Corresponding author(s): Shannon McWeeney, Brian Druker

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using Illumina Real Time Analysis (RTA) 1.18; Data was assembled to fastq files using Illumina Bcl2Fastq2 v2.xx

| Data analysis | Described in detail in the Methods; BWA MEM version 0.7.10-r789; Genome Analysis Toolkit (v3.3); bundled Picard (v1.120.1579); The files contained within the Broad's bundle 2.8 were used including their version of the build 37 human genome (These files were downloaded from: ftp://ftp.broadinstitute.org/bundle/2.8/b37/); • The SAM files were sorted and converted to BAM via SortSam; MarkDuplicates was run, marking both lane level standard and optical duplicates; The reads were realigned around indels from the reads--RealignerTargetCreator/IndelRealigner; Base Quality Score Recalibration; The resulting BAM files were then aggregated by sample and an additional round of MarkDuplicates was carried out at the sample level; Quality control reports were generated using the ReportingTools and qrqc; Mutect v1.1.7; Varscan2 v2.4.1; Indels were produced using Varscan2; Variant Effect Predictor v83 against GRCh37; vcf2maf v1.6.6 tool; Mutect v1.1.7; Varscan2 v2.4.1; Ensembl VEP v83 on GRCh37; vcf2maf (v1.5.0) program; vcf2maf suite; CrossMap; Pindel; UnifiedGenotyper; subjunc aligner (1.5.0-p2; featureCounts (1.5.0-p2); the GRCh37 build provided by the Broad as part of the GATK bundle was used; subjunc -i /path/to/reference/ -u -r fastq1 -R fastq2 -o outputBAMFilename -I 5 -T 7 -d 50 -D 600 -S fr featureCounts -a Homo_sapiens.GRCh37.75.gtf -o output -F GTF -t exon -g gene_id -s 2 -C -T 10 -p -P -d 50 -D 600 -B BAM_files; weighted gene correlation network analysis (WGNCA); conditional quantile normalization procedure; edgeR; RNA-sequencing genotyping protocol (as of GATK v3.3); SplitNCigarReads; RealignerTargetCreator/IndelRealigner; TopHat-Fusion (v2.0.14); 'bicor' correlation setting the max proportion of outliers to .1; dynamicTreeCut; SNPRelate; ComplexHeatmap R package; limma users manual; limma-trend approach; fSVA; Mclust; clusplot; DISCOVER; corrplot; lasso approach; glmnet |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw and processed sequencing data, along with relevant clinical annotations are submitted to dbGaP and Genomic Data Commons. The raw data for clinical annotations, variant calls, gene expression counts, and drug sensitivity that underlie all figures in this manuscript are found in the Supplementary Information. In addition, all data can be accessed and queried through our online, interactive user interface, Vizome, at www.vizome.org.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The sample size was chosen to assure sufficient statistical power to capture genetic events observed at 1% or greater disease frequency in AML. |
| Data exclusions | Algorithms, filtering, and curation for variant calling, gene expression counts, and drug sensitivity data are described in detail in the Methods; Mutect v1.1.7 was run using default parameters except that no limit was placed on the number or frequency of the alternative allele frequency in the normal to help address normal contamination; Varscan2 v2.4.1 was run in somatic mode with the recommended filtering scheme except as shown in the table below
Parameter Current
Initial Calls
Min coverage 3
Min variant Frequency .08
Het P-value .1
Somatic Calls
Min tumor frequency .08
Max normal frequency 1
High confidence P-value .1
Post Processing
Max variant avgrl 0
Max reference avgrl 0
Indels and SNVs were produced for the tumor-only samples again using Mutect without a specified normal for consistency and VarScan2 in mpileup2indel or mpileup2snp mode respectively; These variants were assigned to their most deleterious effect on Ensembl transcripts using Ensembl VEP v83 on GRCh37. This assignment was done using the same VEP parameters as the vcf2maf (v1.5.0) program; Using the runs from MuTect and VarScan, these data were next filtered to keep only the protein impacting SNVs and indels from Mutect and VarScan2 and filtered requiring that the variants had at least 5 reads and either not be seen in the exome aggregation consortium (ExAC) or be seen at a frequency < .01. These data present several additional challenges. First somatic calls cannot be obtained directly from the tumor-only samples, second there is always a possibility of tumor contamination of the skin samples for those samples that were paired. To address |

