

Functional genomic landscape of acute myeloid leukaemia

Jeffrey W. Tyner^{1,2}, Cristina E. Tognon^{2,3,4}, Daniel Bottomly^{2,5}, Beth Wilmot^{2,5,6}, Stephen E. Kurtz^{2,3}, Samantha L. Savage^{1,2}, Nicola Long^{2,3}, Anna Reister Schultz^{1,2}, Elie Traer^{2,3}, Melissa Abel^{1,2}, Anupriya Agarwal^{2,7}, Aurora Blucher^{2,5}, Uma Borate^{2,3}, Jade Bryant^{1,2}, Russell Burke^{2,3}, Amy Carlos^{2,8}, Richie Carpenter^{2,3}, Joseph Carroll^{2,9}, Bill H. Chang^{2,10}, Cody Coblentz^{2,3}, Amanda d'Almeida^{1,2}, Rachel Cook^{2,3}, Alexey Danilov^{2,3}, Kim-Hien T. Dao^{2,3}, Michie Degnin^{2,3}, Deirdre Devine^{2,3}, James Dibb^{2,3}, David K. Edwards^{1,2}, Christopher A. Eide^{2,3,4}, Isabel English^{2,3}, Jason Glover^{2,10}, Rachel Henson^{2,8}, Hiberny Ho^{2,3}, Abdusebur Jemal^{2,10}, Kara Johnson^{2,3}, Ryan Johnson^{2,3}, Brian Junio^{2,3}, Andy Kaempf^{2,11}, Jessica Leonard^{2,3}, Chenwei Lin^{2,8}, Selina Qiuying Liu^{2,3}, Pierrette Lo^{2,3}, Marc M. Loriaux^{2,12}, Samuel Luty^{2,3}, Tara Macey^{2,3}, Jason MacManiman^{1,2}, Jacqueline Martinez^{1,2}, Motomi Mori^{2,11,13}, Dylan Nelson¹⁴, Ceilidh Nichols^{2,3}, Jill Peters^{2,3}, Justin Ramsdill^{5,6}, Angela Rofelty^{1,2}, Robert Schuff^{5,6}, Robert Searles^{2,8}, Erik Segerdell^{2,5}, Rebecca L. Smith^{2,3}, Stephen E. Spurgeon^{2,3}, Tyler Sweeney^{2,3}, Aashis Thapa^{2,3}, Corinne Visser^{2,3}, Jake Wagner^{2,3}, Kevin Watanabe-Smith^{2,3}, Kristen Werth^{2,3}, Joelle Wolf^{2,10}, Libbey White^{2,5}, Amy Yates^{5,6}, Haijiao Zhang^{1,2}, Christopher R. Cogle¹⁵, Robert H. Collins¹⁶, Denise C. Connolly^{17,18}, Michael W. Deininger¹⁹, Leylah Drusbosky¹⁵, Christopher S. Hourigan²⁰, Craig T. Jordan²¹, Patricia Kropf²², Tara L. Lin²³, Micaela E. Martinez²⁴, Bruno C. Medeiros²⁵, Rachel R. Pallapati²⁴, Daniel A. Pollyea²¹, Ronan T. Swords²⁶, Justin M. Watts²⁶, Scott J. Weir^{27,28}, David L. Wiest²⁹, Ryan M. Winters¹⁸, Shannon K. McWeeney^{2,5,6*} & Brian J. Druker^{2,3,4*}

The implementation of targeted therapies for acute myeloid leukaemia (AML) has been challenging because of the complex mutational patterns within and across patients as well as a dearth of pharmacologic agents for most mutational events. Here we report initial findings from the Beat AML programme on a cohort of 672 tumour specimens collected from 562 patients. We assessed these specimens using whole-exome sequencing, RNA sequencing and analyses of ex vivo drug sensitivity. Our data reveal mutational events that have not previously been detected in AML. We show that the response to drugs is associated with mutational status, including instances of drug sensitivity that are specific to combinatorial mutational events. Integration with RNA sequencing also revealed gene expression signatures, which predict a role for specific gene networks in the drug response. Collectively, we have generated a dataset—accessible through the Beat AML data viewer (Vizome)—that can be leveraged to address clinical, genomic, transcriptomic and functional analyses of the biology of AML.

Approximately 21,000 people are diagnosed with AML and over 10,000 AML-related deaths are reported annually in the United States^{1,2}. Cytogenetic and sequencing analyses have revealed at least 11 genetic classes of AML³ and over 20 subsets can be assigned when also considering cell differentiation states of the leukaemic blasts^{4,5}. Deep sequencing of AML by The Cancer Genome Atlas (TCGA) revealed a heterogeneous disease with nearly 2,000 somatically mutated genes observed across 200 patients⁶. Many of the recurrent cytogenetic events and somatic mutations have been shown to carry prognostic importance^{3,7,8}. Some of the most frequent somatic variants can also be observed in myelodysplastic syndromes and myeloproliferative neoplasms^{9–11} that can transform into AML. These same mutations have

also been observed in healthy individuals who have age-related clonal haematopoiesis, which is associated with significant risk for the development of myelodysplastic syndromes, myeloproliferative neoplasms and AML^{12–15}.

A small number of therapies targeted to mutational events have been developed for patients with AML, although the current standard of care remains largely unchanged over the past 30–40 years. The first targeted therapy for AML involved use of all-*trans* retinoic acid in combination with arsenic trioxide for patients with rearrangement of the retinoic acid receptor^{16,17}. More recently, fms-related tyrosine kinase 3 (FLT3) inhibitors have been developed for FLT3 mutational events that occur in approximately 20–30% of patients with AML^{18–21}.

¹Department of Cell, Developmental & Cancer Biology, Oregon Health & Science University, Portland, OR, USA. ²Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. ³Division of Hematology & Medical Oncology, Department of Medicine, Oregon Health & Science University, Portland, OR, USA. ⁴Howard Hughes Medical Institute, Portland, OR, USA. ⁵Division of Bioinformatics and Computational Biology, Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA. ⁶Oregon Clinical & Translational Research Institute, Oregon Health & Science University, Portland, OR, USA. ⁷Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR, USA. ⁸Integrated Genomics Laboratories, Oregon Health & Science University, Portland, OR, USA. ⁹Technology Transfer & Business Development, Oregon Health & Science University, Portland, OR, USA. ¹⁰Division of Hematology and Oncology, Department of Pediatrics, Oregon Health & Science University, Portland, OR, USA. ¹¹Biostatistics Shared Resource, Oregon Health & Science University, Portland, OR, USA. ¹²Department of Pathology, Oregon Health & Science University, Portland, OR, USA. ¹³Oregon Health & Science University-Portland State University School of Public Health, Portland, OR, USA. ¹⁴High-Throughput Screening Services Laboratory, Oregon State University, Corvallis, OR, USA. ¹⁵Department of Medicine, Division of Hematology and Oncology, University of Florida, Gainesville, FL, USA. ¹⁶Department of Internal Medicine/Hematology Oncology, University of Texas Southwestern Medical Center, Dallas, TX, USA. ¹⁷Molecular Therapeutics Program, Fox Chase Cancer Center, Philadelphia, PA, USA. ¹⁸Fox Chase Cancer Center Biosample Repository Facility, Philadelphia, PA, USA. ¹⁹Division of Hematology & Hematologic Malignancies, Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA. ²⁰National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA. ²¹Division of Hematology, University of Colorado, Denver, CO, USA. ²²Bone Marrow Transplant Program, Fox Chase Cancer Center, Philadelphia, PA, USA. ²³Division of Hematologic Malignancies & Cellular Therapeutics, University of Kansas, Kansas City, KS, USA. ²⁴Clinical Research Services, University of Miami Sylvester Comprehensive Cancer Center, Miami, FL, USA. ²⁵Department of Medicine—Hematology, Stanford University, Stanford, CA, USA. ²⁶Department of Hematology, University of Miami Sylvester Comprehensive Cancer Center, Miami, FL, USA. ²⁷Department of Toxicology, Pharmacology and Therapeutics, University of Kansas Medical Center, Kansas City, KS, USA. ²⁸Department of Medicine, Division of Medical Oncology, University of Kansas Medical Center, Kansas City, KS, USA. ²⁹Blood Cell Development and Function Program, Fox Chase Cancer Center, Philadelphia, PA, USA. *e-mail: mcweeney@ohsu.edu; drukerb@ohsu.edu

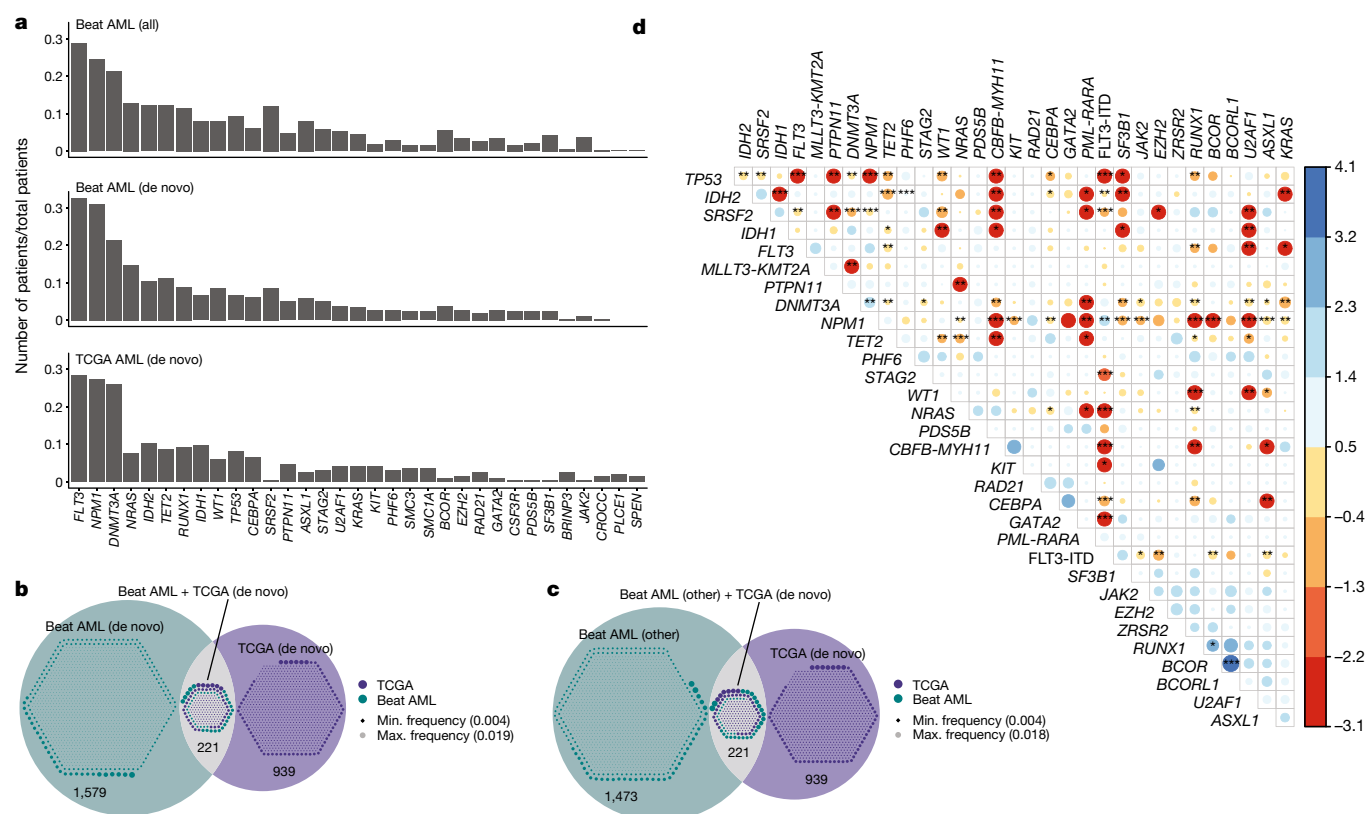


Fig. 1 | Comparative genomic landscape of AML. **a**, Frequency of the 33 mutational events that were cumulatively the most frequent in Beat AML ($n = 531$ patients) and TCGA ($n = 200$ patients) datasets. Top, the full Beat AML cohort; middle, only the de novo Beat AML cases; bottom, de novo cases in the TCGA. Mutations were summarized by gene as was done by TCGA, whereas the FLT3-ITD mutations were kept separate in the rest of this manuscript. **b**, Mutational events at 2% frequency or less in the de novo cases of Beat AML and TCGA were compared for overlap. Venn diagram displays the overlap with the small circles within each

compartment representing a size-scaled frequency of each mutational event. **c**, Analysis as in **b** with only the non-de novo Beat AML cases versus TCGA. **d**, Co-occurrence or exclusivity of the most recurrent mutational events in the Beat AML cohort ($n = 531$ patients) were assessed using the DISCOVER⁴¹ method. The dot plot shows the odds ratio of co-occurrence (blue) or exclusivity (red) using colour-coding and circle size as well as asterisks that indicate FDR-corrected statistical significance. $*P < 0.1$; $**P < 0.05$; $***P < 0.01$.

