SUPPLEMENTARY TABLES:

Supplementary Table S1: Sample mutations
Supplementary Table S2: RNA-seq signature and per-cluster weights
Supplementary Table S3: Median survival in TCGA data by predicted cluster
Supplementary Table S4: Enhancer, mRNA, and SY-1425 EC50 in AML cell lines
Supplementary Table S5: GSEA of transcriptional response to SY-1425 in RARA-high cell lines in perturbation gene sets
Supplementary Table S6: GSEA of transcriptional response to SY-1425 in RARA-high cell lines in hallmark and pathways gene sets
Supplementary Table S7: GREAT analysis of genes near H3K27ac peaks up-regulated by SY-1425 in RARA-high cell lines in perturbation gene sets

Supplementary Figure Legends

Supplementary Figure S1. Identification and refinement of cancer-specific super-enhancers in human AML. A, H3K27Ac landscape at positive control locus MALAT1. Tracks show read density scale normalized ChIP-seq signal. B-C, Distribution of H3K27Ac signal across the 12,656 enhancers in a single AML patient sample (SU041 - left), and the 12,311 enhancers in a sample of healthy CD34+ HSPCs (right). A subset of the enhancers (red, 470 and 830 respectively) have exceptionally high H3K27Ac signal and are called super-enhancers by the ROSE algorithm. D, Heatmap illustrating Pearson correlation of replicate enhancer maps. All enhancers' coordinates were merged across samples, and then all enhancers (including super-enhancers) were compared pairwise with Pearson correlation. The rows and columns are clustered with Euclidean distance and complete-linkage hierarchical clustering. The final two samples are CD34+ bone marrow cells from two separate healthy donors collected at separate locations. E, H3K27Ac ChIP-seq read density profiles across the admixture experiment samples. 11 Admix samples are shown (from 100% AML to 0% AML in steps of 10% from top to bottom). The three loci are “Ubiquitous”: the area over the ANKRD44 gene, “PBMC-specific”: the area around the IL10RA gene, and “AML-specific”: the area around the CDK6 gene. F, The percentage of PBMC (red) or AML (blue) -specific SEs recovered as super-enhancers are shown as a function of the percentage of blasts in each admixed sample. G, The percentage of PBMC (red) or AML (blue) -specific SEs recovered in the top 1000 enhancers (including SEs) are shown as a function of the percentage of blasts in each admixed sample.

Supplementary Figure S2. Mapping of super-enhancers to genes. A, Schematic demonstrating the difficulty in assigning function to an enhancer in the absence of three-dimensional chromatin architecture data. B-D, Correlation of super-enhancer scores for an enhancer near the (B) NEDD9 (r²=0.75, p<2E^-16), (C) TMEM170B (r²=0.45, p<5E^-10), and (D) SMIM13 (r²=0.17, p<0.0014) genes with the promoter H3K27Ac signal of each gene across all AML patients. A p value cutoff of 1.8E^-6 was used to determine whether a given enhancer acts on a given promoter. This is equivalent to 0.1 multiplied by the number of tested enhancer-promoter pairs genome-wide. E, Bar plot showing the number of SE-gene links per SE.

Supplementary Figure S3. AML SE profiles show pronounced variation in enhancers related to myeloid differentiation and enable de novo stratification into six distinct epigenomic subgroups.
A, Scatterplot showing the Euclidean distance of each sample from an HSPC centroid derived from the most HSPC- or monocyte-associated SEs from Figure 1D (x-axis), as well as the distance of each sample from the centroid of the monocyte samples (y-axis). B, Scatterplot showing the predicted HSPC signature from Figure 1D (x-axis) against the first independent component from the ICA in Figure 1E (y-axis) across all patient samples. C, Heatmap showing the super-enhancer score (with a ceiling at 2) of the top 40 ICA2-associated SEs by R² value. The rows are clustered by Euclidean distance and complete-linkage hierarchical clustering, and the columns are ordered by the ICA2 loading for each sample (top row, blue to red). The bottom row shows the type of sample: HSPC, light blue; primary AML, gray; monocyte, purple. Example gene links for the SEs are displayed on the right. D, Coefficient matrix from the best (lowest reconstruction error) NMF. Cells show the non-negative weight put onto each factor for each sample. Samples are ordered by cluster label and rows (factors) are hierarchically clustered. E, Basis matrix from the best (lowest reconstruction error) NMF. Cells show the non-negative weight put into each factor for each SE. SEs (rows) and factors (columns) are ordered by hierarchical clustering with complete linkage. F, Pearson correlation heatmap showing positively (red) and negatively (blue) correlated samples by the top 200 most variable SEs with at least one strong patient score. Samples are ordered identically to the NMF-C distance matrix with the exception of the HSPC and monocyte samples. G, A scatter plot showing ICA1 vs ICA2 as in Figure 1E, colored by cluster (patient samples) or cell type (FACS-purified normal samples).