these issues and maximize comparability we used an iterative approach. The following was done separately for the two genotypers:
• An initial set of higher confidence somatic mutations were retrieved from the paired tumor/normal samples requiring tumor variant allele frequency (VAF) >= 8% and normal VAF <= 5% in addition to the significance tests already performed by the programs.
• A list of all candidate mutations was collated requiring that a mutation was either seen in the high confidence somatic set, the set of variants from Jaiswal et al or from the lifted over set of variants from the TCGA AML paper.
• Mutations from the overall set were kept if:
o The overall number of calls in the paired samples was not more than twice the number of high confidence somatic calls
o The tumor-only frequency for the calls was less than 50% greater than the number of calls in the paired samples
o The mutation was seen in Jaiswal/TCGA list
• High Confidence somatic mutations were kept regardless
The data from the two genotypers were combined along with FLT3-ITD calls from Pindel. Comparing our variant lists from whole exome sequencing versus our custom capture validation sequencing, we noticed, similar to others, that low allele frequency C->A variants (< 15%) tended to have poor concordance (7.7%; data not shown) between the initial run and the technical validation run. These variants were removed in these data and along with a curated 'blacklist' of known problematic variants/genes including mitochondrial DNA variants. In addition, all variants that were seen in a cumulative list of BeatAML normal samples at a frequency greater than 1% were removed. Cumulatively, of this set, 94% of covered single-nucleotide variants were validated with 82% of insertion/deletion calls also being confirmed with validation sequencing.
Manual review was then carried out in the following steps:
a.) The addition back of all Jaiswal flagged rows.
b.) Reviewed all TCGA flagged rows for VAF pattern that matched or did not match with known drivers in same specimen. Some TCGA variants were added back based on convincing VAF pattern and known pathogenic role, other TCGA variants were kept excluded based on VAF pattern unlike known drivers in same specimens.
c.) Other variants were added back based on other specimens that had the same variant that was still on the include list and VAF pattern looked convincing for inclusion.
d.) All Jaiswal genes with only frameshift/nonsense variants were manually reviewed and missense mutations were manually removed.
e.) Genes/variants that were on both the include and exclude lists were manually reviewed and were removed if c to a with over 15% VAF, did not validate, and/or VAF pattern unlike known drivers in same specimen
f.) Further review of all genes in summary sheet with cohort frequency of 8 or more (1% of more). Any that were not familiar from knowledge of AML literature were manually reviewed for VAF patterns that did or did not match known drivers within same specimens. Those that did not match were manually removed.
After this manual review, additional curated mutations from the UnifiedGenotyper run were added back in along with a curated set of variants from tumor-only patients in Jaiswal et al genes; Gene expression count data were collated from featureCounts matrices and all genes with no counts across the samples were excluded. Genes with duplicate gene symbols and those where the counts were < 10 for 90% or more of the samples were additionally removed prior to normalization similar to the approach suggested for weighted gene correlation network analysis (WGNCA). Samples for which their median expression was less than 2 standard deviations below the average were removed from the dataset (N=10); Quality Control
The UnifiedGenotyper runs for both the WES and RNA-sequencing were combined into a single VCF file using the GATK CombineVCFs functionality. This combined VCF file was converted to a GDS file using SNPRelate (1.12.2). Note the version is an upper bound as several versions were used across the entire project). The overall similarity of the genotypes of each pair of samples were computed, termed identity by state (IBS) and a hierarchical clustering was performed using one minus this similarity. From this clustering and visualization we had devised hard cutoffs for further inspection based on the types of data being compared. For instance samples not meeting the specified IBS thresholds (DNA-DNA=.9; RNA-RNA=.83; DNA-RNA=.89) were subject to manual review. Based on the dendrogram structure as well as the clinical/lab information, samples were either excluded, assigned to a different patient ID or in rare cases assigned to a different sample. It was observed that bone marrow transplants between sample collections produced a noticeable but milder effect in these dendrograms and such samples were flagged for removal in RNA-sequencing analysis and for treatment as tumor-only samples in the WES analysis as is described in the 'WES Variant Detection' section; Ex vivo Functional Drug Screen Data Processing -- A given sample was run on one or more panels and within each panel, the majority of drugs were run without within-panel replicates. Two steps were performed to harmonize these data prior to model fitting:
1. A 'curve-free' AUC (integration based on fine linear interpolation between the 7 data points themselves) was calculated for those runs with within-panel replicates after applying a ceiling of 100 and a floor of 0 for the normalized viability. The maximum change in AUC amongst the replicates was noted and those runs with differences > 100 were removed.
2. Remaining within-plate replicates had their normalized viability averaged and subject to a ceiling of 100 and floor of 0. An additional set of 'curve-free' AUCs was computed for sample-inhibitor pairs run on multiple panels. The maximum change in AUC amongst the across-panel replicates was noted and those runs with differences > 75 were removed.
At this point, the within and across plate replicates for the normalized viability were averaged together and a ceiling of 100 was applied. From the steps above, the floor was already at 0.
Based on the methodology used in our prior drug combination study{Kurtz, 2017 #917}, a probit regression was fit to all possible run groups using the model:
(normalized_viability / 100) ~ 1 + log10(concentration)
Where for all groups there were N=7 dose-response measurements.
The summary measures of curve fit were inspected and cutoffs were devised removing all runs with an AIC > 12 and deviance > 2. For inhibitors that were run using multiple concentration ranges, only the latest concentration range was kept. Finally, these data were compared to the AUC values from third order polynomial fits. Those runs that were discrepant in terms of sensitive/resistant calls were manually reviewed as subject to removal; Co-occurrence/mutual exclusivity -- Only mutations seen in at least 10 patients were kept. The DISCOVER method was used to determine significant mutual exclusivity and co-occurrences. A plot of the co-occurrences was generated using corrplot with the odds ratio of the pairwise co-occurrence used to color and scale the circle sizes; Sensitive/Resistant differential expression -- For each drug, it was required that at least 3 sensitive and 3 resistant samples using the 20%/20% criteria outlined in the 'Drug Analysis' section; Integration of both mutation and RNA-Sequencing with Ex vivo Functional Drug Screen -- Mutations (0/1 encoding) and the module eigengenes from the WGCNA analysis were used separately and combined together in regression models with coefficients selected using the lasso approach as implemented in glmnet. For each datatype and the combination, only drugs with at least 200 patients samples were tested. The 3 datasets were initially randomly separated into training (75%) and test (25%) sets. Similar to a previous approach, a bootstrap aggregation approach was used where the 1,000 bootstraps of the training dataset was generated and for each one, the lasso trained using 10 fold cross-validation. Predictions were formed for the test dataset over these bootstrap models and the predicted AUC was averaged. R2 values were computed for these aggregated predictions relative to the test AUC values. As performance was seen to be dependent on the initial test/training split, we repeated the entire process 100 times, recording the mean and standard deviation of the R2 value as well as the