When FLT3 inhibitors were used as single agents, responses of only 2–6 months were obtained^{22–25}. Midostaurin, a broad-spectrum FLT3 inhibitor, was recently approved for use in newly diagnosed patients with AML with FLT3 mutations, in combination with standard of care chemotherapy²⁶; however, relapse was still prevalent in this setting. Targeting of a mutant form of isocitrate dehydrogenase (NADP(+)) 1 and 2, cytosolic (IDH1 and IDH2)²⁷, has shown clinical benefit leading to approval of the IDH2 inhibitor, enasidenib, and the IDH1 inhibitor, ivosidenib^{28,29}. Additional proposed strategies have included inhibition of epigenetic modifiers, such as enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2)³⁰, lysine demethylase 1A (KDM1A)³¹, and DOT1-like histone lysine methyltransferase (DOT1L)³², based on the direct mutation of these targets or synthetic lethality in the context of drug combinations (all-*trans* retinoic acid and KDM1A inhibitors) or specific genetic features (lysine methyltransferase 2A (KMT2A)-gene rearrangement for DOT1L inhibitors). Hypomethylating agents have been used in patients with AML, for which better responses have been reported for certain genetic subsets, such as those with mutations in TET2³³ or tumour protein 53 (TP53)³⁴. Most recently, an inhibitor of the BCL2 apoptosis regulator (BCL2), venetoclax, showed an approximately 20% response rate when used as a single agent in patients with a relapse³⁵ and higher response rates (around 60%) were reported in combination with hypomethylating agents in newly diagnosed, elderly patients with AML³⁶.

Comparative genomic landscape of AML

To better understand genetic or transcriptional markers and mechanisms of drug sensitivity and resistance in AML, we developed a cohort

of 672 primary specimens from 562 patients with AML and we performed extensive functional and genomic analyses on these samples. Detailed clinical annotations, including diagnostic information, clinical laboratory values, treatments, responses and outcomes were curated from electronic medical records and are reported in Supplementary Tables 1–5.

We performed exome sequencing on 622 of the specimens from the cohort representing 531 different patients. The final, high-confidence variant list (Supplementary Table 7) revealed a range of 1–80 somatic variants per patient (cohort median of 13 somatic variants) (Extended Data Fig. 1). Comparison of the top 33 most commonly mutated genes across Beat AML and TCGA⁶ showed generally similar frequencies. Higher frequency of mutations in serine- and arginine-rich splicing factor 2 (SRSF2) were seen in Beat AML than in TCGA, and this difference was conserved when only the de novo cases in Beat AML and TCGA were compared (Fig. 1a). By contrast, mutational events that were seen with a frequency of less than 2% across Beat AML and TCGA were much more divergent; variants in 221 mutant genes were called in both datasets, 939 mutant genes were called only in TCGA and around 1,500 mutant genes were called only in Beat AML, irrespective of whether we compared only de novo or non-de novo cases (Fig. 1b, c). Most of these divergent mutational events were observed only in single patients; however, there were mutations in 11 genes that were called in 1% or more of patients in Beat AML, but were not observed in previous AML sequencing studies (Extended Data Fig. 1). Finally, co-occurrence and exclusivity of the most frequent variants were computed and reveal significant patterns of mutational co-segregation, suggesting biological cooperation between certain mutational events (Fig. 1d).

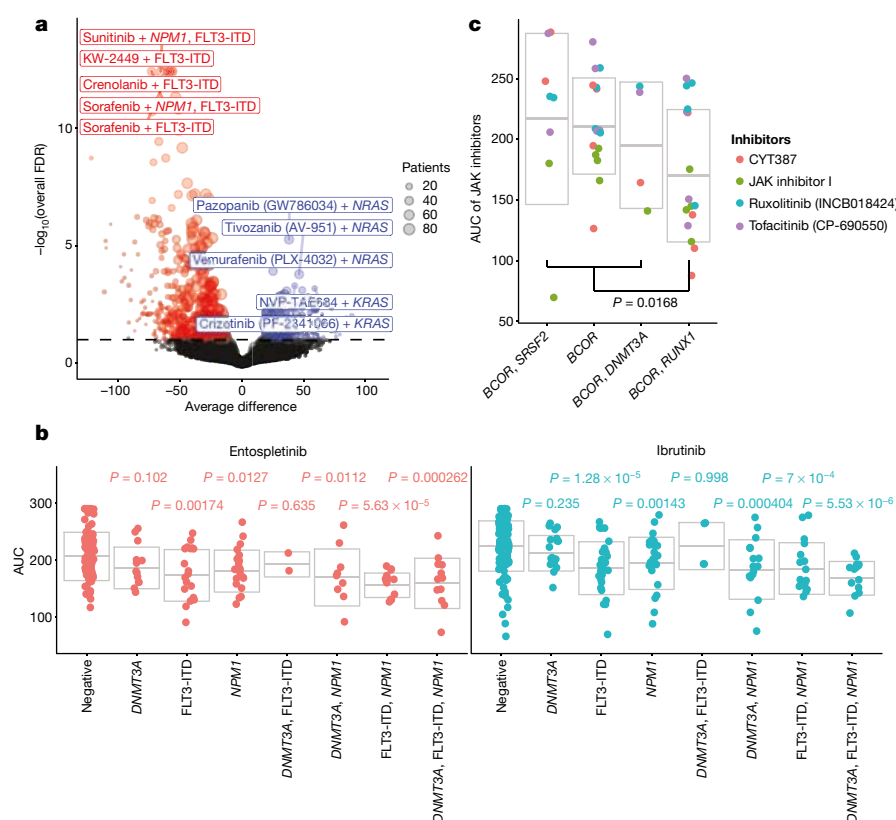


Fig. 2 | Integration of genetic events with drug sensitivity. **a**, Average difference in AUC drug response between mutant and wild-type cases was determined using a Student's two-sided *t*-test from a linear model fit (*x* axis). The *P* values were corrected using the Benjamini–Hochberg method over all the drugs (*y* axis). The number of samples used to correlate each mutational event with drug sensitivity is reported in the Supplementary Table 17. Expanded and interactive plots are available in our online data browser (<http://www.vizome.org/> and http://vizome.org/additional_figures_BeatAML.html). **b**, AUCs for ibrutinib or entospletinib ($n = 277$ or 168 patient samples, respectively) were plotted for cases with single, double or triple mutations in *NPM1* and *DNMT3A* as well as

FLT3-ITD. Data are mean \pm s.d. An ANOVA was conducted using the Bonferroni approach (statistical results and sample size for all groups are reported in Supplementary Tables 18, 19). **c**, Inhibitors of JAK family kinases were assessed for activity against cases with *BCOR* mutations alone or *BCOR* mutations in combination with mutations in *SRSF2*, *RUNX1* or *DNMT3A*. The AUC values are plotted per case; data are mean \pm s.d. There was a significant difference in AUC; two-sided Student's *t*-test ($t_{42} = -2.489$, $P = 0.0168$, 95% confidence interval -73.018 to -7.643) between mutations in *BCOR* and *RUNX1* ($n = 16$) versus the average of *BCOR* mutations alone ($n = 16$), mutations in *BCOR* and *SRSF2* ($n = 8$), and mutations in *BCOR* and *DNMT3A* ($n = 4$).

Drug response and gene expression

RNA sequencing was performed on 451 specimens from 411 patients. Clustering of the 2,000 most variably expressed genes across the cohort revealed gene expression signatures that were associated with many of the prominent genetic and cytogenetic disease groups (Extended Data Fig. 2). To understand the profile of sensitivity and resistance to a variety of small-molecule inhibitors, we profiled primary tumour cells from 409 specimens derived from 363 patients against a panel of 122 small-molecule inhibitors using an ex vivo drug sensitivity assay³⁷. Drug sensitivity patterns were analysed with respect to clinical and genetic features of tumours (Extended Data Fig. 3). We compared the average area-under-the-curve (AUC) values for each drug between samples from cases with de novo AML and cases with AML that had transformed from myelodysplastic syndromes or myeloproliferative neoplasms, through a series of single-factor analysis of variance (ANOVA) tests. Generally, transformed cases showed less sensitivity than de novo cases to most drugs. Of the 122 drugs tested, 64 were significantly (false discovery rate (FDR) < 0.1) more sensitive in the de novo samples, whereas only one drug—panobinostat (an HDAC inhibitor)—was significantly more sensitive in the transformed cases (Extended Data Fig. 4). In addition, we analysed the concordance of drug sensitivity patterns with respect to predicted target gene, gene family or pathway for each drug (drug assignments to target families can be found in Supplementary Table 11). This analysis revealed drug targets and/or drug families that were highly concordant among constituent members, as well as drug families that were quite discordant

(Extended Data Fig. 5). To create a global view of overall sensitivity or resistance for each case, we generated a heat map of binary sensitive or resistant calls for each sample to each drug. We then annotated the sensitive or resistant fraction of each case against the European Leukaemia Net (ELN) 2017⁵ (Extended Data Fig. 6a) and WHO (World Health Organization) 2016⁴ (Extended Data Fig. 6b) classifications.

Gene signatures of drug responses

We performed a cohort analysis to assess the correlation between drug sensitivity patterns and mutational events or gene expression levels. Correlation analyses between drug sensitivity and mutational events were performed by assessing the range of sensitivity of cases with a mutation in an individual gene (as well as co-occurring mutational events) versus cases with the wild-type sequence for that same gene. Broad summaries of full cohort results are displayed in Circos and Manhattan plots (Extended Data Figs. 7a, 8). Individual associations between drugs and mutations are displayed as a Volcano plot, in which the differences in drug sensitivity between mutant and wild-type genes and the FDR-corrected significance values³⁸ are plotted (Fig. 2a). Some of the associations with the highest levels of statistical significance involved FLT3 internal tandem duplications (FLT3-ITD) and these showed sensitivity to FLT3 pathway inhibitors, which serves as a proof-of-principle as FLT3 inhibitors are known to be more effective against FLT3-ITD AML. However, to reveal associations between drugs and mutations that were not biased by the co-occurrence with FLT3-ITD, we also plotted the same analysis using only cases that had

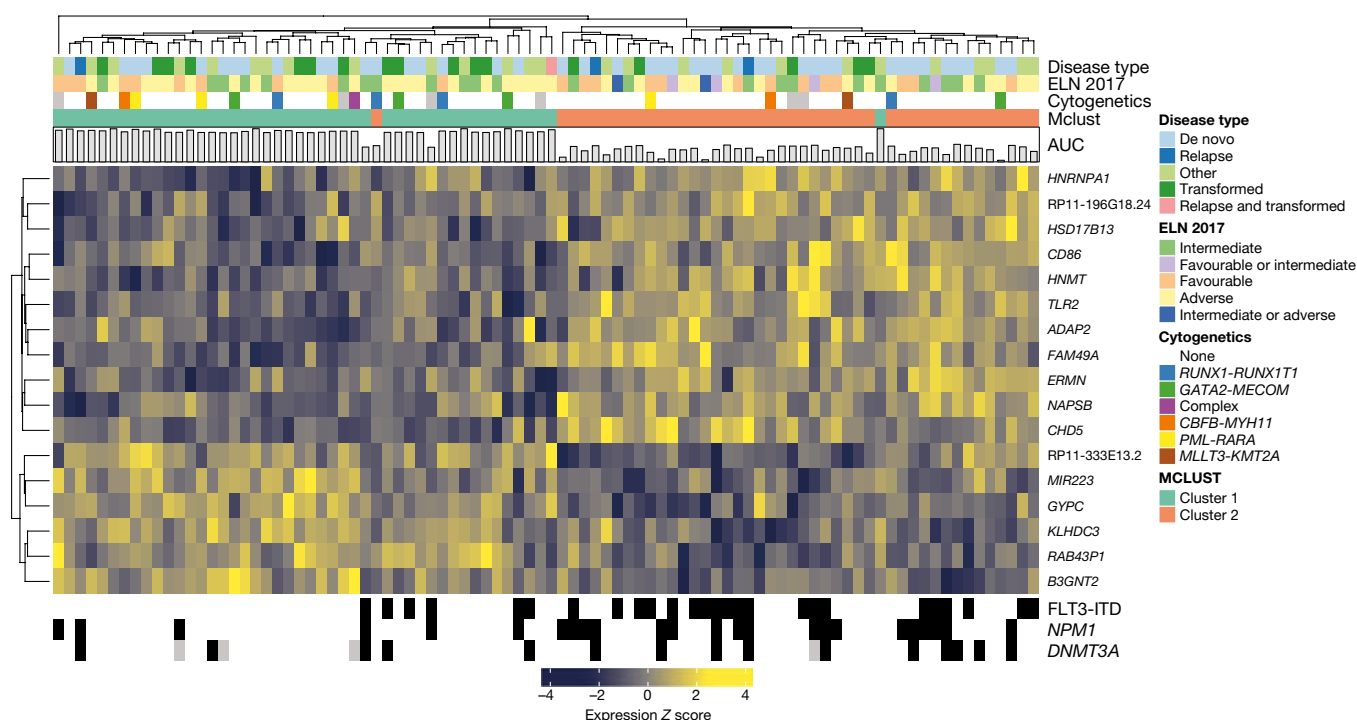


Fig. 3 | Integration of gene expression and drug sensitivity patterns. Differential gene expression signature distinguishing the 20% most ibritinib-sensitive ($n = 46$) from the 20% most resistant ($n = 44$) specimens. Heat maps for all other drugs are available in our online data

browser (http://vizome.org/additional_figures_BeatAML.html). For the number of samples used to correlate each drug with gene expression, see Supplementary Table 17. RP11-333E13.2 is a lincRNA; RP11-196G18.24 is a pseudogene.

wild-type *FLT3* (Extended Data Fig. 7b). We also created Volcano plots that were specific to each individual drug (versus all mutational events) and for each individual gene (versus all drugs tested). All Volcano plots can be found with interactive features in our online data browser (<http://www.vizome.org/> and http://vizome.org/additional_figures_BeatAML.html).

