Supplementary methods

Next generation sequencing data analysis

Data were aligned using BWA (version 0.7.4-r385)\(^1\) and then pre-processed using GATK (version 2.7-2-g6bda569) per recommended best practices recommended by the Broad Institute\(^2\-^4\). The somatic SNV analysis was performed using Mutect (version 1.1.4)\(^5\), Varscan (version 2.3.6; samtools version 0.1.19)\(^6\), and somaticSniper (version 1.0.0.1)\(^7\). Specifically, the mutect command is: “java -Xmx2g -jar muTect-1.1.4.jar -rf BadCigar --analysis_type MuTect --reference_sequence hg19.fa --cosmic b37_cosmic_v54_120711_chr.vcf --dbsnp dbsnp_137.hg19.sorted.vcf --input_file:normal normal.bam --input_file:tumor tumor.bam --out output.mutect.out.txt --coverage_file output.coverage.wig.txt -vcf output.vcf input.bam”. The varscan command is: “java varscan somatic <(samtools mpileup -q 20 -f hg19.fa normal.bam) <(samtools mpileup -q 20 -f hg19.fa tumor.bam) output --output-snp output.snp.vcf --output-indel output.indel.vcf --min-coverage 10 --min-coverage-normal 10 --min-coverage-tumor 10 --min-var-freq 0.05 --min-freq-for-hom 0.75 --normal-purity 1.00 --tumor-purity 1.00 --p-value 0.99 --somatic-p-value 0.05 --strand-filter 1 --output-vcf 1”. The somaticSniper command is: “bam-somaticsniper -f hg19.fa -q 20 -Q 15 -F vcf tumor.bam normal.bam output.snp.ss.vcf”.

Somatic mutations were determined by comparing mutations detected in the lymphocyte germline DNA to the mutations detected at each time point. The thresholds for somatic variant calls required that the variant be observed on reads from both strands, with a minimum tumor variant allele frequency (VAF) of 5%. The significance of allele frequency in varscan difference was calculated by Fisher's exact test. The p-value
threshold for calling variants was < 0.05. For varscan calling, the results were further filtered using the following three commands for SNVs and INDELs: “java -d64 -Xms4g -Xmx4g -jar VarScan.v2.3.6.jar somaticFilter output.snp.vcf --min-coverage 8 --min-strands2 2 --min-var-freq 0.05 --p-value 0.05 --indel-file output.indel.vcf --output-file output1.snp.vcf” and “java -d64 -Xms4g -Xmx4g -jar VarScan.v2.3.6.jar processSomatic output1.snp.vcf --min-tumor-freq 0.05 --max-normal-freq 0.02 --p-value 0.05” and “java -d64 -Xms4g -Xmx4g -jar VarScan.v2.3.6.jar processSomatic output.indel.vcf --min-tumor-freq 0.05 --max-normal-freq 0.02 --p-value 0.05”. For Varscan SNVs and INDELs, only high confidence somatic mutations output from these commands were included. For mutect callings we removed all the rejected mutations. For somaticsniper calling, the somatic score is at least 40 and the mapping quality is at least 40.

For somatic SNVs, we used a minimal average base quality for variant supporting reads of 20, minimum read coverage of at least 10 reads in both tumor and matched normal samples, minimum tumor alternative allele frequency of at least 5%, and normal sample alternative allele frequency less than 5%. Only somatic SNVs detected by at least two callers were used. For somatic indels, we used minimum read coverage of at least 10 reads in both tumor and matched normal samples, minimum tumor alternative allele frequency of at least 5%, and normal sample alternative allele frequency less than 2%.

Mutations located in the overlapping regions between NimbleGen SeqCap EZ Human Exome Library v3.0 and Agilent Human Exon V3 were kept. We then keep the mutations in the loci of somatic mutations with at least 10 reads in more than 108 samples sequenced. Mutations were further filtered if not located in targeted regions by Agilent SureSelect Human All Exon V4, unless the loci were covered by at least 10 reads in at
least 140 samples sequenced. For Whole genome sequencing (WGS) data, only somatic mutations with alternative allele frequency greater than 20% were included. The somatic SNVs were used to for calculating the somatic mutation burden and allele frequency density for WGS data of AML_130.