count non-zero coefficients
All of these criteria were pre-established prior to execution of the analyses.

Replication | All data analysis pipelines filtering, exclusions and quality control steps are described above and in detail in the Methods. Each analytical approach and result was replicated successfully.

Randomization | All samples were assigned numerical identifications with no association to any features of the sample, and for all sequencing batches samples were randomized into capture library groups and flow cells.

Blinding | All samples were assigned numerical identifications that bore no relevance to sample features or attributes, and all data analyses were performed using these de-identified specimen ID numbers

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	Unique biological materials
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology
☒	Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics | This is all documented in great detail in the Methods and in Supplementary Information. Below is a list of covariate population characteristics:
labId
patientId
consensus_sex
inferred_sex
inferred_ethnicity
centerID
CEBPA_Biallelic
ageAtDiagnosis
isRelapse
isDenovo
isTransformed
finalFusion
specificDxAtAcquisition_MDSMPN
nonAML_MDSMPN_specificDxAtAcquisition
priorMalignancyNonMyeloid
priorMalignancyType
cumulativeChemo
priorMalignancyRadiationTx
priorMDS
priorMDSMoreThanTwoMths
priorMDSMPN
priorMDSMPNMoreThanTwoMths
priorMPN
priorMPNMoreThanTwoMths
dxAtInclusion
specificDxAtInclusion
ELN2017
ELN2008
dxAtSpecimenAcquisition
specificDxAtAcquisition
ageAtSpecimenAcquisition
timeOfSampleCollectionRelativeToInclusion
specimenGroups
specimenType
rnaSeq
exomeSeq
totalDrug

rnaSeqAnalysis
analysisExomeSeq
analysisDrug
cumulativeTreatmentTypeCount
cumulativeTreatmentTypes
cumulativeTreatmentRegimenCount
cumulativeTreatmentRegimens
cumulativeTreatmentStageCount
cumulativeTreatmentStages
responseToInductionTx
typeInductionTx
responseDurationToInductionTx
mostRecentTreatmentType
currentRegimen
currentStage
mostRecentTreatmentDuration
vitalStatus
overallSurvival
causeOfDeath
any_different_labs
any_different_labs_also_beataml
different_lab_ids
different_id_karyotype_interval
%.Basophils.in.PB
%.Blasts.in.BM
%.Blasts.in.PB
%.Eosinophils.in.PB
%.Immature.Granulocytes.in.PB
%.Lymphocytes.in.PB
%.Monocytes.in.PB
%.Neutrophils.in.PB
%.Nucleated.RBCs.in.PB
ALT
AST
Albumin
Creatinine
FAB/Blast.Morphology
Hematocrit
Hemoglobin
Karyotype
LDH
MCV
Other.Cytogenetics
Platelet.Count
Surface.Antigens.(Immunohistochemical.Stains)
Total.Protein
WBC.Count
any_different_cgs
any_different_cgs_also_beataml
different_cgs_lab_ids
FLT3-ITD
NPM1
ABL1
ASXL1
ASXL2
ATM
BCOR
BCORL1
BRAF
BRCA2
CALR
CBL
CCND2
CCND3
CD36
CEBPA
CHEK2
CIITA
CREBBP
CSF3R
CTCF
CUX1
DNMT3A
EP300
ETV6

EZH2
FBXW7
FLT3
GATA1
GATA2
IDH1
IDH2
IKZF1
JAK1
JAK2
JAK3
KDM6A
KIT
KMT2A
KMT2D
KRAS
MEN1
MPL
MUTYH
MYD88
NF1
NOTCH1
NRAS
PAX5
PDGFRB
PHF6
POT1
PRDM1
PTPN11
RAD21
ROS1
RUNX1
SETBP1
SF3B1
SMC1A
SOCS1
SRSF2
STAG2
STAT3
SUZ12
TCL1A
TET2
TP53
TYK2
U2AF1
WT1
ZRSR2

Recruitment | All patients with a diagnosis of acute myeloid leukemia at any of the partner institutions were eligible for and consented for the study. No exclusionary criteria existed.