Mutations in several genes, most notably *TP53* or ASXL transcriptional regulator 1 (*ASXL1*), were shown to cause a broad pattern of drug resistance. Notably, a few drugs trended to be more sensitive to cases with *TP53* mutations (such as elesclamol) or with mutations in *ASXL1* (such as panobinostat), suggesting candidates for further exploration of cases with AML that have these poor prognostic features. Mutations in the NRAS proto-oncogene, GTPase (*NRAS*) or KRAS proto-oncogene, GTPase (*KRAS*) also correlated with resistance to most drugs, although mutations in these genes did show the predicted sensitivity to MAPK inhibitors. Of particular note, there was a stronger association between mutations in *NRAS* and MAPK sensitivity than between mutations in *KRAS* and MAPK sensitivity. *IDH2* mutations conferred sensitivity to a broad spectrum of drugs, whereas mutations in *IDH1* conferred resistance to most drugs. Mutations in *RUNX1* correlated with sensitivity to PIK3C and mTOR inhibitors (such as BEZ235) and to the multi-kinase vascular endothelial growth factor receptor (VEGFR) inhibitor, cediranib. Mutation in the spliceosome components U2 small nuclear RNA auxiliary factor 1 (*U2AF1*) and zinc-finger CCCH-type, RNA-binding motif and serine/arginine rich 2 (*ZRSR2*) correlated with sensitivity to several drugs. The mechanisms that underlie these latter sensitivity correlations (and many others in the dataset) merit further investigation. A significant association was seen between mutations in *FLT3*, *NPM1* and *DNMT3A* and sensitivity to the Food and Drug Administration (FDA)-approved drug, ibritinib. Because these mutations exhibit a significant pattern of co-occurrence, we next examined every combination of single, double or triple mutant genes with respect to ibritinib. We observed that *DNMT3A* alone or *DNMT3A* and *FLT3* double-mutant cases were not significantly different from cases with wild-type genes, whereas cases with *FLT3*-ITD alone or any combination with a mutation in *NPM1* (including cases in which all

three genes were mutated) were significantly more sensitive than cases with wild-type genes (Fig. 2b). Ibrutinib is an inhibitor of BTK and TEC family kinases, although it can exhibit broad off-target effects when maintained in continuous culture with target cells. We noted that another kinase inhibitor with high specificity for spleen-associated tyrosine kinase (SYK)—entospletinib—showed a similarly significant pattern of sensitivity in cases with *FLT3*-ITD and mutations in *NPM1* (Fig. 2b), potentially pointing to an operationally important target for this disease subset. Indeed, previous studies have suggested that SYK is an interacting target of *FLT3*-ITD in AML³⁹. Finally, we performed an additional analysis that leveraged multiple inhibitors with common targets to see whether this approach could identify additional associations. We focused on correlations between the four selective Janus kinase (JAK) inhibitors in our drug panel (mometinib, ruxolitinib, tofacitinib and JAK inhibitor I) and mutations in the *BCL6* corepressor (*BCOR*) alone or mutations in *BCOR* together with mutations in *DNMT3A*, *RUNX1* or *SRSF2*. By plotting the average difference of each JAK inhibitor between mutant and wild-type groups for these four categories and performing a one-way ANOVA of the four groups, we found that mutations in both *BCOR* and *RUNX1* correlated with increased sensitivity to all four JAK kinase inhibitors, whereas *BCOR* mutations alone or mutations in *BCOR* together with mutations in *DNMT3A* or *SRSF2* showed no difference in sensitivity to the JAK kinase inhibitors—although *BCOR* mutations alone did show sensitivity to other drugs, such as the tankyrase/WNT inhibitor XAV-939 and the multi-kinase inhibitor crizotinib (Fig. 2c). Collectively, these data suggest that dysregulation of the JAK pathway may represent a vulnerability within certain settings of specific combinations of mutations and not in others.

We also performed an integrative analysis of drug sensitivity data with respect to patterns of gene expression, comparing the 20% of samples with the lowest AUC versus the 20% with the highest AUC and assessing the most differentially expressed genes between those sample sets. This analysis revealed significant ($FDR < 0.05$) expression signatures for 78 testable drugs in the panel (78 out of 119; 65.5%). As an example, the 20% most and least sensitive cases to ibritinib could be clearly distinguished by an expression signature of 17 genes

(Fig. 3 and Extended Data Fig. 9). Expression-signature heat maps for each drug can be found in our online data browser (http://vizome.org/additional_figures_BeatAML.html).

Finally, to assess the joint contributions of both mutation and system-level co-expression patterns (based on de novo network inference) to predicting drug response, multivariate modelling was performed. This integrated analysis allows us to move beyond the significant associations of single mutations (such as FLT3-ITD and mutations in *NPM1*). We performed a weighted correlation network analysis of RNA-sequencing data that identified 14 sets of genes for which the gene expression patterns showed significant clustering across the cohort (clusters contained both increased and decreased gene expression events). We performed regularized regression modelling (LASSO)⁴⁰ to understand how strongly any mutational event or any of these 14 gene expression clusters correlated with sensitivity or resistance to any of the drugs on the panel. We identified numerous, novel co-occurrences of mutations and expression clusters that were associated with drug sensitivity or resistance, with co-occurrences with ibrutinib shown as an example in Extended Data Fig. 9. For ibrutinib, these co-occurrences included a co-expression cluster of 345 genes (using a colour-labelling scheme and shown in ‘turquoise’) that correlated with drug sensitivity and frequently co-occurred with FLT3-ITD, which also correlated with drug sensitivity. There was significant overlap between this ‘turquoise’ gene expression cluster and the 17-gene signature in Fig. 3 (indicated by the ratio of observed overlap to expected overlap, or the representation factor, which was 13.6; $P < 1.734 \times 10^{-4}$). It is important to note that network analysis of this gene expression cluster highlighted enrichment of a number of immune-related pathways, which was not detected within the 17-gene signature (displayed at http://vizome.org/additional_figures_BeatAML.html). We also identified a 110-gene subnetwork (labelled in ‘magenta’ in Extended Data Fig. 9), which was associated with resistance to ibrutinib and was significantly associated with adverse ELN 2017 risk. To look more broadly at associations between mutations and gene expression clusters, we summarized drugs at family level to assess the frequency with which mutations and gene expression clusters were selected in iterative regression modelling (displayed at http://vizome.org/additional_figures_BeatAML.html).

Discussion

In summary, we report a large functional genomic dataset of primary tumour biopsies. We present a cohort of specimens from patients with AML for which we have performed detailed clinical annotations, genomic and transcriptomic analyses and ex vivo drug sensitivity studies, and we provide analytical approaches for data integration. Each of these datasets alone has revealed new information about the biology and potential translational approaches in AML, and the integration of these datasets has revealed new markers and mechanisms of drug sensitivity and resistance that merit further study. These data have all been made publicly available through the NIH/NCI dbGaP and Genomic Data Commons (GDC) resources, and we have developed tools to facilitate user-interfacing with the dataset (<http://www.vizome.org/>). We hope and expect that this public data release will stimulate further use of the data, such that novel findings can be derived and turned into new clinical approaches for treatment of AML.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0623-z>.

Received: 4 April 2018; Accepted: 14 August 2018;
Published online 17 October 2018.

1. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J. Clin.* **60**, 277–300 (2010).
2. SEER. Cancer stat facts: leukemia — acute myeloid leukemia (AML). *National Cancer Institute* <https://seer.cancer.gov/statfacts/html/amyl.html> (2018).

3. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
4. Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).
5. Döhner, H. et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).
6. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
7. Byrd, J. C. et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* **100**, 4325–4336 (2002).
8. Patel, J. P. et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).
9. Haferlach, T. et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241–247 (2014).
10. Lundberg, P. et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* **123**, 2220–2228 (2014).
11. Deininger, M. W. N., Tyner, J. W. & Solary, E. Turning the tide in myelodysplastic/myeloproliferative neoplasms. *Nat. Rev. Cancer* **17**, 425–440 (2017).
12. Busque, L. et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179–1181 (2012).
13. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
14. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
15. Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
16. Huang, M. E. et al. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **72**, 567–572 (1988).
17. Shen, Z. X. et al. Use of arsenic trioxide (As₂O₃) in the treatment of acute promyelocytic leukemia (APL): II. Clinical efficacy and pharmacokinetics in relapsed patients. *Blood* **89**, 3354–3360 (1997).
18. Nakao, M. et al. Internal tandem duplication of the FLT3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911–1918 (1996).
19. Tse, K. F., Mukherjee, G. & Small, D. Constitutive activation of FLT3 stimulates multiple intracellular signal transducers and results in transformation. *Leukemia* **14**, 1766–1776 (2000).
20. Yamamoto, Y. et al. Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* **97**, 2434–2439 (2001).
21. Yokota, S. et al. Internal tandem duplication of the FLT3 gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia* **11**, 1605–1609 (1997).
22. Knapper, S. et al. A phase 2 trial of the FLT3 inhibitor lestaurtinib (CEP701) as first-line treatment for older patients with acute myeloid leukemia not considered fit for intensive chemotherapy. *Blood* **108**, 3262–3270 (2006).
23. O’Farrell, A. M. et al. An innovative phase I clinical study demonstrates inhibition of FLT3 phosphorylation by SU11248 in acute myeloid leukemia patients. *Clin. Cancer Res.* **9**, 5465–5476 (2003).
24. Smith, B. D. et al. Single-agent CEP-701, a novel FLT3 inhibitor, shows biologic and clinical activity in patients with relapsed or refractory acute myeloid leukemia. *Blood* **103**, 3669–3676 (2004).
25. DeAngelo, D. J. et al. Phase 1 clinical results with tandutinib (MLN518), a novel FLT3 antagonist, in patients with acute myelogenous leukemia or high-risk myelodysplastic syndrome: safety, pharmacokinetics, and pharmacodynamics. *Blood* **108**, 3674–3681 (2006).
26. Stone, R. M. et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N. Engl. J. Med.* **377**, 454–464 (2017).
27. Mardis, E. R. et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
28. Wang, F. et al. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).
29. Rohle, D. et al. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626–630 (2013).
30. Fiskus, W. et al. Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood* **114**, 2733–2743 (2009).
31. Schenk, T. et al. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat. Med.* **18**, 605–611 (2012).
32. Daigle, S. R. et al. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* **20**, 53–65 (2011).
33. Itzykson, R. et al. Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias. *Leukemia* **25**, 1147–1152 (2011).
34. Welch, J. S. et al. TP53 and decitabine in acute myeloid leukemia and myelodysplastic syndromes. *N. Engl. J. Med.* **375**, 2023–2036 (2016).
35. Konopleva, M. et al. Efficacy and biological correlates of response in a phase II study of venetoclax monotherapy in patients with acute myelogenous leukemia. *Cancer Discov.* **6**, 1106–1117 (2016).
36. DiNardo, C. D. et al. Safety and preliminary efficacy of venetoclax with decitabine or azacitidine in elderly patients with previously untreated acute myeloid leukemia: a non-randomised, open-label, phase 1b study. *Lancet Oncol.* **19**, 216–228 (2018).

37. Tyner, J. W. et al. Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. *Cancer Res.* **73**, 285–296 (2013).
38. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
39. Puissant, A. et al. SYK is a critical regulator of FLT3 in acute myeloid leukemia. *Cancer Cell* **25**, 226–242 (2014).
40. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
41. Canisius, S., Martens, J. W. M. & Wessels, L. F. A. A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. *Genome Biol.* **17**, 261 (2016).

Acknowledgements We thank all of our patients at all sites for donating precious time and tissue. DNA and RNA quality assessments, library creation and short read sequencing assays were performed by the OHSU Massively Parallel Sequencing Shared Resource. S. Sheela, C. Lai, K. Lindblad and K. Oetjen helped with study coordination at NIH. B. Sawicki and C. Cline helped with study coordination at the University of Florida. S. Ravencroft helped with patient sample shipping and data entry and K. Schorno provided project management and support of activities at the University of Kansas Cancer Center. J. Taw helped with patient sample shipping and S. Patel helped with data entry at Stanford University. Funding for this project was provided, in part, by a Therapy Acceleration Grant to B.J.D. and J.W.T. from The Leukemia & Lymphoma Society and by support provided by the Knight Cancer Research Institute (Oregon Health & Science University, OHSU). Supported by grants from the National Cancer Institute (1U01CA217862, 1U54CA224019, 1U01CA214116, 3P30CA069533-18S5) and NIH/NCATS CTSA UL1TR002369 (S.K.M., B.W.). A.S.B. was supported by the National Library of Medicine Informatics Training Grant (T15LM007088). J.W.T. received grants from the V Foundation for Cancer Research, the Gabrielle's Angel Foundation for Cancer Research, and the National Cancer Institute (1R01CA183947). C.R.C. received a Scholar in Clinical Research Award from The Leukemia & Lymphoma Society (2400-13), was distinguished with a Pierre Chagnon Professorship in Stem Cell Biology and Blood & Marrow Transplant and a UF Research Foundation Professorship. This work was supported in part by the Intramural Research Program of the National Heart, Lung, and Blood Institute of the National Institutes of Health.

