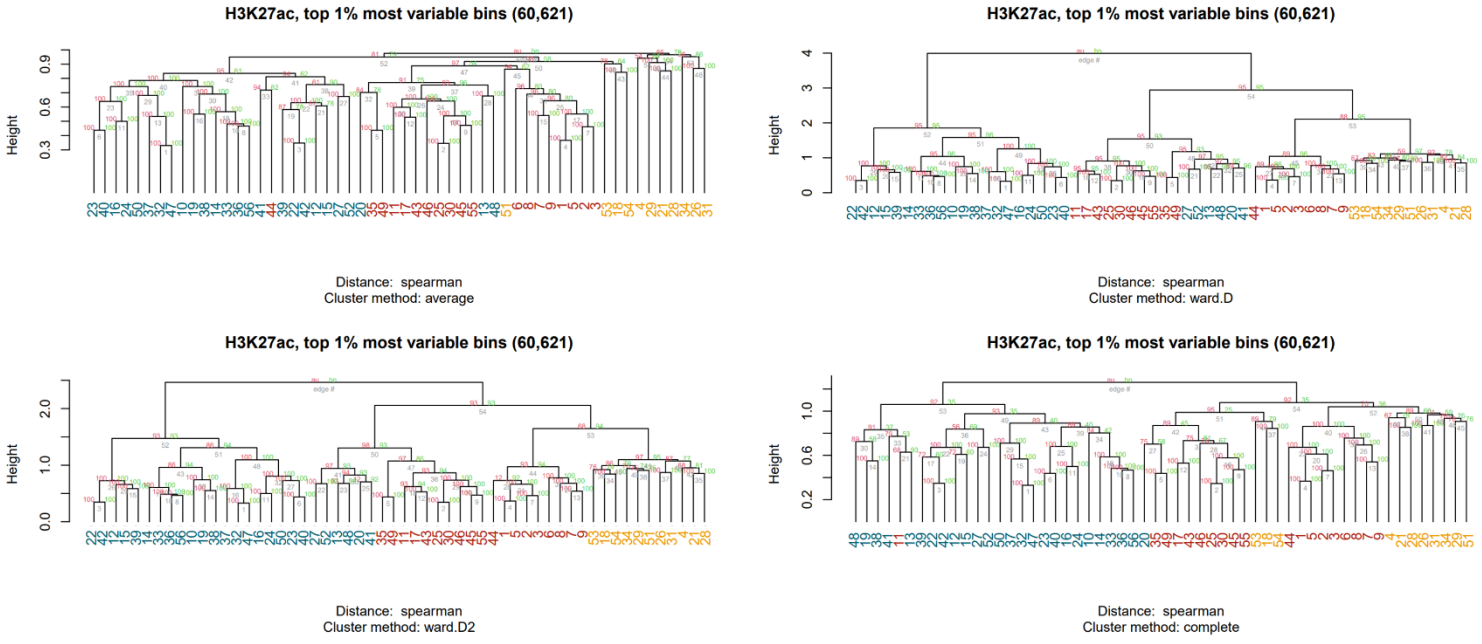
Supplementary Figures



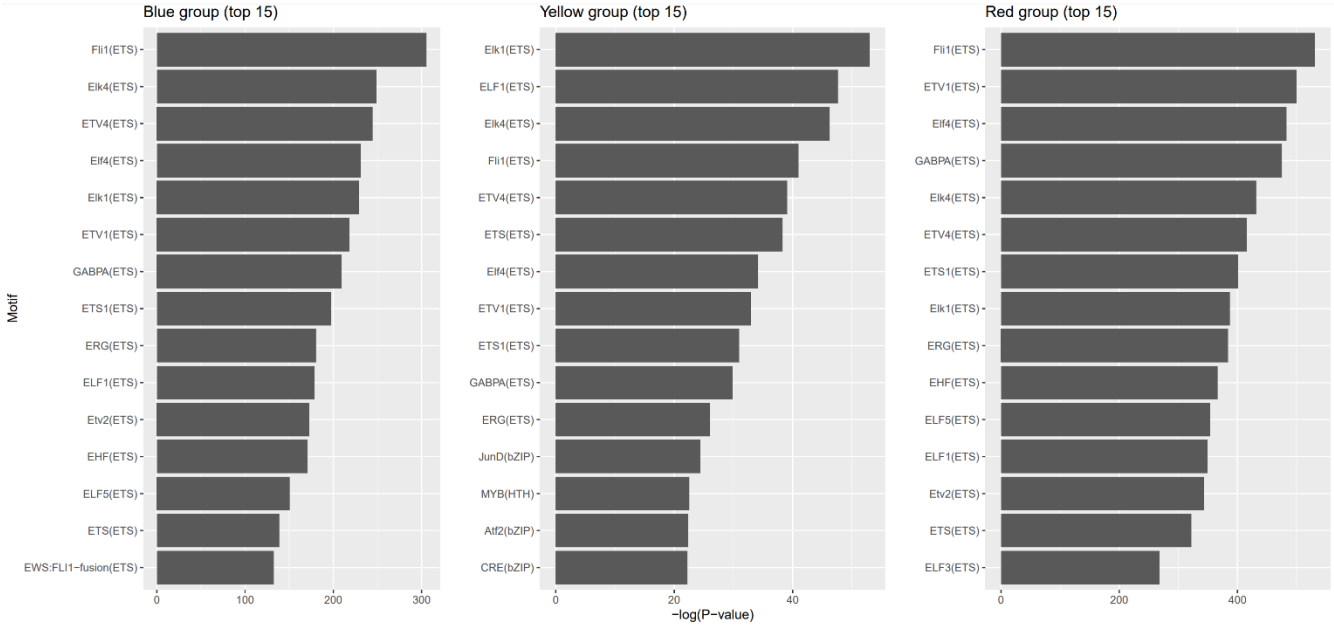
Supplementary Figure 1: Hierarchical clustering of AML patient samples using regions marked by peaks in multiple samples. Dendrograms displaying hierarchical clustering of AML patient samples based on H3K27ac. Genome was binned into 500bp bins and the percentage overlap between MACS2 peaks in each sample and the bins was calculated. Spearman distance was used as the distance metric and the number of bootstrap replications was 30. Only bins marked by a peak in at least **a-d)** 3 **e)** 6 **f)** 11 **g)** 17 samples were used.



Supplementary Figure 2: Hierarchical clustering of AML patient samples using enhancer and promoter regions. Dendrogram displaying hierarchical clustering of AML patient samples based on H3K27ac. Percent overlap between MACS2 peaks in each sample and set of promoter and enhancer regions was calculated. Spearman distance was used as the distance metric and the number of bootstrap replications was 30.



Supplementary Figure 3: Hierarchical clustering of AML patient samples using most variable regions. Dendrogram displaying hierarchical clustering of AML patient samples based on H3K27ac. Percent overlap between MACS2 peaks in each sample and set of promoter and enhancer regions was calculated Genome was binned into 500bp bins and the percentage overlap between MACS2 peaks in each sample and the bins was calculated. Spearman distance was used as the distance metric and the number of bootstrap replications was 30. Only top 1% most variable bins were used.



Supplementary Figure 4: Top 15 enriched motifs for blue, red, and yellow shared regions. Box plot depicting the negative common log of the p-values for the top 15 enriched motifs in the shared regions for each of the blue, red and yellow groups.