Supplementary Figure S4. Recapitulation of SE-defined clusters in AML data from the TCGA. A, Bar graph showing the frequency of each mutation within our AML cohort. B, Overall survival within our cohort of 62 AML patients. Solid line: survival, dashed lines: confidence interval. C, Heatmap showing the 1,710 genes used in the nearest shrunken centroid classifier. Expression is row scaled by mean and standard deviation. Genes (rows) are ordered by which cluster they correspond to. Samples (columns) are ordered by cluster. D, Heatmap showing the contribution of each gene to each centroid. White indicates no contribution. Blue indicates a negative contribution and red indicates a positive contribution. E, Clustering was predicted for all non-M3 TCGA patient samples by scoring them against RNA-seq derived cluster centroids. Survival in the TCGA cohort significantly associates with projected cluster membership with a cox proportional hazards p value < 0.014. F, The mutation status of each of the TCGA AML samples is shown, ordered and colored by their predicted cluster. A bold mutation name indicates a nominally significant association (p < 0.05) between the mutation and the clustering (Fisher’s exact). Similarly, bold and italic font indicates significance after multiple hypothesis testing correction (p < 0.001). Only the most recurrent mutations (n > 2) are shown here. rMLL = MLL rearrangement; Mito = Mitochondrial genes.

Supplementary Figure S5 An SE is present at the RARA locus in a subset of AML samples. A, Metatrack plots of the H3K27Ac landscape at the RARA locus in all AML clusters and healthy monocytes and HSPCs. B, Correlation of RARA enhancer score with RARα mRNA levels (by RNA-seq). Spearman rho ~ .47; p < 0.0008. C, Three-dimensional interactions in the local region surrounding the RARA gene locus from promoter capture Hi-C data. H3K27Ac ChIP-seq from one AML patient (SU070) and CTCF ChIP-seq from monocytes (Encode) are shown for reference. The RARA SE is highlighted by a blue box. Interactions of the RARA promoter with the RARA SE are shown in red. Only unique interactions with a CHiCAGO score > 5 are shown for clarity. Region represents chr17:38455078-38579467. D, Analysis of RARα mRNA from a
public data set (1) reveals a significantly higher level of RARα mRNA in a subset of MDS patients blast cells compared to non-malignant CD34+ (two tail T-test with Welch’s correction). Both cohorts are CD34+ cells sorted from bone marrow aspirates. E, H3K27ac ChIP-seq tracks for two AML patients with RARA among their top 100 enhancers ( crimson), two MDS patients with RARA among their top 100 enhancers ( red), and two non-malignant blast cell samples from healthy donors (gray). F, Expression of mRNA by RNA-seq (log2 TPM) across a set of primary AML patients, AML cell lines, and AML PDXs.

**Supplementary Figure S6** SY-1425 is a selective and potent RARα agonist. A, Biochemical assay plot of SY-1425 (tamibarotene) concentration (log nM) versus degree of association between co-activator and RAR protein from FRET assay expressed as % maximal ATRA signal. Data are for tamibarotene effects on RARα (red circles), RARβ (blue squares), and RARγ (green triangles). The horizontal dashed line indicates the tamibarotene Emax on RARα and the vertical dashed line the EC50 value. B, Cellular reporter plot of degree of association between co-activator and RAR protein from FRET assay expressed as % maximal ATRA signal (Emax) induced by SY-1425 (tamibarotene) as a function of concentration (log nM). Data are for SY-1425 on RARα (red circles) and RARγ (green triangles).

**Supplementary Figure S7** Comparison of SY-1425 response to RARA expression and enhancer score and the effect of an RARα antagonist. Correlations of enhancer, mRNA, and sensitivity to SY-1425 are color coded by cell line as indicated. A, Plot of RARα mRNA for AML cell lines versus SY-1425 anti-proliferative effect. B, Plot of RARA enhancer score for AML cell lines versus SY-1425 anti-proliferative effect. C, Plot of normalized enhancer versus expression. D, Cellular proliferation by ATPlite in RARA-high MV4-11 (red, circles), RARA-high OCI-AML3 (crimson, triangles), RARA-low OCI-M1 (blue, diamonds) and RARA-low Kasumi-1 (black, squares). Open shapes for BMS195614, a selective RARα antagonist, and closed for SY-1425. E, Anti-proliferative effect of SY-1425 (red circles) or idarubicin (black squares) with (open symbol) or without (closed symbol) 1µM BMS195614, a RARα selective antagonist, on the RARA-high cell line SigM5.

**Supplementary Figure S8** SY-1425 in vivo pharmacokinetics. A, Chemical structure of SY-1425 and plot of time versus SY-1425 concentration in plasma from mice given 3mg/kg SY-1425 (black circle) vs human SY-1425 levels (white diamond) 6mg/m2. Tables of PK properties for SY-1425 in mice are shown (top right) with comparator data from Kanai et al. Hem. Int. 2014. B, Chemical structure of ATRA with PK plot. 3mg/kg SY-1425 (black circle), 4mg/kg ATRA (white diamond), or human ATRA levels 45mg/m2. Tables of PK properties for ATRA are shown (bottom right) with comparator data (2).