Mutations were annotated using SnpEFF (version 3.6)\(^8\). The SnpEFF command is “snpEff.jar -s output.stat hg19 -v input.vcf -cancer -oicr -sequenceOntology output.eff.vcf”.

For cluster 1 and 3 determined by eloci proportion, the proportion of somatic mutations for diagnosis and relapse were compared using Wilcoxon signed rank test within each cluster to measure the significance of dominance, respectively. The number of somatic mutations between cluster 1 and 3 were compared by Wilcoxon rank sum test at diagnosis and relapse stage, respectively. The proportions of somatic mutations with increased or decreased VAF by at least 10% from diagnosis to relapse were compared between cluster 1 & 3 using Wilcoxon rank sum test for those loci that covered by at least 10 reads in tumor samples and matched normal samples for each patients.

Copy number aberrations (CNA) were obtained from Varscan (version 2.3.6) and XHMM\(^9\) (version 1.0). The Varscan CNA commands are: “java -Xms4g -Xmx4g -jar VarScan.v2.3.6.jar copynumber <(samtools mpileup -q 1 -f hg19.fa normal.bam tumor.bam | awk 'NF==9 && $4!=0') output --mpileup 1 --min-base-qual 20 --min-map-qual 20 --min-coverage 30 --min-segment-size 100 --max-segment-size 10000 --p-value 0.01 --data-ratio 1.0”, java -Xms4g -Xmx4g -jar VarScan.v2.3.6.jar copyCaller output.copynumber --output-file output.copynumber.called --output-homdel-file output.copynumber.called.homdel”. The results were further refined by binary
segmentation using the DNAcopy library\textsuperscript{10} (version 1.42.0) from BioConductor. The XHMM parameter is: exome-wide CNA rate = 1e-08, mean number of targets in CNA = 6, mean distance between targets within CNA (in KB) = 70, mean of deletion z-score distribution = -3, standard deviation of deletion z-score distribution = 1, mean of deploid z-score distribution = 0, standard deviation of deploid z-score distribution = 1, mean of duplication z-score distribution = 3, standard deviation of duplication z-score distribution = 1. The CNA results from Varscan and XHMM were merged to achieve a comprehensive CNA prediction.

Genetic clonal evolution analysis was performed using sciClone (version 1.0.7)\textsuperscript{11} to infer the clonality based on CNAs and SNVs with maximum of 6 clusters at minimum read depth of 50. The sciClone command was executed in R: “sciClone(snv, copyNumberCalls = cnv,sampleNames =’sample’,cnCallsAreLog2 = FALSE, minimumDepth = 50, maximumClusters = 6)”. Only samples with at least 6 or more somatic SNVs in the copy number neutral regions were included for sciClone analysis (n = 40). For those loci that only have somatic mutations called at one stage, the other stage’s coverage of all read and alternative allele reads was calculated using GATK UnifiedGenotyper (version 2.7-2). For those locations not called by UnifiedGenotyper, total read coverage was profiled using bedtools coverage, and alternative allele reads were assigned as zero. Clonal complexity was measured by the number of subclone clusters predicted by sciClone. A sample has an increased clonal complexity if there are more subclone clusters predicted by sciClone at relapse stage than diagnosis stage. A cluster was detected in one stage if the mean alternative allele frequency of the cluster at the stage is at least 5%. A sample has a decreased clonal complexity if there are fewer
subclone clusters at relapse stage than diagnosis stage. Otherwise, the sample’s clonal complexity is defined as stable. The proportions of increased, stable, and decreased clonal complexity between epigenetic clusters were compared using chi-square test. The somatic SNVs with neutral copy number mutations were used to measure the proportion of somatic mutations with increase or decreased variant allele frequency by at least 10% at relapse. This proportion of somatic mutations was then compared between eloci clusters 1 and 3 using Wilcoxon rank sum test.
Supplementary table and figure legends

**Supplementary Table 1**: Summary of patient characteristics. Summary table of clinical characteristics of patients in study cohort. Plus minus values are means ± standard deviation. Percentages may not total 100 because of rounding. MRC = Medical Research Council risk stratification; ELN = European Leukemia Net risk stratification; Ara-C = cytarabine arabinoside; AlloSCT = Allogeneic stem cell transplant; AutoSCT = Autologous stem cell transplant.