Supplemental Tables

ID	EpiRR	Project	Cell Markers
1	IHECRE00003824.4	CEEHRC	
2	IHECRE00003829.4	CEEHRC	
3	IHECRE00003830.4	CEEHRC	
4	IHECRE00003833.4	CEEHRC	
5	IHECRE00003835.4	CEEHRC	
6	IHECRE00003843.4	CEEHRC	
7	IHECRE00003844.4	CEEHRC	
8	IHECRE00003846.4	CEEHRC	
9	IHECRE00003847.4	CEEHRC	
10	IHECRE00000014.2	BLUEPRINT	CD33+
11	IHECRE00000026.3	BLUEPRINT	
12	IHECRE00000031.3	BLUEPRINT	
13	IHECRE00000041.3	BLUEPRINT	
14	IHECRE00000053.3	BLUEPRINT	CD33+
15	IHECRE00000175.3	BLUEPRINT	
16	IHECRE00000249.2	BLUEPRINT	CD34+
17	IHECRE00000252.2	BLUEPRINT	gate 35%: CD34- CD33+ CD13+ CD14+ CD64+ CD117-
18	IHECRE00000260.2	BLUEPRINT	CD33+
19	IHECRE00000263.2	BLUEPRINT	CD33+
20	IHECRE00000267.2	BLUEPRINT	CD33+
21	IHECRE00000270.2	BLUEPRINT	CD33+
22	IHECRE00000277.2	BLUEPRINT	CD33+
23	IHECRE00000295.2	BLUEPRINT	CD34+
24	IHECRE00000296.2	BLUEPRINT	CD34+
25	IHECRE00000311.2	BLUEPRINT	92% blasts: CD45_dim CD33++ CD64+ CD56+ CD34- CD14-; 7% blasts: CD45_dim CD33++ CD64+ CD56+ CD34- CD14+
26	IHECRE00000315.2	BLUEPRINT	CD33+
27	IHECRE00000316.2	BLUEPRINT	CD33+
28	IHECRE00000321.2	BLUEPRINT	CD33+
29	IHECRE00000339.2	BLUEPRINT	CD33+
30	IHECRE00000343.2	BLUEPRINT	92% blasts: CD45_dim CD33++ CD64+ CD56+ CD34- CD14-; 7% blasts: CD45_dim CD33++ CD64+ CD56+ CD34- CD14+
31	IHECRE00000348.2	BLUEPRINT	CD33+
32	IHECRE00001238.1	BLUEPRINT	CD34
33	IHECRE00001239.1	BLUEPRINT	CD33
34	IHECRE00001244.1	BLUEPRINT	CD33+
35	IHECRE00001246.1	BLUEPRINT	CD33+
36	IHECRE00001275.1	BLUEPRINT	CD33+
37	IHECRE00001288.1	BLUEPRINT	CD34
38	IHECRE00001306.1	BLUEPRINT	CD33+
39	IHECRE00001326.1	BLUEPRINT	CD33+
40	IHECRE00001339.1	BLUEPRINT	CD34
41	IHECRE00001378.1	BLUEPRINT	CD33+
42	IHECRE00001410.1	BLUEPRINT	CD34
43	IHECRE00001413.1	BLUEPRINT	
44	IHECRE00001423.1	BLUEPRINT	
45	IHECRE00001436.1	BLUEPRINT	MLL-AF9 positive 62% blasts high BRE
46	IHECRE00001469.1	BLUEPRINT	MLL-AF9 positive 96% blasts normal BRE
47	IHECRE00001473.1	BLUEPRINT	CD34
48	IHECRE00001482.1	BLUEPRINT	CD33+
49	IHECRE00001501.1	BLUEPRINT	
50	IHECRE00001507.1	BLUEPRINT	CD34
51	IHECRE00001508.1	BLUEPRINT	CD33+
52	IHECRE00001518.1	BLUEPRINT	CD33+
53	IHECRE00001536.1	BLUEPRINT	CD33+
54	IHECRE00001542.1	BLUEPRINT	CD33+
55	IHECRE00001554.1	BLUEPRINT	
56	IHECRE00001557.1	BLUEPRINT	CD33+

Supplementary Table 1: Sample information. Table containing the EpiRR, project and cell markers used for sorting for each sample.