**Supplementary Figure S9** SY-1425 survival and tissue effects on AML PDX models. A, RARA mRNA levels for the four non-APL AML PDX models assessed by RNA-seq with RARA-high indicated in red and RARA-low indicated in blue. B-C, Tumor burden in blood, bone marrow, and spleen was assessed by human CD45 positive cell percentage by FACS for (B) AM5512, and (C) AM8096. SY-1425 treated (6 mg/kg) animals are shown in red to indicate RARA-high status while vehicle treated animals are shown in gray. D, Survival plot for AM5512 PDX model with significant prolongation of survival (0.03 Mantel-Cox test). E, Survival plot for AM8096 PDX model with significant prolongation of survival (0.02 Mantel-Cox test). F-G, Tumor burden in
blood, bone marrow, and spleen was assessed by human CD45 positive cell percentage by FACS for (F) AM7577, and (G) AM7440. SY-1425 treated (6 mg/kg) animals are shown in blue to indicate RARA-low status while vehicle treated animals are shown in gray. **H**, Survival plot for AM7577 PDX model with no significant prolongation of survival (0.3 Mantel-Cox test). **I**, Survival plot for AM7440 PDX model.

**Supplementary Figure S10** SY-1425 shows greater potency than ATRA in vitro and in vivo. **A**, Plot of EC50 values for SY-1425 versus ATRA for AML cell lines with equipotency (solid), three fold stronger SY-1425 (dotted) and ten-fold stronger SY-1425 (dashed) lines indicated. **B-C**, Mice in AM5512 RARA-high PDX model were treated with 3mg/kg BID for SY-1425 or 4mg/kg BID for ATRA. Tumor burden in (B) spleen and (C) bone marrow are shown. Vehicle is shown in black; ATRA treatment in blue; and SY-1425 in dark red for mice sacrificed during the course of the study and light red for mice assessed at the end of two months of treatment. SY-1425 shows significant reduction versus ATRA (two tail T-test with Welch’s correction). **D**, Plot of percent human CD45 positive in vehicle (black) and SY-1425 3mg/kg BID (red) for peripheral blood (closed circle) or bone marrow (open squares).

**Supplementary Figure S11** Induction of markers of myeloid differentiation in RARA-high AML cell lines upon SY-1425 treatment. **A-B**, Induction of mRNA expression measured by microarray in cell lines treated with vehicle or SY-1425 50nM for 24 hours for (A) CD11c and (B) CD11b. **C**, FACS measurement of percent positive in vehicle or SY-1425 50nM 72 hours for CD66. The APL cell line NB4 is in blue, RARA-high cell line MV4-11 in red, and RARA-low cell line OCI-M1 in gray. **D,** Percentage of CD38 high cells in RARA-high versus RARA-low ex vivo samples after 48 hours. Significance by two tail T-test with Welch’s correction.

**Supplementary Figure S12** Comparison of SY-1425 expression and molecular response to APL. **A**, Canonical pathways (MSigDB C2.CP) and Hallmark (from MSigDB) gene sets enriched by GSEA in SY-1425 response in the RARA-high cell lines (Supplementary Table S6). Top 10 gene sets by FDR are shown. Positive normalized enrichment scores (NES) represent gene sets enriched in genes that are higher upon SY-1425 treatment, negative NES values represent gene sets enriched in genes that are lower upon SY-1425 treatment. **B**, Number of genes bound by RARα in NB-4 cell line in common with AML cell lines. Genes with RARα ChIP-seq peak (top 4000 peaks per cell line) in the gene regulatory region (promoter+1Mb up to neighboring gene’s promoter) are considered bound by RARα. NB4 bar indicates total number of genes bound in NB4. **C**, Heatmap of spearman correlation of super-enhancer (SE) response to SY-1425 in each cell line. Only SEs differential (FDR<0.05) in at least one cell line are included. Clustering is by Ward method. **D**, SE response to SY-1425 in NB-4 and RARA-high cell lines. Only SEs differential in NB-4 (FDR<0.05) (n=1967) are shown.

**Supplementary Figure S13** SY-1425 expression and molecular response. **A**, Volcano plots of H3K27ac response to SY-1425 for the union of H3K27ac peaks in the vehicle and treatment conditions by cell line. **B**, Percentage of H3K27ac peaks in each cell line that are higher upon treatment with SY-1425. Red bars indicate RARA-high lines and blue bars indicate RARA-low lines. **C**, Percentage of H3K27ac peaks in each set that are up-regulated by SY-1425 (FDR<0.05 and log2 fold change >1) in each cell line. RARα bound peaks overlap a RARα ChIP-seq peak (top 4000 per cell line), and background peaks are drawn from peaks with similar H3K27ac level.
to RARα bound peaks. Error bars are from bootstrapping. Numbers in the RARα bound bars indicate the number of H3K27ac peaks up-regulated and bound by RARα in that cell line.