Reviewer information Nature thanks P. Campbell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.W.T., C.E.T., D.B., B.W., S.E.K., S.L.S., N.L., A.R.S., E.T., S.K.M. and B.J.D. contributed equally to this work. J.W.T., C.E.T., S.K.M., A.A. and B.J.D. provided project oversight for experimental design, data management, data analysis and interpretation, methods development and development of the Vizome platform; B.W. conceived the Vizome platform, provided oversight for development of the platform and helped to provide project oversight for experimental design, workflow development, data processing, management, analysis and dissemination, and methods development; E.T. acquired patient samples, and structured, collected and analysed clinical data; S.L.S., A.R.S., M.A., I.E. and A.R. helped with processing of patient samples, ex vivo drug screening, DNA/RNA extractions and submissions of sequencing samples; N.L., R.Ca., J.D. and C.V. curated and annotated the clinical data of patients; S.E.K. provided variant confirmation and analysed data; D.B. led the workflow development for pre-processing and analysis of RNA sequencing, exome sequencing and ex vivo drug screening data, multivariate modelling and integration and helped with clinical data curation and integration, methods development and the back-end development of the Vizome platform, including the integration of Hitwalker; L.W. wrote all of the software for the Vizome platform, developed the platform as well as novel visualizations for data integration and display; J.R., R.Sc. and A.Y. provided clinical data integration from OHSU Research Data Warehouse; P.K. helped with the recruitment of patients and collection of samples for analysis; C.R.C. and R.T.S. were co-investigators for the repository protocol, and edited and provided feedback on the manuscript; L.D., C.T.J. and D.A.P. collected samples for the repository protocol and provided feedback on the manuscript; M.E.M. and R.R.P. consented patients and collected samples for the repository protocol and aided with clinical annotation; D.N. helped with the creation of drug-screening replicate plates; C.A.E., K.W.-S. and H.Z.

helped with data analysis; D.K.E. analysed and curated data and developed analytical processes; A.K. and M.M. helped with the development of the ex vivo drug screening processing workflow; S.J.W. enabled, facilitated and mentored basic, translational and clinical research activities arising from data generated by Beat AML at the University of Kansas Cancer Center site, participated in the Beat AML Symposia to share research and create new research projects; A.C., R.H., C.L. and R.Se. helped with the creation of exome-sequencing and RNA-sequencing libraries and with sequencing and data processing; D.D., C.N. and J.P. helped with genomic isolation, curation of samples and entry of patient clinical annotations; E.S. helped with the whole-exome sequencing and RNA-sequencing data processing, data management and developed the data dissemination workflows; D.L.W. served as a liaison for sample acquisition; M.W.D. was a local principal investigator for the repository protocol, consented patients, collected samples for the repository protocol, aided with clinical annotation, and edited and provided feedback on the manuscript; R.M.W. served as a manager for their repository protocol, consented patients and collected samples for the repository, aided with clinical annotation, and edited and provided feedback on the manuscript; M.M.L. helped with patient sample acquisition, IRB protocol development and maintenance, clinical data structure, collection and analysis; U.M., B.H.C., R.Co., A.D., K.-H.T.D., J.L. and S.E.S. helped with the acquisition of patient samples; A.d'A., J.B., R.B., C.C., M.D., J.G., H.H., A.J., K.J., R.J., S.Q.L., S.L., J.Mac., J.Mar., R.L.S., T.S., A.T., J.Wa. and J.Wo. helped with patient sample processing and ex vivo drug screening; C.S.H. was a principal investigator for their local protocol; consented patients and collected samples for the repository protocol, aided with clinical annotation, and edited and provided feedback on the manuscript; J.M.W. was a principal investigator for their local repository protocol, and edited and provided feedback on the manuscript; B.C.M. was a principal investigator for their local repository protocol, consented patients and collected samples for the repository protocol, aided with clinical annotation, and edited and provided feedback on the manuscript; D.C.C., T.L.L. and R.H.C. were principal investigators for their local repository protocol, consented patients and collected samples for the repository protocol, aided with clinical annotation, and edited and provided feedback on the manuscript; T.M. provided regulatory oversight; B.J. and K.W. provided regulatory oversight, and helped with the curation and entry of clinical annotations of patients; A.B. provided targetome overlay; J.C. and P.L. helped with technology transfer and project development.

Competing interests J.W.T. receives research support from Agios, Aptose, Array, AstraZeneca, Constellation, Genentech, Gilead, Incyte, Janssen, Seattle Genetics, Syros and Takeda, and is a co-founder of Vivid Biosciences. M.W.D. serves on the advisory boards of and/or as a consultant for Novartis, Incyte and BMS, and receives research funding from BMS and Gilead. C.S.H. receives research funding from Merck. T.L.L. consults for Jazz Pharmaceuticals and receives research funding from Tolero, Gilead, Precient, Ono, Bio-Path, Mateon, Genentech/Roche, Trovogene, Abbvie, Pfizer, Celgene, Imago, Astellas, Karyopharm, Seattle Genetics and Incyte. D.A.P. receives research funding from Pfizer and Agios and served on the advisory boards of Pfizer, Celyad, Agios, Celgene, AbbVie, Argenx, Takeda and Servier. B.J.D. serves on the advisory boards of Gilead, Aptose, and Blueprint Medicines and is a principal investigator or coinvestigator on Novartis and BMS clinical trials. The Oregon Health & Science University (on behalf of B.J.D.) has contracts with these companies to pay for patient costs, nurse and data manager salaries and institutional overhead. B.J.D. does not derive salary, nor does his laboratory receive funds from these contracts. The authors certify that all compounds tested in this study were chosen without input from any of the industry partners.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0623-z>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0623-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.K.M. or B.J.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Patient samples. All patients gave informed consent to participate in this study, which had the approval and guidance of the Institutional Review Boards (IRB) at Oregon Health & Science University (OHSU), University of Utah, University of Texas Medical Center (UT Southwestern), Stanford University, University of Miami, University of Colorado, University of Florida, National Institutes of Health (NIH), Fox Chase Cancer Center and University of Kansas (KUMC). Samples were sent to the coordinating centre (OHSU; IRB 9570; 4422; NCT01728402), where they were coded and processed. Specific names of centres associated with each specimen were coded and centres providing less than five samples were aggregated together and given one centre identifier. Mononuclear cells were isolated by Ficoll gradient centrifugation from freshly obtained bone marrow aspirates or peripheral blood draws. Cell pellets were snap-frozen in liquid nitrogen for subsequent DNA isolation (Qiagen, DNeasy Blood & Tissue Kit), freshly pelleted cells were lysed immediately in guanidinium thiocyanate (GTC) lysate for subsequent RNA isolation (Qiagen, RNeasy Mini Kit), and freshly isolated mononuclear cells were plated into an ex vivo drug sensitivity assays within 24 h (described in 'Ex vivo functional drug screens'). Skin punch biopsies were collected at the site of Jamshidi needle insertion for subsequent bone marrow biopsies and genomic DNA was isolated for use as matched normal controls for exome sequencing (Qiagen, DNeasy Blood & Tissue Kit). Clinical, prognostic, genetic, cytogenetic and pathologic laboratory values as well as treatment and outcome data were manually curated from the electronic medical records of the patient. Patients were assigned a specific diagnosis based on the prioritization of genetic and clinical factors as determined by WHO guidelines. To prevent re-identification, any patient over the age of 90 was placed into a >90 aggregated age bracket. Genetic characterization of the leukaemia samples included results of a clinical deep-sequencing panel of genes commonly mutated in haematologic malignancies (Sequenome and GeneTrails (OHSU); Foundation Medicine (UTSW); Genoptix; and Illumina).

Whole-exome sequencing and custom-capture validation sequencing. For whole-exome sequencing, Illumina Nextera RapidCapture Exome capture probes and protocol were used, which provided coverage of 37 Mb of genomic DNA-coding regions. In brief, following initial quality control on a TapeStation (Agilent), 50 ng of intact genomic DNA was fragmented and tagged (tagmentation) in a single step. Following clean-up, the tagmented DNA was amplified by 10 cycles of PCR, which added the indexed adaptors for clustering and sequencing. Libraries were hybridized to capture pools in 12 sample sets with two rounds of hybridization performed to increase specificity. Libraries recovered with streptavidin magnetic beads were amplified by 10 cycles of PCR, unincorporated reagents were removed with AMPure beads (Agencourt), and validated on the TapeStation. Quantification of capture pools was done using real-time PCR (Kapa). Libraries were denatured, flow cells set up using the cBot (Illumina), and run on a HiSeq 2500 using paired-end 100-cycle protocols. For the AML tumour samples, 5 or 6 lanes were run per capture group. For the matched skin biopsy samples, 3 lanes (or equivalent) were run per capture group. The instrument and chemistry for all capture groups are provided in Supplementary Tables 12, 13.

For validation of sequencing results, an 11.9-Mb custom-capture library was developed that provided coverage of all variants previously reported in AML as well as all new variants detected from exome sequencing in this study (genes, variants and capture regions for this custom library can be found in Supplementary Table 14). This capture library was then applied to sequence 96 specimens that had previously been subjected to whole-exome sequencing for validation of variant calls.

Whole-exome sequencing data processing. We developed customized analytical pipelines that combined published algorithms with novel filtering, curation and quality-control steps. Detailed depictions of analytical workflows can be found in Supplementary Table 6 and on our online browser (http://vizome.org/additional_figures_BeatAML.html). Initial data processing and alignments were performed with commonly used analytical tools. For each flowcell and each sample, the FASTQ files were aggregated into single files for read 1 and 2. During this process, these reads were trimmed by 3 on the 5' end and 5 on the 3' end. BWA MEM version 0.7.10-r789⁴² was used to align the read pairs for each sample-lane FASTQ file. As part of this process, the flowcell and lane information was kept as part of the read group of the resulting SAM file. The Genome Analysis Toolkit (version 3.3) and the bundled Picard (version 1.120.1579) were used⁴³ for alignment post-processing. 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 following steps were performed per sample-lane SAM file generated for each CaptureGroup: (1) the SAM files were sorted and converted to BAM via SortSam; (2) MarkDuplicates was run, marking both lane level standard and optical duplicates; (3) the reads were realigned around insertions and deletions (indels) from the reads using RealignerTargetCreator/IndelRealigner; (4) base quality score recalibration was performed.

The resulting BAM files were then aggregated by sample and an additional round of MarkDuplicates was performed out at the sample level. Quality-control reports were generated using the ReportingTools⁴⁴ and qrc⁴⁵ Bioconductor R packages along with sequencing core and alignment output files. Each AML sample BAM was paired with its skin biopsy pair and an additional round of indel realignment was carried out to ensure consistency of genotypes between the two samples. If an AML sample did not have a pair, the indel realignment was instead done at the sample level.

Whole-exome sequencing variant detection. For genotyping, each AML-skin biopsy pair was realigned at the sample level and then genotyped for single-nucleotide variations using Mutect version 1.1.7⁴⁶ and VarScan2 version 2.4.1⁴⁷. Indels were produced using VarScan2. Each variant call format (VCF) file was annotated using the Variant Effect Predictor version 83 against GRCh37⁴⁸. The resulting VCF files were filtered to include only those annotated to a gene and were converted to mutation annotation format (MAF) format using the vcf2maf version 1.6.6 tool⁴⁹.

Mutect version 1.1.7⁴⁶ was run using default parameters, except that no limit was placed on the number or frequency of the alternative allele frequency in the normal condition to help to address normal contamination.

VarScan2 version 2.4.1⁴⁷ was run in somatic mode with the recommended filtering scheme⁵⁰, except as shown in Supplementary Table 23.

Indels and single-nucleotide variants (SNVs) were produced for the tumour-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 version 83 on GRCh37. This assignment was done using the same VEP parameters as the vcf2maf (version 1.5.0) program.

The TCGA AML variants⁶ in MAF form were downloaded from the GDC archive site: <https://portal.gdc.cancer.gov/legacy-archive/files/c410d927-d49c-4d4f-8356-601bee563ebe>. The MAF files were converted to VCF files using the vcf2maf suite⁵¹. The resulting VCF files were lifted over from NCBI36 to GRCh37 of the human genome using CrossMap⁵². Only those variants that successfully lifted over were kept.

Variants from supplementary table 2 from a previously published study¹⁴ were extracted and further processed, removing variants that were ambiguous in terms of external sources and could not be found in their whole-exome sequencing variants. The unique set of Beat AML variants was annotated relative to RefSeq transcripts using Ensembl VEP similar to above and all consequences were kept. This set of variants and consequences was searched against the set of processed variants from the previously published study¹⁴.

Using the runs from MuTect and VarScan2, these data were next filtered to keep only the protein-influencing 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)⁵³ dataset or be seen at a frequency <0.01. These data present several additional challenges. First, somatic calls cannot be obtained directly from the tumour-only samples. Second, there is always a possibility of tumour 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. (1) An initial set of higher-confidence somatic mutations were retrieved from the paired tumour-normal samples requiring tumour variant allele frequency (VAF) ≥ 8% and normal VAF ≤ 5%, in addition to the significance tests already performed by the programs. (2) 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 the previously published study¹⁴ or from the lifted-over set of variants from the TCGA AML paper⁶. (3) Mutations from the overall set were kept if: (a) the overall number of calls in the paired samples was not more than twice the number of high-confidence somatic calls; (b) the tumour-only frequency for the calls was less than 50% greater than the number of calls in the paired samples; (c) the mutation was seen in list of the previous study¹⁴ or TCGA dataset⁶. (4) High-confidence somatic mutations were kept regardless.