**Supplementary Table 2**: Detailed description of patient characteristics and genomics assays performed. Clinical characteristics are detailed for each patient included in the study cohort. Genomic assays performed on each sample are indicated. WBC = peripheral white blood cell count; BM = bone marrow; WGA (°) = whole genome amplified; * = cytoreduction; Protocol: 1 = NimbleGen SeqCap EZ Human Exome Library v3.0, 2 = Agilent Human Exon V3 (Exon 50Mb), 3 = Agilent SureSelect Human All Exon V4 (51 MB); AlloSCT = Allogeneic Stem Cell Transplant; AutoSCT = Autologous Stem Cell Transplant; ATRA = All Trans Retinoic Acid; KW-2449 = investigational FLT3 tyrosine kinase inhibitor.

**Supplementary Table 3**: Sequencing statistics from ERRBS. Sequencing parameters from all ERRBS libraries prepared and sequenced. Sample ID = assigned sample ID (Dx = Diagnosis; Rel = Relapse; AML = Acute Myeloid Leukemia; NBM = normal bone marrow CD34+). SRA biosample_id = assigned biosample identification code in dbGap accession number phs001027.v1.p1. CpG covered = number of CpGs covered per sample sequenced at a minimum of 10X. BS = Bisulfite.

**Supplementary Table 4**: Sequencing statistics from Genomic Sequencing. Sequencing parameters from Whole Exome and Whole Genome libraries prepared and sequenced. Sample ID = assigned sample ID (Dx = Diagnosis; Rel = Relapse; AML = Acute Myeloid Leukemia; G = patient matched germline sample). Protocol = sequencing protocol executed (WES = Whole Exome Sequencing; WGS = Whole Genome
Sequencing). SRA biosample_id = assigned biosample identification code in dbGap accession number phs001027.v1.p1. Mapping_Rate = read mapping rate. Coverage_Rate_x15 = percent targets covered at a minimum of 15X. Average_Coverage = average coverage per target.

**Supplementary Table 5**: Patterns of EPM and genetic changes in epigenetic clusters 1 and 3. Median EPM (Epialleles per Million loci), somatic mutations and VAF changes between diagnosis and relapse AML patient specimens annotated into epigenetic clusters 1 and 3 (Wilcoxon rank sum test).

**Supplementary Table 6**: Somatic mutations in recurrently affected AML genes. Somatic mutation (SNVs and indels) determined from exome capture data analysis in genes recurrently affected in AML\textsuperscript{14,15} Sample ID = assigned sample identification number; Ref = sequence present in the reference genome; Alt = sequence detected in tumor sample; Func.refGene = genic region in which mutation detected; Gene.refGene = gene symbol; ExonicFunc.refGene = mutation type; AAChange.refGene = predicted amino acid change in translated gene; VAF\_diagnosis = alternative variant allele frequency in diagnostic sample; VAF\_relapse = alternative variant allele frequency in relapsed sample; N.D.= not determined due to inability to determine the variant allele frequency accurately; N.A.= not available due to lack of coverage in data generated.