**Supplementary Methods**

**Preparation of samples for ChIP-seq**

After thawing or processing, cells were resuspended in room-temperature (RT) PBS to a concentration of 1 million cells per mL. Freshly made fixation solution (50 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 11% Formaldehyde) was added at 100 ul per mL of cell mixture for a final formaldehyde concentration of 1%. Fixation was carried out at RT for 8 minutes with end-over-end mixing. To quench fixation, 55 uL of 2.5 M glycine was added per mL of original cell mixture for a final glycine concentration of 0.125 M. Quenching was carried out at RT for 5 minutes with end-over-end mixing. Cells were then centrifuged for 5 minutes at 4°C at 1500 relative centrifugal force (RCF). The supernatant was removed and the cells were washed with 1 mL of ice-cold 0.125 M glycine in PBS prior to a second spin as before. The final supernatant was removed using two pipetting steps and the cell pellet was flash frozen in liquid nitrogen and stored at -80°C.

**High-throughput ChIP-seq of H3K27ac and RARα**

Briefly, AML cells were resuspended in cold Lysis Buffer 1 (140mMNaCl, 1mMEDTA, 50mM HEPES, 10% glycerol, 0.5% NP-40, 0.25% Triton-X-100), incubated 10 min at 4 degrees C, and pelleted. Lysed AML samples were resuspended in Lysis Buffer 2 (10 mM TRIS [pH 8.0], 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), incubated for 10 min at 4 degrees C, and cell nuclei was then pelleted and resuspended in sonication buffer (10 mM TRIS [pH 8.0], 1 mM EDTA, 0.1% SDS). Cell nuclei were sonicated using a Covaris E220 sonication water bath for 5 min total time. Sheared chromatin was diluted 1:1 in dilution buffer (300 mM NaCl, 2 mM EDTA, 50 mM TRIS [pH 8.0], 1.5% Triton-X, 0.1% SDS) and incubated with either anti-H3K27ac (Active Motif, #39133) or anti-RARα (Santa Cruz, sc-551) conjugated Protein G Dynal beads (Invitrogen) overnight (8–16 hr, rotating) at 4 degrees C, and then washed two times with wash buffer 1 (20 mM TRIS [pH 8.0], 150 mM NaCl, 2 mM EDTA, 1% Triton-X, 0.1% SDS, 0.05% DOC), two times with high-salt wash buffer (20 mM TRIS [pH 8.0], 500 mM NaCl, 2 mM EDTA, 1% Triton-X, 0.1% SDS, 0.05% DOC), one time with LiCl buffer (250mM LiCl, 10mM TRIS 8.0, 1mM EDTA, 0.5% NP-40) and one time with TE-NaCl buffer (10 mM Tris [pH 8.0], 1 mM EDTA, 50 mM NaCl). Samples were eluted from beads for 1 hr at 65 degrees C in elution buffer (50 mM TRIS [pH8.0], 10 mM EDTA, 1% SDS) and supernatant reverse-crosslinked at 65 degrees C for 6–16 hr. Samples were diluted 1:1 with TE buffer (50 mM TRIS [pH 8.0], 1 mM EDTA) and treated with RNase A (0.2 mg/ml) for 2 hr at 37 degrees C and then Proteinase K (0.2 mg/ml) for 2 hr at 55 degrees C. DNA was isolated by phenol-chloroform extraction and ethanol precipitation or by AMPure XP beads (Agencourt) according to manufacturer instructions.

Libraries were constructed with the Illumina Tru-Seq library preparation kit using a target fragment size of 200–400 bp and multiplexing barcodes. Libraries were sequenced using Illumina HiSeq 2000 with single-end reads for 40 cycles. Sequences were demultiplexed and aligned using Bowtie2 against the human hg19 genome. This resulted in genome-wide BAM files summarizing the alignment of both the ChIP input and immunoprecipitate sequencing experiments.
**ChIP-seq of cell lines with treatment:**
On the day of the experiment, cell density was measured with a Countess Automated Cell Counter (ThermoFisher). For each AML cell line, 500,000 cells in 1 mL of the appropriate growth medium was added to each well of a 6-well plate. An additional 1 mL of growth medium or growth medium containing SY-1425 (100 nM, for a final concentration of 50 nM) was added to each well and incubated for 24 hours.

**ChIP-seq analysis**

**RARA ChIP**
For RARA ChIP-seq data, we first called peaks genome-wide using a pipeline combining MACS v2 (3) and IDR (3,4). First, aligned BAM files from the replicate input samples for each experiment were combined. Next, peaks were called on the aligned BAMs from each IP replicate using MACS2 with the combined input sample and a relaxed p-value threshold of 1e-3. In parallel, the BAM files from the individual IP replicates were combined and then randomly split into two pseudo-replicates. Peaks were then called on each of these pseudo-replicates and the combined input sample in the same manner. The IDR software was then run on the true replicates and the pseudo-replicates separately. The number of final peaks was determined by the maximum of the number of peaks from the true replicates with IDR threshold of 0.02 and the number of peaks from the pseudo-replicates with IDR threshold of 0.0025, up to a maximum of 25000. The top peaks up to the number of final peaks were taken from the IDR calls on the pseudo-replicates.