The data from the two genotypers were combined along with FLT3-ITD calls from Pindel⁵⁴. Comparing our variant lists from whole-exome sequencing and custom-capture validation sequencing, we noticed—similar to others⁵⁵—that low allele frequency C-to-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, along with a curated 'blacklist' (Supplementary Table 15) of known problematic variants and/or genes, including mitochondrial DNA variants. In addition, all variants that were seen in a cumulative list of normal samples from Beat AML at a frequency greater than 1% were removed. Cumulatively, of this set, 94% of covered SNVs were validated with 82% of indel calls also being confirmed with validation sequencing. Manual review was then carried out in the following steps. (1) The addition back of all flagged rows from the previous study¹⁴. (2) The review of all TCGA-flagged rows for VAF pattern that matched or did not match with known drivers in the 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 a VAF pattern that was unlike known drivers in the same specimens. (3) Other variants were added back based on other specimens that had the same variant that were still on the include list, and if the VAF pattern looked convincing for inclusion. (4) All genes from the previous study¹⁴ with only frameshift and/or nonsense variants were manually reviewed and missense mutations were manually removed. (5) Genes and/or variants that were on both the include and exclude lists were manually reviewed and they were removed if they were C-to-A with over 15% VAF, they did not validate and/or the VAF pattern was unlike known drivers in same specimen. (6) Further review of all genes in the 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 the 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 tumour-only patients for genes from the previous study¹⁴. The tumour-only AML samples used in another study (H.Z. et al., manuscript submitted) were removed and used for that study.

Detection of internal FLT3-ITD and NPM1 mutations. A subset of samples was tested for FLT3-ITD and NPM1 mutation status using an internally run PCR assay and capillary electrophoresis. Genomic DNA (gDNA) was extracted from fresh blood or bone marrow aspirates of patients with AML and was used to detect the presence or absence of FLT3-ITD and NPM1 4-bp insertion mutations^{56,57}. Primers for FLT3 spanned approximately 330 bp to include the common internal duplication site⁵⁶. Primers for NPM1 spanned approximately 170 bp to cover the clustered multiple insertional or indel sites^{57,58}. Primers were HPLC-purified by the manufacturer. The multiplex PCR reaction solution⁵⁹ consisted of 100 ng gDNA, 10 pmol of the respective forward and reverse primers for FLT3 and NPM1, 25 mmol l⁻¹ MgCl₂, 2.5 mmol l⁻¹ dNTPs, 5 µl 10× PCR buffer, 0.2 µl AccuTaq DNA polymerase and water in a total volume of 50 µl. The PCR conditions were: initial denaturing for 5 min at 94 °C, followed by 30 cycles at 94 °C for 30 s, 56 °C for 45 s and 72 °C for 30 s with a final cycle of 10 min at 72 °C. The PCR products were diluted 1:10 and analysed by capillary electrophoresis on a QIAxcel high-resolution DNA cartridge according to the manufacturer's protocol. Forward primer FLT3: 5'-AGCAATTAGGTATGAAAGCCAGCTA-3'; reverse primer FLT3: 5'-CTTTCAGCATTTTGACGGCAACC-3'. Forward primer NPM1: 5'-GTTTCTTTTTCCTCCAGGCTATCAAG-3'; reverse primer NPM1: 5'-CACGGTAGGGAAGTTCTCACTCTGC-3'.

Derivation of FLT3-ITD and NPM1 consensus calls. Consensus FLT3-ITD and NPM1 mutation calls found in the clinical summary table (Supplementary Table 5) were determined by comparing the internal capillary PCR test (internal; according to the methods described in 'Detection of internal FLT3-ITD and NPM1 mutations'), the Clinical Laboratory Improvement Amendments/College of American Pathology (CLIA/CAP) laboratory run test (Sequenome, GeneTrails, Foundation Medicine, Genoptix, Illumina). The internal test was used for the sample consensus call when available, as it was performed on the exact sample that was used for further ex vivo drug sensitivity assays. Where discordance existed between the internal test and the CLIA laboratory test results, the sample was flagged for manual review. The trace file for the internal test was visually inspected and if discordance with the CLIA/CAP test results persisted, the whole-exome sequencing data were then used to help to determine the consensus call.

Derivation of CCAAT enhancer binding protein α (CEBPA) biallelic consensus calls. N-terminal and C-terminal mutations have been described to occur on opposing alleles and patients with CEBPA biallelic mutations have been shown to fall into a favourable risk category⁶⁰. Patients were scored positive for biallelic CEBPA mutation if described in the clinical notes as biallelic or double-positive. Patients were also scored as CEBPA biallelic if both N-terminal and C-terminal mutations were identified in the whole-exome sequencing data.

RNA sequencing and data processing. All samples were sequenced using the Agilent SureSelect Strand-Specific RNA Library Preparation Kit on the Bravo robot (Agilent). In brief, poly(A)⁺ RNA was chemically fragmented. Double-stranded cDNAs were synthesized using random hexamer priming with 3' ends of the cDNA adenylated, after which indexed adaptors were ligated. Library amplification was performed using three-primer PCR using a uracil DNA glycosylase addition for strandedness. Libraries were validated with the Bioanalyzer (Agilent) and combined to run 4 samples per lane, with a targeted yield of 200 million clusters. Combined libraries were denatured, clustered with the cBot (Illumina) and sequenced on the HiSeq 2500 using a 100-cycle paired-end protocol. In addition to the AML samples, there was also a sample of purified CD34 molecule (CD34)⁺ cells from healthy control bone marrow, which was included in each sample group (for a total of 12 times sequencing this control RNA). This control served as both a healthy control and a quality check on inter-group batch effects. In addition, 21 individual healthy bone marrow samples were also included, two of which were

CD34-selected (17-00053 and 17-00056) with the other 19 being whole mononuclear bone marrow cells from healthy donors.

Workflows for processing and analysis of RNA-sequencing data to generate gene counts and gene fusions for each sample are shown on our online browser (http://vizome.org/additional_figures_BeatAML.html) with processed gene expression values for each specimen listed in Supplementary Tables 8, 9. For each flowcell and each sample, the FASTQ files were aggregated into single files for read 1 and read 2 (if not already done by the sequencing core). During this process, these reads were trimmed by 3 on the 5' end and 5 on the 3' end. Alignment of reads was performed using the subunc aligner (version 1.5.0-p2)⁶¹. BAM files obtained from subunc were used as inputs into featureCounts (version 1.5.0-p2)⁶² and gene-level read counts were produced. For a reference genome, the GRCh37 build provided by the Broad as part of the GATK bundle was used. Gene assignments were based on the Ensembl build 75 gene models on GRCh37. The following parameters were used for the software:

```
subunc -i /path/to/reference/ -u -r fastq1 -R fastq2 -o outputBAMFilename -I 5 -T 7 -d 50 -D 600 -S fr and 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.
```

The data were collated from featureCounts matrices and all genes with no counts across the samples were excluded. Genes with duplicate gene symbols and those for which the counts were <10 for 90% or more of the samples were additionally removed before normalization similar to the approach suggested for weighted gene correlation network analysis (WGCNA)⁶³. Samples for which the median expression was less than 2 standard deviations below the average were removed from the dataset ($n = 10$). Normalization was performed using the conditional quantile normalization procedure⁶⁴, which produced GC-content-corrected log₂ reads per kilobase per million mapped reads (RPKM) values. This procedure produces both offsets to be used in conjunction with edgeR as well as a matrix of log₂-normalized RPKM values for clustering.

In addition, the subunc BAM files were processed using the RNA-sequencing genotyping protocol (as of GATK version 3.3), which was similar to the whole-exome sequencing protocol described in the 'Whole-exome sequencing data processing' section, including the following steps for each sample: (1) MarkDuplicates; (2) SplitNCigarReads; (3) RealignerTargetCreator/IndelRealigner; (4) base quality score recalibration. The resulting BAM files were used to produce RNA genotypes using the UnifiedGenotyper for the purposes of quality control and ethnicity estimation.

Gene fusion data were additionally generated using the TopHat-Fusion (version 2.0.14) program using default parameters⁶⁵.

Coexpression network formation. We formed coexpression modules using the WGCNA procedure on the RNA-sequencing data from the 'RNA sequencing and data processing' section. All RNA-sequencing samples were used to form the set of modules. Owing to the heterogeneity of the clinical expression data, we generated 'signed hybrid' networks using the 'bicor' correlation⁶⁶, setting the maximum proportion of outliers to 0.1. We ran the procedure multiple times, varying several parameters to choose the most relevant set for further analysis. The WGCNA procedure was run on datasets formed from the top 2,000 and 5,000 most variable genes. For each dataset, we set the 'power' variable to either 2 or 3. For each of these runs, we varied the module detection parameters of dynamicTreeCut⁶³, namely the deepSplit parameter was set to 0 or 2 and the pamStage parameter was set to TRUE or FALSE. For each of these sets of modules we computed a series of module quality statistics⁶⁷, mean correlation, mean adjacency, mean maximum adjacency ratio (MAR), mean correlation with the module eigengene (KME), proportion of variability explained and the mean cluster coefficient. Significance of modules was determined by computing a z score of each of these values relative to the mean and standard deviation of those from 100 random assignment of modules. We chose the set of modules to use in our analyses as those that were most correlated with the 'specimenSpecificDx' using module quality as a tie-breaker. The analysis set of modules was chosen to be the version using the 5,000 most variable genes, power set to 2 and modules formed using deepSplit = 2 and pamStage = F. Of this set of modules, only the grey module did not have a summary z statistic (median across the four density measures) of at least 2. In addition, after correcting the data using the estimated principal components⁶⁸, the module structure did not change appreciably (data not shown).

Quality control. The UnifiedGenotyper runs for both the whole-exome sequencing and RNA-sequencing data were combined into a single VCF file using the GATK CombineVCFs functionality. This combined VCF file was converted to a GDS file using SNPRelate (version 1.12.2)⁶⁹. Note, the version is the most recent version as several versions were used across the entire project. The overall similarity of the genotypes of each pair of samples was computed, termed identity by state (IBS) and hierarchical clustering was performed using one minus this similarity. From this clustering and visualization we had devised hard cut-offs for further inspection based on the types of data being compared. For instance, samples not meeting the specified IBS thresholds (DNA-DNA = 0.9; RNA-RNA = 0.83;

DNA–RNA = 0.89) were subject to manual review. On the basis of the dendrogram structure as well as the clinical/laboratory 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 from the RNA-sequencing analysis and for treatment as tumour-only samples in the whole-exome sequencing analysis as described in ‘Whole-exome sequencing variant detection’.

Fusion annotation for analysis. Fusions calls were determined from a consensus of three datatypes, a specific diagnosis categorization at the time of sample acquisition, current set of clinical karyotypes and fusions detected in RNA-sequencing data by TopHat-Fusion. All sources were limited to the same set of known fusions: *RUNX1-RUNX1T1*, *CBFB-MYH11*, *MLLT3-KMT2A*, *DEK-NUP214*, *GATA2-MECOM* and *PML-RARA*. It was determined that the RNA-sequencing calls did not provide additional resolution in detecting these known fusions and was not performed on all the samples, so the consensus was limited to the clinical karyotype calls as well as the specific diagnosis categorization (which was determined based on karyotype and other cytogenetic clinical tests). Overall, the calls were based on the karyotype data except in 10 cases; for 3 cases the karyotype and diagnosis was sufficiently complex to warrant a separate ‘complex’ categorization. The remaining 7 of these cases were set to the specific diagnosis classification. It should be noted that there was additional support from the RNA-sequencing data for several of these cases.

Ethnicity. The combined RNA and whole-exome sequencing VCF files from the ‘Quality control’ step were merged with a set of Hapmap genotypes⁷⁰ lifted over to build 37. The SNPRelate package was used to convert the VCF to GDS, perform linkage disequilibrium (LD) pruning using an LD threshold of 0.2, MAF cut-off of 0.05 and allowing a missing rate of 0.3 and calculation of the principal components. Previously published methodology⁷¹ was used to assign admixture proportions relative to the HapMap samples using the principal components. Each sample was assigned to an ethnicity group based on the group with the maximum admixture proportion. If the maximum was 50% or less, we labelled it ‘Admixed’. As we had observed previously that the clustering of ethnicities for RNA-sequencing samples are more diffuse than exome sequencing, we assigned the final inferred ethnicity to each patient based on the distinct whole-exome sequencing calls if available, deferring to RNA sequencing only if not available. If multiple exome sequencing samples were present with discrepant calls, the data of the patient were manually reviewed. The only patient for which this occurred was patient 4043, a self-identified Hispanic who had two RNA samples and an exome sequencing sample inferred as ‘White’ and one exome sample inferred to be ‘Hispanic’. Only the ‘White’ call for the exome sequencing had an admixture proportion over 0.5. The patient was kept consistent with the self-identification and labelled as ‘Hispanic’.

Sex. For DNA, coverage was first computed over the Y chromosome and the counts for each sample were added up and log₁₀-transformed (after adding 1 to all the counts). *k*-means clustering was used to assign samples to two clusters with the cluster with the lower mean labelled as the ‘Female’ cluster.

For RNA sequencing, counts were converted to counts per million after applying the Trimmed Mean of M scaling normalization⁷². A set of 28 genes were chosen to successfully discriminate the genders using DE analysis over multiple studies (data not shown) and were used in conjunction with *k*-means clustering to form two clusters. The ‘Female’ cluster was labelled based on high *XIST* expression.