**Supplementary Table 7**: Sequencing statistics from RNA-Sequencing. Sequencing parameters from RNA-sequencing libraries prepared and sequenced. Sample ID = assigned sample ID (Dx = Diagnosis; Rel = Relapse; AML = Acute Myeloid Leukemia; scRNAseq = single cell RNA-sequencing). Protocol = sequencing protocol executed. SRA biosample_id = assigned biosample identification code in dbGap accession number phs001027.v1.p1. Mapping_Rate = read mapping rate. Total_Read_Pairs = Number of reads per sample. RefSeq_Gene_Number = number of Refseq defined genes covered per sample. Batch_information: 1 = first batch of bulk sample RNA-seq libraries prepared and sequenced; 2 = second batch of bulk sample RNA-seq libraries prepared and sequenced. Epigenetic_Cluster = epigenetic cluster designation (1 versus 3).
Supplementary Table 8: Differentially expressed genes between epigenetic clusters 3 and 1. Table of differentially expressed genes (DEGs) determined using bulk cell RNA-sequencing data from patients in epigenetic cluster 1 compared to patients in epigenetic cluster 3 using DESeq2. DEGs were defined by an absolute log2 fold change greater than 1.2 and adjusted $P < 0.05$. Positive Log2 fold change represents up-regulated genes in Cluster 1 patients. The normalized gene expression determined using the DESeq2 variance stabilizing transformation approach (Mean_VSD) is included for each gene in each cluster.

Supplementary Table 9: GO term enrichment analysis of differentially expressed genes between epigenetic clusters 1 and 3. DEGs determined between epigenetic cluster 1 and 3 patients were analyzed using GO term enrichment analysis.

Supplementary Table 10: Somatic mutations gained in AML_130 at first relapse time point (T2). VAF = Variant Allele Frequency.

Supplementary Table 11: Somatic mutations detected in AML_130 at T1 – T5 time points. VAF = Variant Allele Frequency. T1 = Diagnosis; T2 = relapse; T3 = relapse 2; T4 = relapse 3; T5 = relapse 4.

Supplementary Table 12: Genes associated with eloci linked to clinical outcomes. Eloci were determined between diagnosis and NBM samples in patients with longer (n=68) or shorter (n=69) times to relapse based on the median value. The frequency of promoter-associated eloci in the patient groups was assessed using an odds ratio and significance was determined using a fisher's exact test for each gene. Eloci with a $P$ less than 0.05 were included in the table. Multiple hypothesis testing was performed (Benjamini-Hochberg). Chromosome, Start and End columns designate the eloci genomic coordinates. Symbol = gene symbol. Number_Short_Relapse_Patient = number of eloci detected in the shorter time to relapse group. Total_Number_Short_Relapse_Patient = total number of loci covered by ERRBS in the shorter time to relapse group.
Number_Long_Relapse_Patient = number of loci detected in the longer time to relapse group. Total_Number_Long_Relapse_Patient = total number of loci covered by ERRBS in the longer time to relapse group.
Supplementary Figure legends

Supplementary Figure 1: Schematic diagram of epiallele shift measurement using methyl-cytosine sequencing data. Open circles indicate unmodified cytosines while filled circles represent methylated cytosines. Epialleles are defined as the collection of phased DNA methylation patterns at genomic loci with four adjacent CpG sites. Eloci are loci with high epiallele compositional change (epiallele shift; shown is a representative elocus sample from relapse (R) versus diagnosis (D) sample). Overall differences between two methylomes are defined by the number of eloci normalized by the number of loci covered in the respective sequencing library (EPM = Eloci Per Million Loci).

Supplementary Figure 2: Representative results from methyl-capture assays of AML_103 patient specimens. Roche (a-e) and Agilent (f-j) capture assay results were compared to ERRBS results. Scatter plots of delta entropy measured for epiallele shift using ERRBS compared to results from Roche (a) or Agilent (f) capture assays. (b-e) Four loci with high epiallele shift validated by Roche capture approach. (g-j) Four loci with high epiallele shift validated by Agilent capture approach.

Supplementary Figure 3: Somatic mutations in low and high EPM patient groups. Violin plot of the number of somatic mutations in high and low EPM patients from Fig. 1 (Wilcoxon rank sum test).