**H3K27ac ChIP**
For H3K27ac ChIP-seq data, we first called peaks genome-wide in the aligned H3K27Ac read data with MACS v1.4 using the aligned IP BAM as the ChIP-seq foreground data and the aligned IN BAM as the control background data. We used a stringent p value cutoff of 1E-9 but otherwise used the default parameters. We refer to this set of peaks as the MACS peaks or H3K27ac peaks. These peaks were then merged together if they had less than or equal to 12,500 base pairs between them in the human reference genome. We refer to this set of peaks as the ROSE peaks, and the rank of the highest scoring ROSE peak overlapping the RARα transcript for a given patient is recorded as “RARA ROSE Rank” for that patient.

The set of ROSE peaks was then merged together between all of the AML samples by taking the union of the coordinates of each peak from a given patient with all of the peaks that overlapped it from the other patients. This created a universal map of H3K27Ac enrichment. We then quantified each enriched region within this universal map within each patient by, for a given region, summing the number of IP reads that mapped within the region and dividing by the number of reads mapping in the full experiment multiplied by a million (“reads per million”, or RPM). We calculated a similar RPM score for the IN reads. We then subtracted the IN RPM from the IP RPM to achieve the overall score for a given region within the universal map for a given patient. We then fit a negative binomial distribution to the scores for a given patient using the fitdistr function in the R MASS library v7.3.45. We located the tail of the distribution as the point where the cumulative distribution function of this negative binomial crossed 0.985 (equivalent to a p values of 0.025). We divided the overall scores for all of the patient’s regions by this point, so that any region of enrichment that scored in the bottom 97.5% of the fitted negative binomial distribution scores below 1 (i.e. is a “typical enhancer”) and any region that scores above the 97.5% mark scores above 1 (i.e. is a “super-enhancer”). These scores are termed the SE scores for each patient.
against a universal map. Each patient’s SE scores were then normalized together using quantile normalization.

**Differential ChIP-seq binding**
To determine regions of differential binding of H3K27ac between treatments, the Bioconductor R package **DiffBind** (5) was used on either the called H3K27ac peaks from MACS or super-enhancers (SEs). The analysis was run on each cell line separately, using a unique set of regions per cell line merged from the two treatment conditions (SY-1425 and vehicle), and filtered for blacklist regions (5,6). **DiffBind** was run using the **DESeq2** method. The **DiffBind** output was used for the log fold-change values for H3K27ac response (either H3K27ac peaks or SEs). Differential binding of RARA was performed using the same procedure on peaks called using MACS v2.

**Meta-peaks**
For meta-peak analysis, the MACS peaks from each H3K27ac ChIP-seq sample were used. The peaks from each cell line (across both treatments and all replicates) were merged to create a peak map for each cell line. Each peak was then scaled to be 4kb wide. The reads per million from each IP sample was then calculated in each 100bp window in the 4kb peak. These were then normalized by dividing the values in each window by the ratio of the mean RPM in that sample in that window to the mean RPM in all samples in that window. This produces the same overall peak shape for each sample.

    The mean RPM in each window for each set of samples was calculated, and a smoothing spline was then applied.

**RNA-seq**

**RNA-seq experimental procedure**
AML samples were homogenized in 1ml Trizol and RNA isolated by RNAeasy (Ambion) according to manufacturer instructions. RNA libraries were prepared and sequenced as described previously (7).

**RNA-seq Analysis**
After sequencing the RNA as described previously (Loven et al. 2012), reads were aligned to the HG19 transcriptome using rssem v1.2.21 software (rssem-calulate-expression, parameters = --samtools-sort-mem 3G --ci-memory 3072 --bowtie-chunkmbs 1024 --bowtie2 --strand-specific) and then mRNA quantification was done using the same rssem suite (rssem-parse-alignments, rssem-build-read-index, rssem-run-em) and reported in transcripts per million (TPM), which represents an estimate of the number of transcripts for a given gene for every million transcripts detected. All protein coding genes were then extracted for each sample and their scores were quantile normalized together. We processed and normalized all AML patient samples from PDX models, primary AML patient samples, and AML cell lines together to create a universal mRNA score.

**Mapping SEs to genes**
SEs were mapped to genes by identifying the strongest Pearson correlations between the SE score and the amount of H3K27Ac in the promoter (-5 kb - +1 kb around the TSS, input subtracted, normalized through DESeq2 utilizing variance stabilizing transformation), for all promoters within 500 kb. A p-value cutoff for the correlations was established at .1 * the total number of
comparisons. All promoters surpassing this cutoff were linked to the SE. Any SEs without any links were linked to the best correlated promoter, or the two best if the second had a p-value within .01 of the first.

Validation of correlation links with PHi-C
To validate the SE to gene links in our AML samples, we used a promoter-capture HiC dataset (8) in monocytes. We only considered links with a CHiCAGO score of at least 5 in monocytes. A SE to gene link from our SE correlation linking method was considered validated if a link existed between an “other-end” that fell in the SE and the bait end corresponded to a gene linked to that SE. For each of the 72 primary H3K27ac ChIP-seq samples (62 AML, 4 HSPC, and 4 monocytes), we determined how many of the SEs in that sample were linked to a gene using the SE correlation linking method that was also validated using PChi-C. We used the average percent of SEs with validated links across all 72 samples as the total percent of links validated.