ELN 2017 classification. This procedure is based on the categorization in table 5 of the 2017 ELN update paper⁵. Karyotypes in the clinical file were first cleaned and parsed into clones or subclones and distinct abnormalities using standard conventions⁷³. The current representation was corrected for nomenclature type (for example, ‘idem’ versus ‘sl’) in a basic manner. For instance, ambiguous events, such as chromosomal loss (for example, –15), were not corrected for whether the preceding clone had a counteracting gain. Also, additional ‘+’ or ‘–’ symbols in conjunction with valid karyotype operators in a separate clone (for example, +del(12)(q15) or –del(12)(q15)) were treated separately with gains (+) being kept in the unique count of events and losses (–) being removed.

Abnormalities were first checked for the following categories: (1) *RUNX1-RUNX1T1*; (2) *CBFB-MYH11*; (3) *MLLT3-KMT2A*; (4) *DEK-NUP214*; (5) *KMT2A*–*; (6) *BCR-ABL1*; (7) *GATA2-MECOM*; (8) –5/del(5q); (9) –7; (10) –17. Categories 1–8 were further considered to be WHO recurrent fusions.

The number of unique abnormalities (across clones) was then computed. Whether or not a karyotype was considered to be polyploid was also recorded (at least 60 chromosomes or ‘≥ 3n’ or labeled). *NPM1*, *FLT3-ITD* and biallelic *CEBPA* were derived from consensus calls. *FLT3-ITD* allelic ratios were determined only for the samples with an internal assay. The MAF values of the internal assay were converted to a ratio using the formula $MAF/(1 - MAF)$. *RUNX1*, *ASXL1* and *TP53* were derived from the clinical genotypes. Abnormal 17 calls were manually curated from the karyotype data and clinical genotype calls.

The determination of ELN 2017 categories proceeds by assigning true/false/not available values to one or more of the five columns (three ELN and two ambiguous) in the following manner. (1) ‘isFavourable’ is considered true if a sample has at least one of the following: (a) *RUNX1-RUNX1T1*; (b) *CBFB-MYH11*; (c) positive *NPM1* and negative *FLT3-ITD*; (d) positive *NPM1* and positive *FLT3-ITD* with allelic ratio <0.5; (e) biallelic *CEBPA*. (2) ‘isFavourableOrIntermediate’ when *NPM1* is positive and *FLT3-ITD* is positive but the allelic ratio is not available. (3) ‘isAdverse’ is considered true if a sample has at least one of the following: (a) *DEK-NUP214*; (b) *KMT2A*–*; (c) *BCR-ABL1*; (d) *GATA2-MECOM*; (e) –5/del(5q); (f) –7; (g) –17; (h) abn_17; (i) three or more abnormalities and no WHO recurrent fusions; (j) one monosomy (autosomal) and at least one additional abnormality except for *CBFB-MYH11*; (k) positive *RUNX1* or *ASXL1* and not considered to be ‘isFavourable’ or ‘isFavourableOrIntermediate’; (l) positive *TP53*; (4) ‘isIntermediate’ is considered true if a sample has at least one of the following: (a) *MLLT3-KMT2A*; (b) *NPM1* is positive and *FLT3-ITD* is positive with allelic ratio ≥0.5; (c) *NPM1* is negative and *FLT3-ITD* is negative or has a low allelic ratio (<0.5); (d) at least one abnormality and is not considered ‘isFavourable’ or ‘isAdverse’; (5) ‘isIntermediateOrAdverse’ when *NPM1* is negative and *FLT3-ITD* is positive without an allelic ratio. Calls were annotated as ‘not available’ in the absence of *FLT3-ITD* or *NPM1* calls.

Samples for which the specific diagnosis at inclusion indicated ‘acute promyelocytic leukaemia with t(15;17)(q22;q12)’ were automatically set to ‘Favourable’. Any overlaps in the categories were resolved based on manual expert review.

Ex vivo functional drug screens. Ex vivo functional drug screens were performed on freshly isolated mononuclear cells from AML samples. In brief, 10,000 cells per well were arrayed into three, 384-well plates containing 122 small-molecule inhibitors. This panel contained graded concentrations of drugs with activity against two-thirds of the tyrosine kinase as well as other non-tyrosine kinase pathways, including mitogen-activated protein kinases (MAPKs), the pathway involving phosphatidylinositol-4,5-bisphosphate 3-kinase, AKT serine/threonine kinase 1 and mechanistic target of rapamycin kinase (PIK3C–AKT–MTOR); protein kinase AMP-activated (AMPK, also known as PRKAA1), ATM serine/threonine kinase (ATM), Aurora kinases, calcium/calmodulin-dependent protein kinases (CAMKs), cyclin-dependent kinases (CDKs), serine/threonine protein kinase 3 (GSK3), IκB kinase (IκK), cAMP-dependent protein kinase (PKA), protein kinase C (PKC), polo-like kinase 1 (PLK1) and RAF proto-oncogene serine/threonine kinase (RAF). In addition, the library contained small-molecule inhibitors with activity against the BCL2 family, bromodomain containing 4 (BRD4), Hedgehog, heat shock protein 90 (HSP90), NOTCH/γ-secretase, proteasome, survivin, signal transducer and activator of transcription 3 (STAT3), histone deacetylase (HDAC), and WNT/β-catenin. Drug plates were created using inhibitors purchased from LC Laboratories and Selleck Chemicals and master stocks were reconstituted in dimethyl sulfoxide (DMSO) and stored at –80 °C. Master plates were created by distributing a single agent per well in a seven-point concentration series, created from threefold dilutions of the most concentrated stock resulting in a range of 10 μM to 0.0137 μM for each drug (except dasatinib, ponatinib, sunitinib and YM-155, which were plated at a concentration range of 1 μM to 0.00137 μM). DMSO-control wells and positive-control wells containing a drug combination of flavopiridol, staurosporine and velcade were placed on each plate, with the final concentration of DMSO ≤0.1% in all wells. Daughter plates were created using a V&P Scientific 384-well pin tool head operated by the Caliper Sciclone ALH 3000 and equipped with 0.457-mm diameter, 30-nl, slotted stainless-steel pins (FP1NS30). Daughter and destination plates were sealed with peelable thermal seals using a PlateLoc thermal sealer. Destination plates were stored at –20 °C for no more than three months and thawed immediately before use. Primary mononuclear cells were plated across single-agent inhibitor panels within 24 h of collection. Cells were seeded into 384-well assay plates at 10,000 cells per well in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with fetal bovine serum (FBS) (10%), l-glutamine, penicillin–streptomycin, and β-mercaptoethanol (10–4 M). After three days of culture at 37 °C in 5% CO₂, MTS reagent (CellTiter96 Aqueous One; Promega) was added, the optical density was measured at 490 nm, and raw absorbance values were adjusted to a reference blank value and then used to determine cell viability (normalized to untreated control wells).

Ex vivo functional drug screen data processing. A workflow in which the data normalization, curve fit parameters and quality assurance/quality control steps are summarized can be found on our online browser (http://vizome.org/additional_figures_BeatAML.html) with processed drug response data for each specimen listed in Supplementary Table 10. 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 before model fitting.

First, a ‘curve-free’ AUC (integration based on fine linear interpolation between the seven 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 among the replicates was noted and those runs with differences >100 were removed.

Second, the remaining within-plate replicates had their normalized viability averaged 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 among 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. On the basis of the methodology used in our prior drug-combination study⁷⁴, a probit regression was fit to all possible run groups using the model: (normalized viability / 100) $\sim 1 + \log_{10}(\text{concentration})$. For all groups there were $n = 7$ dose-response measurements.

The summary analyses of curve fit were inspected and cut-offs were devised removing all runs with an Akaike information criteria (AIC) > 12 and deviance > 2. For inhibitors that were run using multiple concentration ranges, only the most-recent 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 or resistant calls were manually reviewed as subject to removal.

Ex vivo functional drug screen analysis. For all drug analyses that required a call of sensitivity or resistance (for example, the gene expression signatures), sensitivity or resistance was determined by the lowest and highest 20% of the AUC values for each drug.

Correlations between drugs in families. For each inhibitor in the study, available data on targets of the inhibitors were pulled from a variety of online resources and published studies, many of which were aggregated in the Cancer-Targetome^{75,76}. Activity of each inhibitor for targets was then distilled into a five-tier system to afford comparison across drugs with differing degrees of potency and/or for which differing assays were used to measure drug and/or target activity. Well-represented genes, gene families and pathways were then filtered for drugs that have activity in the top 3 tiers for one or more member of the gene family or pathway. These lists were then manually curated to generate the final list of high-confidence drug target families (shown in Supplementary Table 11). For each inhibitor assigned to at least one drug target family, the Pearson's correlation was computed against all other drugs assigned to at least one drug target family for the AUC values of all available samples shared between the two drugs.

Correlations between drugs and samples. Drugs were first filtered to require greater than two hundred samples per drug. Additional samples were removed accordingly to allow correlations to be computed between all present samples using available AUC data and between all drugs.

Summary drug response scores. For each patient sample, a binary score (1/0) was used for each drug based on the same threshold as for the gene signatures (for example, sensitivity or resistance was determined by the lowest and highest 20% of the AUC values for each drug). Individual scores were computed for resistance and sensitivity separately and presented as the proportion over all screened drugs for each patient sample.

Expression analysis and integration with ex vivo functional drug screen. For all the below analyses the earliest sample was chosen for each patient.

Expression heat map. The top 2,000 most variable genes were extracted. The expression values were centred and scaled across patients and complete-linkage hierarchical clustering was performed using the ComplexHeat map R package⁷⁷.

Sensitive or resistant differential expression. For each drug, it was required that at least three sensitive and three resistant samples using the 20%/20% criteria outlined in the 'Ex vivo functional drug screen analysis' section. Patient samples were limited to those labelled as sensitive or resistant. Next, genes were limited based on their expression, for which at least half the patients used for analysis had to have greater than one count per million (an approach suggested in the limma users manual)⁷⁸. The normalized expression as in the data described in the 'RNA sequencing and data processing' section with the chosen samples and genes was used for differential expression analysis. Because the data had not been batch-corrected at this point, surrogate variable analysis (SVA)⁷⁹ was used to infer covariates for correcting out technical confounders. Next, the linear model fitting for each gene was performed using the limma-trend approach⁸⁰, testing whether the average expression was different between resistant and sensitive correcting for the SVA covariates. Genes with Benjamini-Hochberg³⁸ FDR values of less than 0.05 were kept for the cluster analysis. The expression matrix was corrected with respect to the estimated surrogate variables for consistency with the differential expression procedure using fSVA⁸¹ and MCLUST⁸² was used to determine optimal number of clusters and parameterization. The results were then visualized using a CLUSPLOT⁸³, which displays the clustering results with respect to the first two principal components of the gene expression for the kept genes.

Mutation analysis and integration with ex vivo functional drug screen. For all the below analyses in which groups of samples were compared, the earliest sample was chosen for each patient.

TCGA comparison. The lifted-over TCGA variants from the 'Whole-exome sequencing variant detection' section were annotated using the VEP from Ensembl

build 83, filtered for protein-altering and splice site variants and our 'blacklist' was applied to ensure the variant sets were comparable.

Co-occurrence and mutual exclusivity. Only mutations seen in at least 10 patients were kept. The DISCOVER⁴¹ method was used to determine significant mutual exclusivity and co-occurrence. A plot of the co-occurrences was generated using corrrplot⁸⁴ with the odds ratio of the pairwise co-occurrence used to colour and scale the circle sizes.

Association between mutations and drugs. For each mutated gene in the exome sequencing samples and each recurrent fusion (counting FLT3-ITD as a distinct entity from other FLT3 mutants), we determined all available (at least 5 patients) pairwise and three-way co-occurrence sets. For each drug and each valid set of genes (from one to three genes), we fitted a linear model with AUC as the response and examined the linear contrast (that is, two-sided Student's *t*-test) comparing the AUC of the gene(s) to the appropriate negative. For example the average AUC of the FLT3, DNMT3A and NPM1 mutants would be compared to average AUC of the samples negative for all three genes. FDR was computed using the Benjamini-Hochberg method over all the drugs.

For the ibrutinib and entospletinib comparisons, the presence and absence of the three genes or mutations: NPM1, FLT3-ITD and DNMT3A was collapsed into levels of a single factor. The corresponding single-factor ANOVA was carried out with the 'triple-negative' category set as the reference. Significance of the *P* values of each coefficient was compared to the Bonferroni-corrected 0.05 level.

For the JAK-family analysis, the AUC values were pooled for the four JAK inhibitors (CYT387, tofacitinib (CP-690550), JAK inhibitor I, ruxolitinib (INCB018424)) for each gene mutation set (BCOR, BCOR and DNMT3A, BCOR and RUNX1, BCOR and SRSF2). The contrast of the difference between BCOR and RUNX1 samples and the average of the other three mutation groups was tested.