Supplementary Figure 4: EPM patterns observed at diagnosis and relapse are independent of clinical features. (a-b) Box plots of EPM (log_{10}) detected in diagnosis and relapse patient samples versus NBMs grouped by FAB disease classification. (c-d) Scatter plots of EPM (log_{10}) detected at diagnosis and relapse compared to patient ages. (e-f) Scatter plots of blast purity (post lymphocyte depletion of AML_102 through AML_140 assessed by flow cytometry) compared to EPM (log10) detected at diagnosis (e) and relapse (f). (g,h) Scatter plot of EPM (log_{10}) versus number of somatic mutations (log_{10}) at diagnosis (g) or relapse (h). Pearson correlation was used to calculate r and Hoeffding’s D statistics was used for independency tests.
**Supplementary Figure 5:** Genomic distributions of eloci between diagnosis and relapse AML samples compared to NBM samples. (a) Schematic of CpG islands and flanking regions used for eloci annotation. (b-d) Box plots of the distribution of eloci localized to CpG islands, shores and shelves in diagnostic and relapsed samples compared to NBMs. Wilcoxon signed rank test: $P = 2.9 \times 10^{-3}$ (b), $P = 0.26$ (e), and $P = 0.054$ (d). (e) Schematic of genic components used for annotation of eloci. (f-i) Box plots of the distribution of promoter, exon, intron and intergenic eloci annotations in diagnostic and relapsed samples compared to NBMs. Wilcoxon signed rank tests: $P = 0.027$ (f), $P = 6.5 \times 10^{-8}$ (g), $P = 5.3 \times 10^{-3}$ (h), and $P = 7.6 \times 10^{-4}$ (i). (j-k) Box plots of the distribution of eloci in active and poised enhancers. Wilcoxon signed rank tests: $P = 1.4 \times 10^{-4}$ (j) and $P = 7.5 \times 10^{-3}$ (k). $P$ value definitions: *** = < 0.001, ** = <= 0.01, * = <= 0.05. All the distributions were significantly different from the background loci distributions covered by reduced representation bisulfite sequencing data ($P > 0.01$).

**Supplemental Figure 6:** Associations of K-means clustering of AML patients by eloci abundance at diagnosis and relapse with clinical and genetic parameters. (a) Heatmap of the proportion of eloci that are diagnosis-specific, shared, or relapse-specific (yellow = cluster 1; blue = cluster 2; orange = cluster 3). (b) Age (years) distribution of patients (at diagnosis) in eloci clusters 1-3. (c) Diagnostic sample FAB classification distribution of patients in eloci clusters 1-3. (d) Blast purity distributions of patients in eloci clusters 1-3 (diagnostic sample post lymphocyte depletion of patient samples AML_102 through AML_140). (e) Proportion of patients with higher (High MUT) or lower (low MUT) number of somatic mutations at diagnosis from each eloci cluster (chi-square test). (f) Heatmap of AML-specific recurrent somatic mutations in the epigenetically defined patient clusters 1-3. Light blue = diagnosis-specific mutations; blue = relapse-specific mutations; dark blue = shared mutations between diagnosis and relapse; white = not detected in assay; grey = not covered in assay.

**Supplemental Figure 7:** Mutant allele frequency (VAF) plots of genetic clonal evolution between diagnosis and relapse disease time points as predicted from sciClone analysis.
(a) Cluster 1-assigned patients (patients characterized by predominance of diagnosis-specific eloci). (b) Cluster 2-assigned patients (patients characterized by shared eloci between diagnosis and relapse). (c) Cluster 3-assigned patients (patients characterized by predominance of relapse-specific eloci). Colored symbols are standard sciClone output representing the subclonal cluster predictions.

Supplemental Figure 8: Proportions of evolved genetic subclones as determined from sciClone analysis in the epigenetically defined patient clusters 1-3. Increased (green) = subclones with at least 10% VAF increase from diagnosis to relapse; decreased (red) = subclones with at least 10% VAF decrease from diagnosis to relapse; stable (blue) = subclones with < 10% VAF change between diagnosis and relapse. No significant difference of patient proportions with decreased/increased/stable number of subclones was observed among clusters ($P = 0.915$). $P$ value was calculated using chi-square test.