NMF-C for clustering of patients
6 factor non-negative matrix factorization with “Brunet” optimization implemented in the R NMF library (9) was performed on the matrix of SEs from patients, with SEs chosen such that they were within the top 200 SEs for at least one patient and they were in the top 300 most variable SEs over all patients. NMF was performed 1000 times and a consensus matrix was constructed by observing the number of NMF iterations that grouped samples together in the coefficient matrix by hierarchical clustering into six clusters. This consensus matrix was then clustered with k-medoid clustering, k=6. Samples not included in the factorization (flow sorted samples) were included in the clustering by solving for a new coefficient matrix using the inverse of the basis matrix and the excluded samples SE scores.

Targeted gene sequencing of AML patients
Genotyping of AML samples was performed as described previously (10) by either by whole exome sequencing using the SeqCap EZ Exome SR kit v3.0 (Roche/Nimblegen) or by customized hybrid capture sequencing of the 130 genes most frequently mutated in AML(11) using the SeqCap EZ Choice kit (Roche/Nimblegen). Sequencing was performed on an Illumina HiSeq 2000, HiSeq 2500, or NextSeq 500. Sequence data were aligned to the human reference genome hg19 using BWA (v0.5.9) for global alignment and GATK (v2.8-1) for local realignment. Aligned reads were processed for downstream mutation calling using SAMTools (v0.1.12a). SNPs were called using GATK and Varscan (v2.3.7). All data derived from customized hybrid capture did not have a matched normal genome and was compared instead to the hg19 human reference genome. Putative SNPs were filtered for: 1) minimum sequence depth of 50 reads, 2) less than 90% variant strand bias, 3) non-synonymous, 4) if the SNP is observed in dbSNP, the MAF must be less than 1%, 5) minimum variant frequency of 5%. Insertions and deletions (indels) were called using GATK (12) and Varscan. Putative indels were filtered for: 1) minimum sequence depth of 25, 2) minimum variant frequency of 5%, 3) less than 90% variant strand bias, 4) not observed in dbSNP. Large-scale genomic events such as translocations were called using FACTERA (v1.3) with no additional filtering. FLT3 internal tandem duplications were called using Pindel (v0.2.4) with no additional filtering. Manual observation was used to clarify borderline mutation calls.

Mapping SE clusters to TCGA transcriptomes
An RNA-seq classifier for cluster identity was created using the nearest shrunken centroids algorithm (13). The shrinkage parameter which minimized error on leave-one-out cross validation was selected. Raw RNA-seq from the TCGA AML cohort was realigned identically to the samples described here, and the pamr classifier was used to predict clusters for the TCGA cohort.

Survival analyses
Overall survival was defined as the time from diagnosis to death from any cause. Survival analysis was performed using the Kaplan-Meier estimate method. All patients were included for the analysis regardless of their treatment. P values comparing Kaplan-Meier survival curves were calculated using the log-rank (Mantel-Cox) test.

Permutation analysis of survival differences
To confirm that the differences in survival we identified were not an artifact of our mapping method, we scrambled the cluster labels in our cohort 1000 times. For each of these permutations, we learned a new transcriptome-based cluster classifier, and predicted cluster membership in the TCGA dataset. We took the fraction of permutations which resulted in a p-value at least as extreme as the one discovered on the original clustering as the empirical p-value.

Predicted HSPC Signature
Predicted HSPC signatures were derived for patient samples by reconstructing each patient sample with non-negative least squares regression over the flow sorted HSPCs and monocytes. The total regression weights were summed and the proportion coming from HSPCs constitutes the predicted HSPC signature.

HSPC / Monocyte centroid distance
Using the 20 SEs most associated with HSPCs and the 20 SEs most associated with monocytes from a DESeq2 differential analysis of those two groups, centroids for HSPCs and monocytes were derived. Patients were then scored by distance to these centroids by Euclidean distance.

Motif Analysis
Cluster-specific SEs were derived by identifying those SEs where the median per-cluster scores when hierarchically clustered using complete linkage were separated by more than 1, and where an ANOVA across all patients given the cluster labels was significant at p=0.01. These peaks were searched for motif occurrences in the 1kb centered on the summits of each peak using FIMO (14) with a p-value threshold of 1e-4. Concurrently, we selected a set of background regions comprised of random 1kb genomic intervals pulled from a set of open chromatin regions across 16 ENCODE cell types (15) (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeOpenChromSynth/). Next a one-sided Wilcoxon test was performed for the enrichment of the p-value of the best motif matches in the consensus peaks for each cluster compared to the background regions. The p-values were then adjusted using the Benjamini & Hochberg FDR method.