Integration of both mutation and RNA sequencing with ex vivo functional drug screen. Mutations (0/1 score) and the module Eigengenes from the WGCNA analysis were used separately and combined together in regression models with coefficients selected using the LASSO approach⁴⁰ as implemented in glmnet⁸⁵. For each data type and the combination, only drugs with at least 200 patients samples were tested. The three datasets were initially randomly separated into training (75%) and test (25%) sets. Similar to a previous approach⁸⁶, a bootstrap aggregation approach was used in which the 1,000 bootstraps of the training dataset were generated and for each one, the LASSO was trained using tenfold cross-validation. Predictions were formed for the test dataset over these bootstrap models and the predicted AUC was averaged. *R*² values were computed for these aggregated predictions relative to the test AUC values. As performance was seen to be dependent on the initial split between test and training, we repeated the entire process 100 times, recording the mean and standard deviation of the *R*² value as well as the count non-zero coefficients

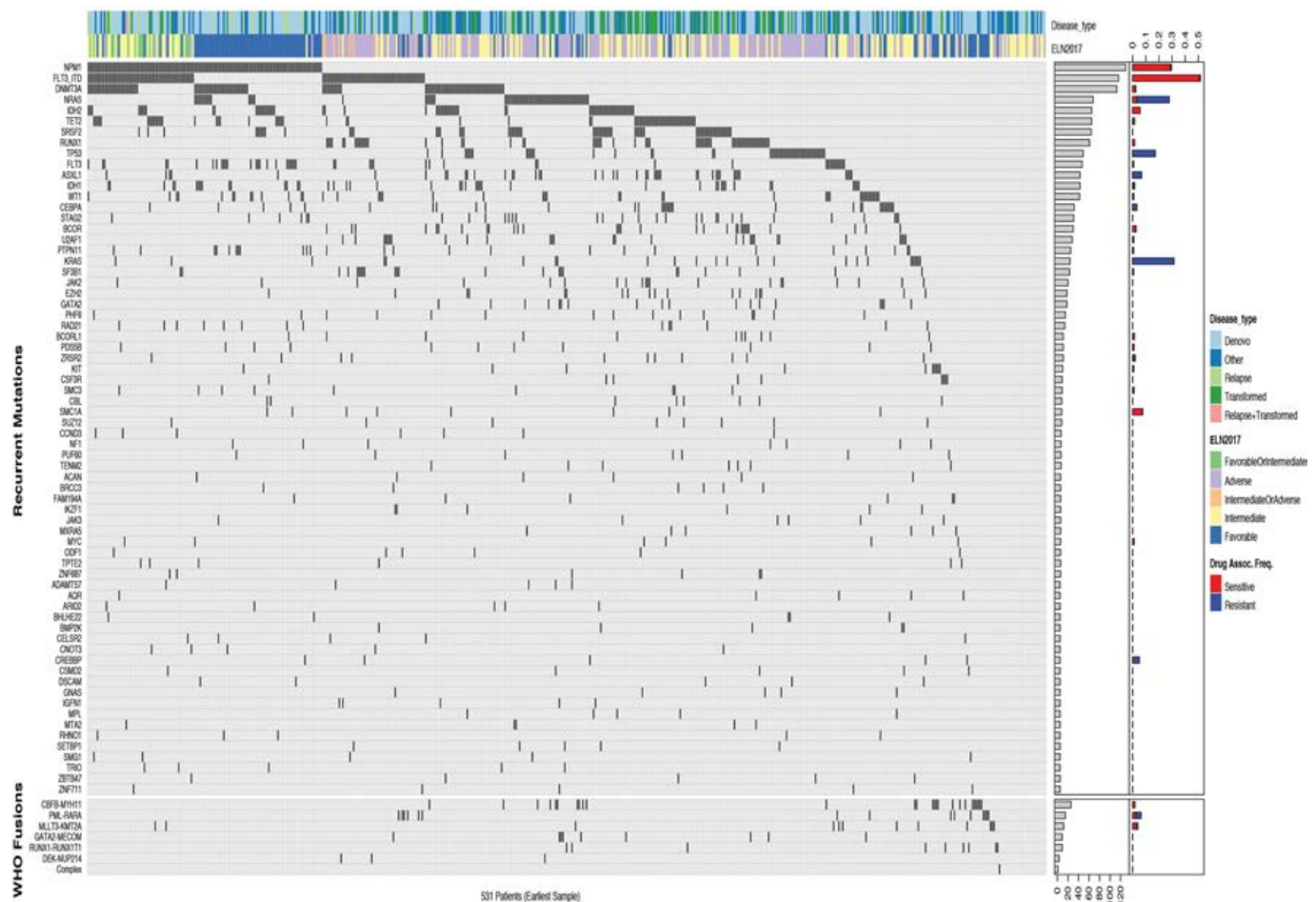
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All raw and processed sequencing data, along with relevant clinical annotations, have been submitted to dbGaP and Genomic Data Commons. The dbGaP study ID is 30641 and accession ID is phs001657.v1.p1. The raw data for clinical annotations, variant calls, gene expression counts and drug sensitivity that underlie Figs. 1-3 and Extended Data Figs. 1-9 are provided as Source Data. In addition, all data can be accessed and queried through our online, interactive user interface, Vizome, at <http://www.vizome.org/>.

42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
43. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).
44. Huntley, M. A. et al. ReportingTools: an automated result processing and presentation toolkit for high-throughput genomic analyses. *Bioinformatics* **29**, 3220-3221 (2013).
45. Buffalo, V. qrc: Quick Read Quality Control. R package version 1.22.0 <http://github.com/vsbuffalo/qrc> (2012).
46. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213-219 (2013).
47. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568-576 (2012).
48. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070 (2010).
49. Memorial Sloan Kettering. vcf2maf. version 1.6.6 <https://github.com/mskcc/vcf2maf/> (2016).
50. Koboldt, D. Release note for VarScan version 2.4.1. <https://github.com/dkoboldt/varscan/blob/master/VarScan.v2.4.1.description.txt> (2015).
51. Memorial Sloan Kettering. maf2vcf. version 1.6.6 <https://github.com/mskcc/vcf2maf/> (2016).
52. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-1007 (2014).

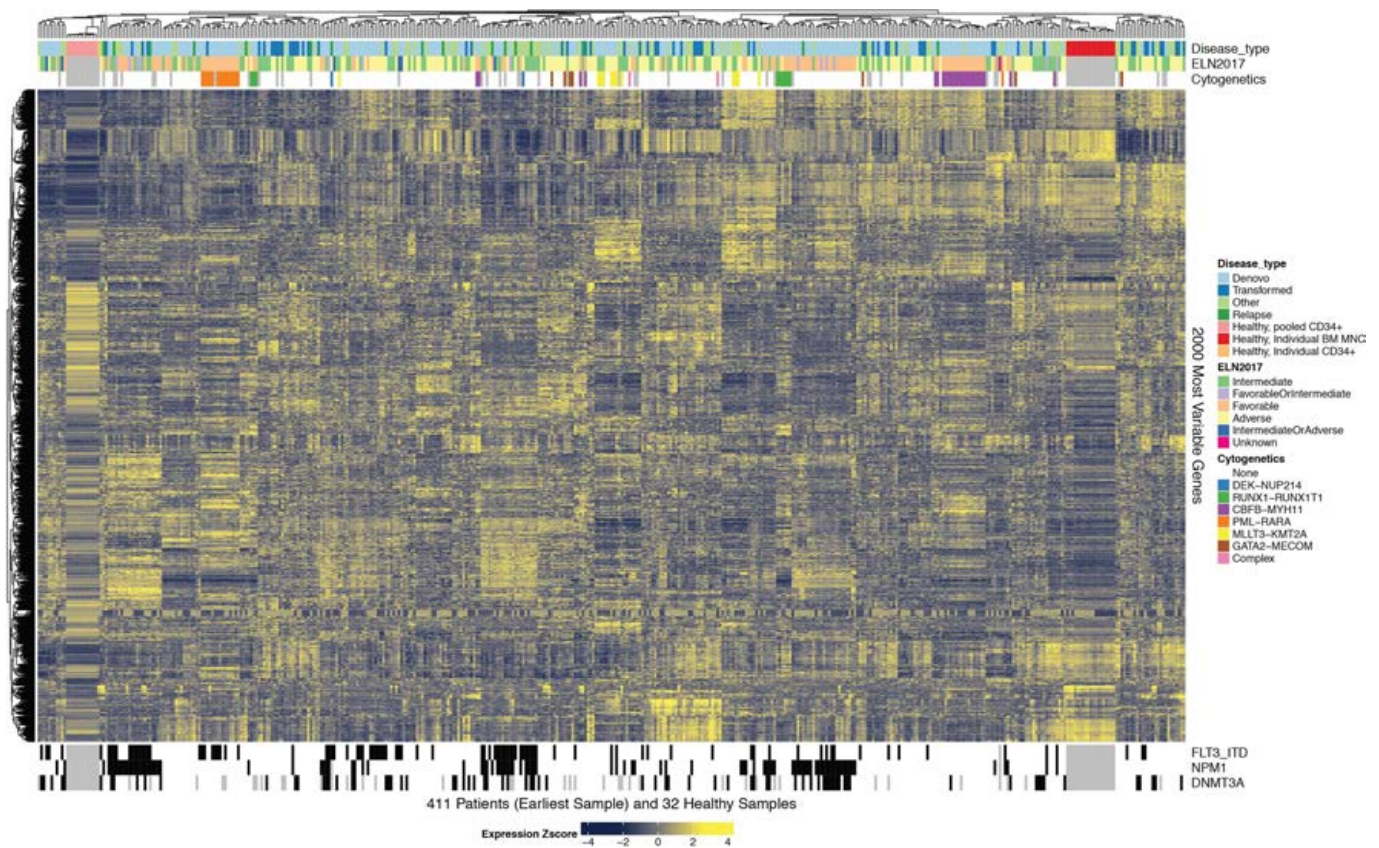
53. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
54. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
55. Costello, M. et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
56. Kottaridis, P. D. et al. The presence of a *FLT3* internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood* **98**, 1752–1759 (2001).
57. Döhner, K. et al. Mutant nucleophosmin (*NPM1*) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* **106**, 3740–3746 (2005).
58. Falini, B., Nicoletti, I., Martelli, M. F. & Mecucci, C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc⁺ AML): biologic and clinical features. *Blood* **109**, 874–885 (2007).
59. Huang, Q. et al. A rapid, one step assay for simultaneous detection of *FLT3*/ITD and *NPM1* mutations in AML with normal cytogenetics. *Br. J. Haematol.* **142**, 489–492 (2008).
60. Wouters, B. J. et al. Double *CEBPA* mutations, but not single *CEBPA* mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* **113**, 3088–3091 (2009).
61. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
62. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
63. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
64. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
65. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).
66. Langfelder, P. & Horvath, S. Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**, 1–17 (2012).
67. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
68. Parsana, P. et al. Addressing confounding artifacts in reconstruction of gene co-expression networks. Preprint at <https://www.biorxiv.org/content/early/2017/10/13/202903> (2017).
69. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
70. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
71. Zheng, X. & Weir, B. S. Eigenanalysis of SNP data with an identity by descent interpretation. *Theor. Popul. Biol.* **107**, 65–76 (2016).
72. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
73. Slovak, M. L., Theisen, A. & Shaffer, L. G. in *The Principles of Clinical Cytogenetics* (eds Gersen, S. L. & Keagle, M. B.) 23–49 (Springer, New York, 2013).
74. Kurtz, S. E. et al. Molecularly targeted drug combinations demonstrate selective effectiveness for myeloid- and lymphoid-derived hematologic malignancies. *Proc. Natl Acad. Sci. USA* **114**, E7554–E7563 (2017).
75. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–1051 (2011).
76. Blucher, A. S., Choonoo, G., Kulesz-Martin, M., Wu, G. & McWeeney, S. K. Evidence-based precision oncology with the cancer targetome. *Trends Pharmacol. Sci.* **38**, 1085–1099 (2017).
77. Gu, Z., Eils, R. & Schlesner, M. Complex heat maps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
78. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
79. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, e161 (2007).
80. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
81. Parker, H. S., Corrada Bravo, H. & Leek, J. T. Removing batch effects for prediction problems with frozen surrogate variable analysis. *PeerJ* **2**, e561 (2014).
82. Fraley, C. & Raftery, A. E. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *J. Classif.* **20**, 263–286 (2003).
83. Pison, G., Struyf, A. & Rousseeuw, P. J. Displaying a clustering with CLUSPLOT. *Comput. Stat. Data Anal.* **30**, 381–392 (1999).
84. Wei, T. et al. corrplot: Visualization of a Correlation Matrix. R package version 0.84 <https://github.com/taiyun/corrplot> (2017).
85. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
86. Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).



Extended Data Fig. 1 | Genomic landscape of the Beat AML cohort.