Supplementary Figure 9: Frequency of shared and unique alleles during tumor progression in AML_130. Pie charts of the frequency of eloci (a) and somatic mutations (b) in five time points (T1-T5) from patient AML_130.

Supplemental Figure 10: DNA methylation heterogeneity in AML. (a) Intra-tumor DNA methylation heterogeneity (as measured by epi-polymorphism; MH) at diagnosis compared to relapse samples (n=138 samples from each group; Wilcoxon signed rank test: $P = 0.005618$). (b) Kaplan–Meier plot comparing the relapse-free survival between AML patients with high MH versus low MH (divided by median MH). MH levels at diagnosis do not segregate patients with distinct clinic outcomes (n = 137). Mantel Cox log rank test was used for the survival analysis ($P = 0.168$; Red: high MH patients; Black: low MH patients).


ERRBS $r = 0.898$

ERRBS $r = 0.868$

Epiallele patterns proportion

Stage

chr1: 166853610-166853619
chr7: 47576885-47576892
chr13: 25506204-25506236
chr20: 62318596-62318613
chr5: 87440240-87440266
chr6: 159589687-159589704
chr19: 719952-719986

ERRBS $r = 0.898$

ERRBS $r = 0.868$

Epiallele patterns proportion

Stage

chr1: 242688494-242688524
chr5: 87440240-87440266
chr6: 159589687-159589704
chr19: 719952-719986
Supplementary Fig. 3

$P = 0.778$

Somatic mutations (log10)

Nature Medicine: doi:10.1038/nm.4125
Supplementary Fig. 4

(a) Diagnosis

Diagnosis vs. NBM EPM (log10)

M1 M2 M4 M5

FAB classification

P = 0.587; n = 56

(b) Relapse

Relapse vs. NBM EPM (log10)

M1 M2 M4 M5

FAB classification

P = 0.926; n = 56

(c) Diagnosis

Diagnosis vs. NBM EPM (log10)

Age

r = 0.116; n = 138

(d) Relapse

Relapse vs. NBM EPM (log10)

Age

r = 0.017; n = 138

(e) Diagnosis

Diagnosis vs. NBM EPM (log10)

Blast purity post separation

r = 0.273; n = 36

(f) Relapse

Relapse vs. NBM EPM (log10)

Blast purity post separation

r = 0.527; n = 32

(g) Diagnosis

Diagnosis vs. NBM EPM (log10)

Somatic mutations (log10)

r = -0.008; P = 0.955

(h) Relapse

Relapse vs. NBM EPM (log10)

Somatic mutations (log10)

r = 0.238; P = 0.103
**Diagnosis-specific**

**Shared**

**Relapse-specific**

Cluster 1
Cluster 2
Cluster 3

**Cluster 1**

**Cluster 2**

**Cluster 3**

**Color Key:**
- Proportion of eloci that are Diagnosis specific, Shared, or Relapse specific.

**FAB classification Cluster**

**Cluster**

**M1**

**M2**

**M3**

**M4**

**M5**

**Blast purity post separation**

**p-value = 0.736; n = 36**

**p-value = 0.441; n = 138**

**p-values >= 0.398; n = 56**

**Eloci cluster proportion**

**P-value = 0.00774; n = 48**

**Somatic mutations that are Diagnosis specific, Shared, or Relapse specific**

**Color Key:**
- Shared
- Relapse-specific
- Diagnosis-specific
- Not detected
- Not covered

**Nature Medicine: doi:10.1038/nm.4125**
Cluster Proportion of patients

decreasing increasing stable

P = 0.915

Nature Medicine: doi:10.1038/nm.4125
Supplementary Fig. 10

(a) Intra-tumour methylation heterogeneity

(b) Relapse-free probability

Intra-tumour methylation heterogeneity

Stage

Relapse-free probability

Years

P = 0.168

Nature Medicine: doi:10.1038/nm.4125