Cell Viability
On the day of the experiment, cell density was measured with a Countess Automated Cell Counter (ThermoFisher). For each AML cell line, sufficient cells were transferred to a flask and adjusted to 60,000 cells/ml with appropriate growth medium. Using a Biotek EL406, 50 μl of cell
suspension was distributed into white 384-well plates (Corning #3570). Immediately after plating cells, compounds were added to plates using a 20-nl 384-well pin transfer manifold on a Janus workstation (PerkinElmer). Compounds from DMSO stock solutions were arrayed by dilution for 10-point quadruplicate concentration-response testing in 384-well compound plates. After addition of compound, plates were incubated for five days at 37°C under 5% CO₂.

On Day 5, the number of viable cells was estimated using ATPlite (PerkinElmer). ATPlite measures the amount of ATP present in each culture well, which reflects the number of cells and their average ATP content. Plates were removed from the incubator and cooled to room temperature prior to measurement. Lyophilized powder of ATPlite reagent was re-suspended in lysis buffer and diluted 1:2 with distilled water. 25 μL of this solution was added to each well using a Biotek liquid handler. Plates were incubated for 15 min at room temperature before the luminescence signal was read on an Envision Plate Reader (PerkinElmer).

Data acquired as described above was stored and grouped in Microsoft Excel. Percent viability was calculated on a per-plate basis by dividing the luminescent signal value of a given treatment well by the average DMSO control value calculated as the average of 20 wells treated with DMSO vehicle only. Data were fit to estimate EC₅₀ and Emax values with GraphPad Prism version 6.0 using four-parameter (Hill slope not assumed to be equal to 1) non-linear regressions with the log₁₀-transformed compound concentrations plotted against the percent viability of the cells normalized to DMSO-only treated wells included on the plate. The equation of the curve fit is:

\[ Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + 10^{(\text{LogEC}_{50} - X) * \text{HillSlope}}} \]

Wells located at the edges of plates were excluded from analysis.

### Mutual Information Network

To construct the mutual information network, first, a set of informative SEs was derived as those SEs where an ANOVA over the log2 score of the SEs showed the cluster term to be significant at a p-value of 0.01. The informative SEs were further filtered by selecting only those where a complete linkage tree of the median cluster SE scores weighted by the number of observations per cluster when cut at a height of 1 resulted in more than one hierarchical cluster. The SE scores for this set of informative SEs across patients was then used as an input matrix to the “mrnet” algorithm in the “minet” BioConductor package using a shrinkage estimator of mutual information and a global-equal-width discretization scheme, with number of bins equal to the floor of the square root of the number of patients. Edges with a mutual information score < .2 were then pruned, and then nodes were labeled if they were a TF. Nodes were colored based on which of the patient clusters were on the high side of the split in the complete linkage tree above. Nodes with motif enrichments that showed preferences for SEs (using methods above) in clusters that correspond to their coloring were outlined in red.

### Cell line expression profiling preparation

On the day of the experiment, cell density was measured with a Countess Automated Cell Counter (ThermoFisher). For each AML cell line, 500,000 cells in 1 mL of the appropriate growth medium was added to each well of a 6-well plate. An additional 1 mL of growth medium or growth medium containing SY-1425 (100 nM, for a final concentration of 50 nM) was added to each well and incubated for 24 hours.

The cell suspensions were transferred to microcentrifuge tubes whereby they were centrifuged for 3 minutes at 300 x g. The supernatant was removed and the pellets washed with l
mL of phosphate-buffered saline (PBS). The pellets were centrifuged again to remove the supernatant and then re-suspended in 200 µL of TRIzol Reagent. 20 µL of miRNA Homogenate Additive (100 µL per 1 mL of TRIzol) from the miRVana miRNA Isolation Kit is mixed in with the TRIzol cell lysate. The mixture was incubated on ice for 10 minutes. 20 µL of bromochloropropane is then mixed in with the lysate. The mixture was incubated at room temperature for 5 minutes after which it was centrifuged at 12,000 x g for 10 minutes at 4°C. 62 µL of the aqueous phase is added to 78 µL of ethanol and transferred to a filter cartridge.

As detailed in the kit’s Total RNA Isolation protocol, the filter cartridge containing the aqueous phase is spun for 15 seconds at 10,000 x g. 700 µL of miRNA Wash Solution 1 was added to the filter cartridge and centrifuged for 10 seconds at the same speed. For a total of two washes, the filter cartridge is washed with 500 µL of Wash Solution 2/3 and centrifuged. 100 µL of pre-heated (95°C) Elution Solution was then added to the center of the filter and spun for 30 seconds at maximum speed. RNA concentration was measured using a NanoDrop 2000c Spectrophotometer (Thermo Scientific).

**Affymetric Array Analysis**

RNA prepped as described above was also sent to the Dana Farber Cancer Institute Molecular Biology Core Facilities for analysis on Affymetrix Prime View arrays. The DFCI MBCF ran the Prime View arrays and provided raw CEL files to Syros for analysis.

These raw CEL files from an affymetrix array were processed using the R Bioconductor package `affy`. The data was first normalized using the justRMA() function and default parameters; this produced a set of expression measurements for each probe in each sample.

Differential expression analysis was performed using the R Bioconductor package `limma`. First, a linear model was fit using the treatment status of the sample as a coefficient and the contrast of interest. Next, a moderated t-test was performed using empirical Bayes moderation to test for a difference in expression of each probe in the two treatment conditions. The output of this test (log fold-change, p-value, and adjusted p-value) was used for subsequent analysis. Volcano plots were produced by plotting the log fold-change of each probe in the two conditions by the –log10(p-value).

To determine which affy probes were differentially expressed upon SY-1425 treatment in AML cell lines, in either RARA-high (OCI-AML3, MV4-11, and SigM5) or RARA-low (OCI-M1, KG-1a, and Kasumi-1) cell lines, the expression data from all three cell lines were used together in the linear model. The linear model only included a coefficient for the treatment used. Gene set analysis was performed using the Gene Set Enrichment Analysis (GSEA) software (Subramanian et al. 2005). The log fold-change value from the differential expression analysis was used as the metric for enrichment.

**Gene regulatory regions**

For associating RARα binding sites to genes, each gene was assigned a regulatory region based on the method used in GREAT (McLean et al. 2010). The basal regulatory region (promoter) for each gene is the region 5kb upstream and 1kb downstream from the transcription start site (TSS). The full regulatory region of each gene comprised of the basal regulatory region plus up to 10kb or 1Mb in each direction until the basal regulatory region of the next closest gene.

**Other statistical analysis**
R was used for all basic statistical analysis. All hierarchical clustering was using the hclust() function with the “complete” method.

**In vitro selectivity studies**
Selectivity studies were performed using the SelectScreen™ Nuclear Receptor Profiling Services (ThermoFisher Scientific) at their site. Biochemical selectivity utilized a cell free LanthaScreen™ FRET based coactivator assay. Cellular selectivity used the GenBLAzer™ reporter based assay.

**In vivo studies**
SY-1425 was prepared as a 100x stock solution in dimethylsulfoxide (DMSO; 60 mg/mL) and stored at -20°C. For administration to mice, this SY-1425 stock solution was diluted to 1% in PBS adjusted to pH 8. To study its effects in AML PDX models, SY-1425 was administered to mice orally once-daily at 6 mg/kg in a 10 ml/kg volume. Mice in the vehicle arm were given the same volume of vehicle on the same dosing schedule.

All mice used in this study were non-obese diabetic severe combined immunodeficiency strain (NOD/SCID) females, except for the AM7440 model which used the NOD/SCID gamma (NSG) strain.

Patient-derived xenograft models were initiated by tail vein intravenous injection of AML cells into each mouse. Engraftment and tumor development was monitored weekly before and through the dosing period by assessing the number and percentage of human CD45-positive cells. For each model, groups of three mice were assigned to vehicle and treatment groups after reaching initiation criteria. Tumor burden by %CD45 positivity was balanced to give equivalent average levels between the two groups. SY-1425 (6 mg/kg) or vehicle was administered daily by oral gavage in 10 ml/kg dose solution. Dosing was continuous throughout the study unless a mouse lost ≥15% of body weight, at which point a dose holiday was allowed prior to resuming treatment. Dose holidays accorded during the study for mice receiving SY-1425 are detailed in the following table. No dose holidays were required for vehicle-treated mice.

**CD45 positivity assay**
CD45 is a cell surface marker useful in detecting leukemia cells. The antibody used here is specific for the human protein (does not bind mouse CD45) and can therefore be used to count the number of human leukemia cells present in a blood or tissue sample. Weekly blood draws were taken from the mice by retro-orbital puncture for determination of CD45-positive cell number and percentage white blood cells. Bone marrow and spleen were collected at the end of dosing for analysis along with blood. Bone marrow was flushed from mouse femurs and gently mixed in a 15 mL tube to homogenize. Spleen was ground with a cell strainer and processed with lysis buffer as was done with blood.

**Flow cytometry**
AML cell lines were treated for 72hour with 50nM tamibarotene or DMSO. Cell surface expression was quantified using CD38 and CD66b fluorescent antibodies on a BD Fortessa FACS Caliber. Antibodies used were FITC-CD38 (BioLegend catalog # 356610), BV711-CD66 (BD Biosciences catalog # 740805).

Primary AML patient samples were obtained from Conversant Bio (samples 1, 2, 3, 5) and iSpecimen (sample 4), thawed and treated with 50nM tamibarotene or DMSO for 24h and 48h. Cell surface expression of CD38 was quantified using CD38-FITC antibody on a BD
Fortessa FACS caliber. Sample 1 and 5 were bone marrow aspirates from Caucasian male subjects aged 80 and 90 respectively. Samples 2 and 3 were PBMCs from Caucasian male MDS patients aged 78 and 77 respectively. Sample 4 was a bone marrow aspirate from an AML patient.

Supplementary References


4. Kundaje A. TF ChIP-seq peak calling using the Irreproducibility Discovery Rate (IDR) framework [Internet]. 2012. Available from: https://sites.google.com/site/anshulkundaje/projects/idr