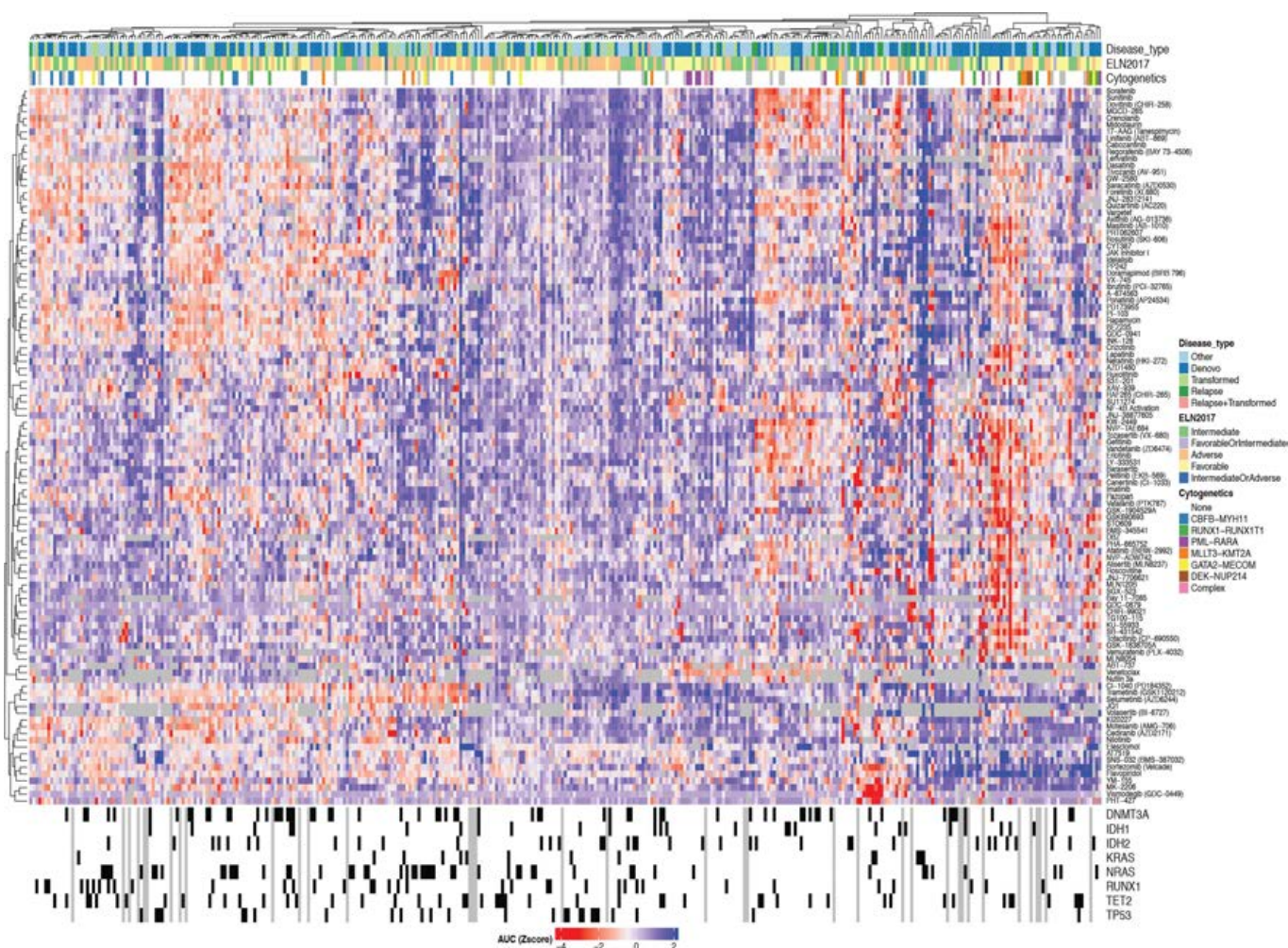
In total, 622 specimens from 531 patients were used for whole-exome sequencing. Automated and manual curation steps (described in the Methods, Supplementary Information and at http://vizome.org/additional_figures_BeatAML.html) were used to obtain a final set of high-confidence variants (annotated in Supplementary Table 7) and the earliest sample for each individual patient was used in this analysis. Clinical cytogenetics and gene fusion calls from RNA sequencing were used to curate recurrent gene rearrangements (Supplementary Information). The mutational profile for each patient is shown for frequency-ranked mutational events (top) and frequency-ranked gene rearrangements (bottom). The mosaic plot is annotated with clinical features of each case, such as diagnosis or relapse and de novo or transformed disease states, and the first bar chart on the right summarizes the cohort frequencies of mutational and gene rearrangement events. The last bar chart on the right

summarizes the frequency of significant drug-mutation associations for the given gene across the cohort with drug sensitivity displayed in red and drug resistance displayed in blue. Eleven genes that have not previously been reported to be somatically mutated in cancer were observed with mutations at approximately 1% cohort frequency: CUB and Sushi multiple domains 2 (*CSMD2*), NAC alpha domain containing (*NACAD*), teneurin transmembrane protein 2 (*TENM2*), aggrecan (*ACAN*), ADAM metalloproteinase with thrombospondin type 1 motif 7 (*ADAMTS7*), immunoglobulin-like and fibronectin type III domain containing 1 (*IGFN1*), neurobeachin-like 2 (*NBEAL2*), poly(U) binding splicing factor 60 (*PUF60*), zinc-finger protein 687 (*ZNF687*), cadherin EGF LAG seven-pass G-type receptor 2 (*CELSR2*) and glutamate ionotropic receptor NMDA type subunit 2B (*GRIN2B*). For the number of samples used to correlate each drug with mutations, see Supplementary Table 17.



Extended Data Fig. 2 | Transcriptomic landscape of the Beat AML cohort. In total, 451 specimens from 411 patients with AML were used for RNA-sequencing analyses. The 2,000 genes with the greatest differential expression across these patients with AML are displayed as a heat map.

The heat map is annotated with disease type, ELN risk stratification groups, and genetic and cytogenetic features of disease as indicated in the key.



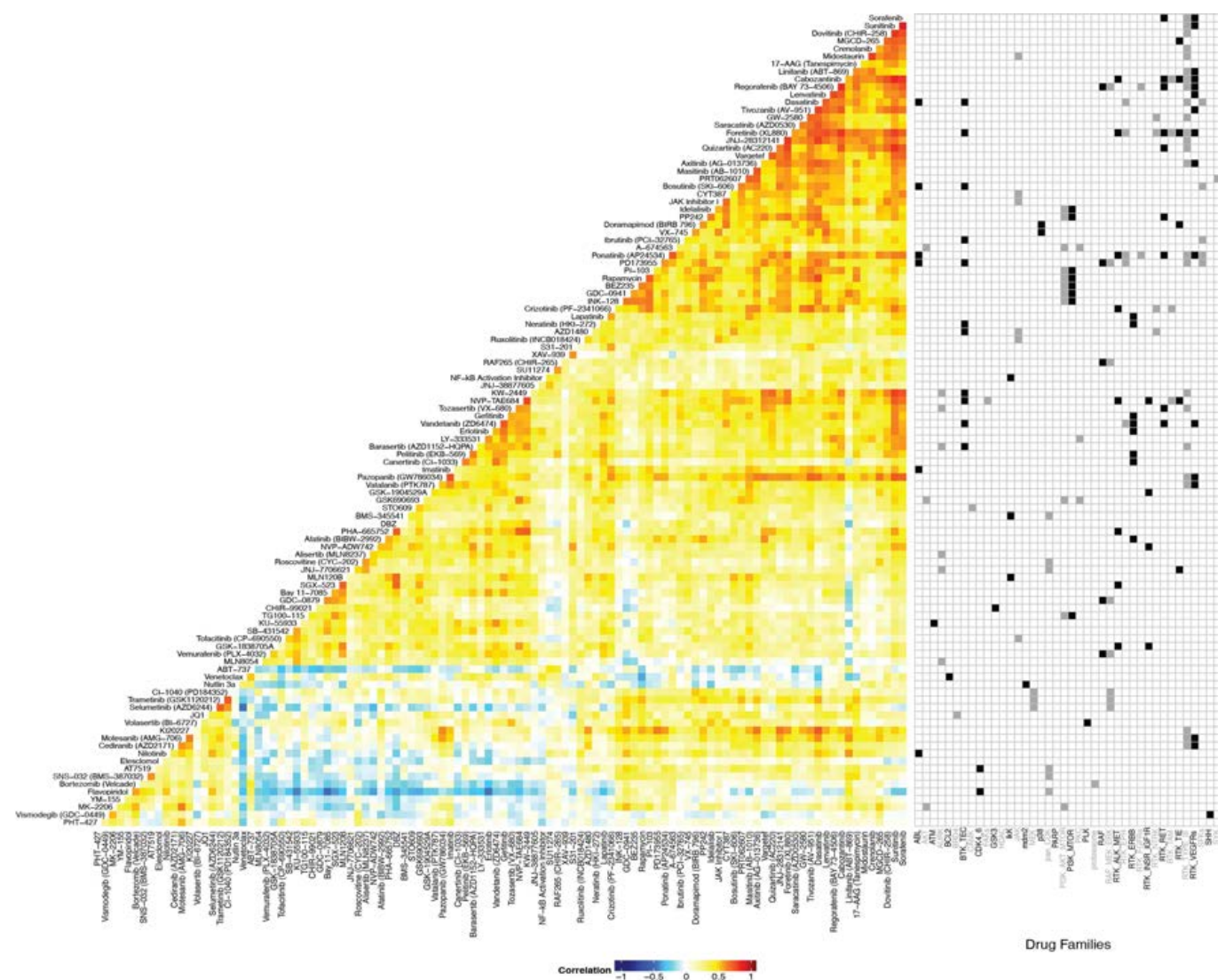
Extended Data Fig. 3 | Functional drug sensitivity landscape of the Beat AML cohort. In total, 409 specimens from 363 patients with AML were subjected to an ex vivo drug sensitivity assay, in which freshly isolated mononuclear cells from blood or bone marrow of patient specimens were incubated with graded concentrations of 122 small-molecule inhibitors (seven dose points in addition to the no drug control). Probit curve fits

were used to compute drug-response metrics, and the z score of area under the dose-response curve is plotted for each individual patient specimen against each drug. Drug sensitivity (blue) and resistance (red) are annotated by a colour gradient, with grey indicating no drug data available. The heat map is annotated at the top and bottom with major clinical, cytogenetic and genetic features of disease as indicated in the key.



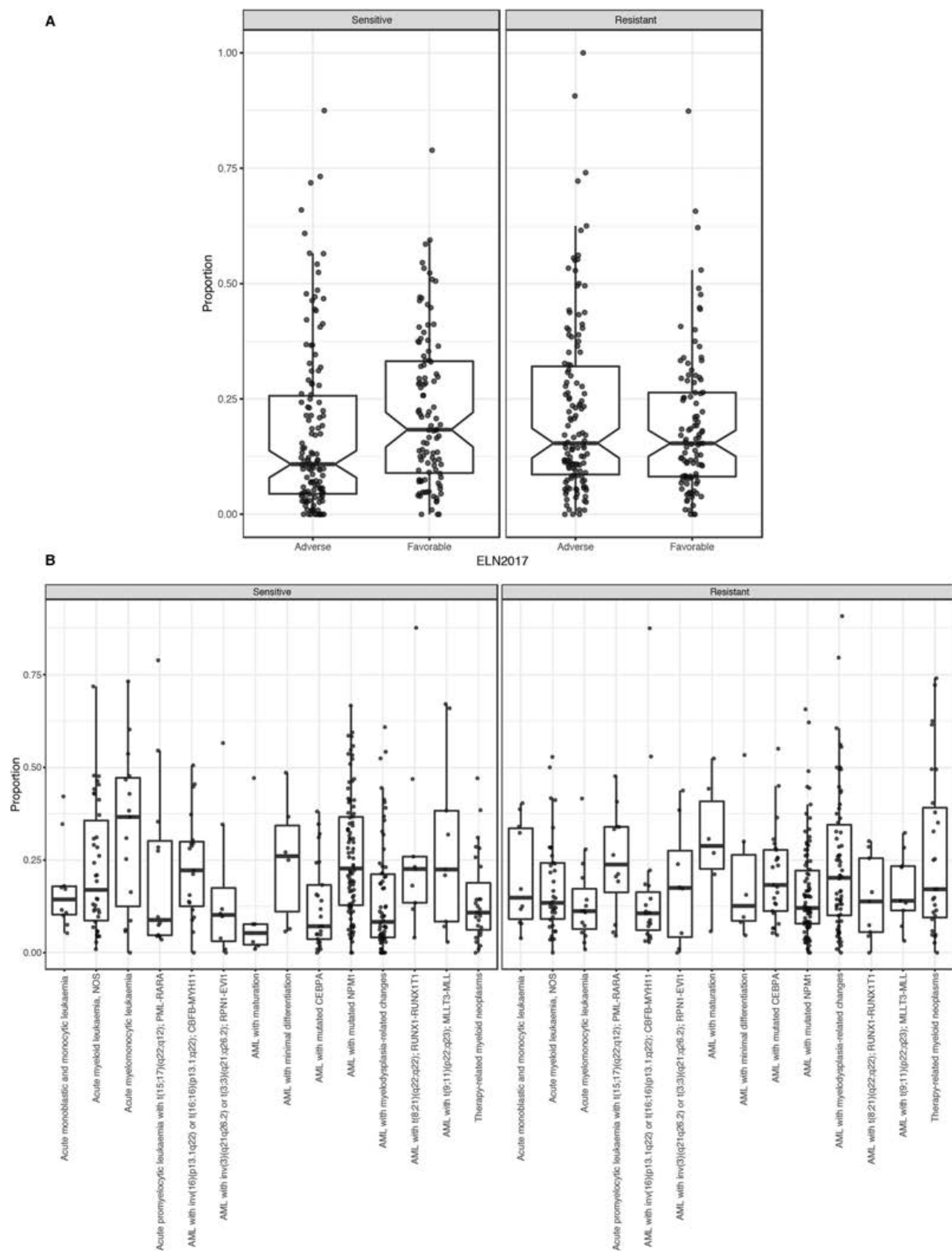
Extended Data Fig. 4 | Drug response in de novo versus transformed AML cases. The average inhibitor response AUCs for all cases that were de novo ($n = 288$) versus all cases that transformed from a background of myelodysplastic syndromes ($n = 111$) were compared for every inhibitor that had at least three cases with evaluable data in each group. The middle

point represents the average difference in AUC between the two groups with the bars representing the 95% confidence interval. For the sample size and statistical results of each drug-sample group correlation, see Supplementary Table 20.



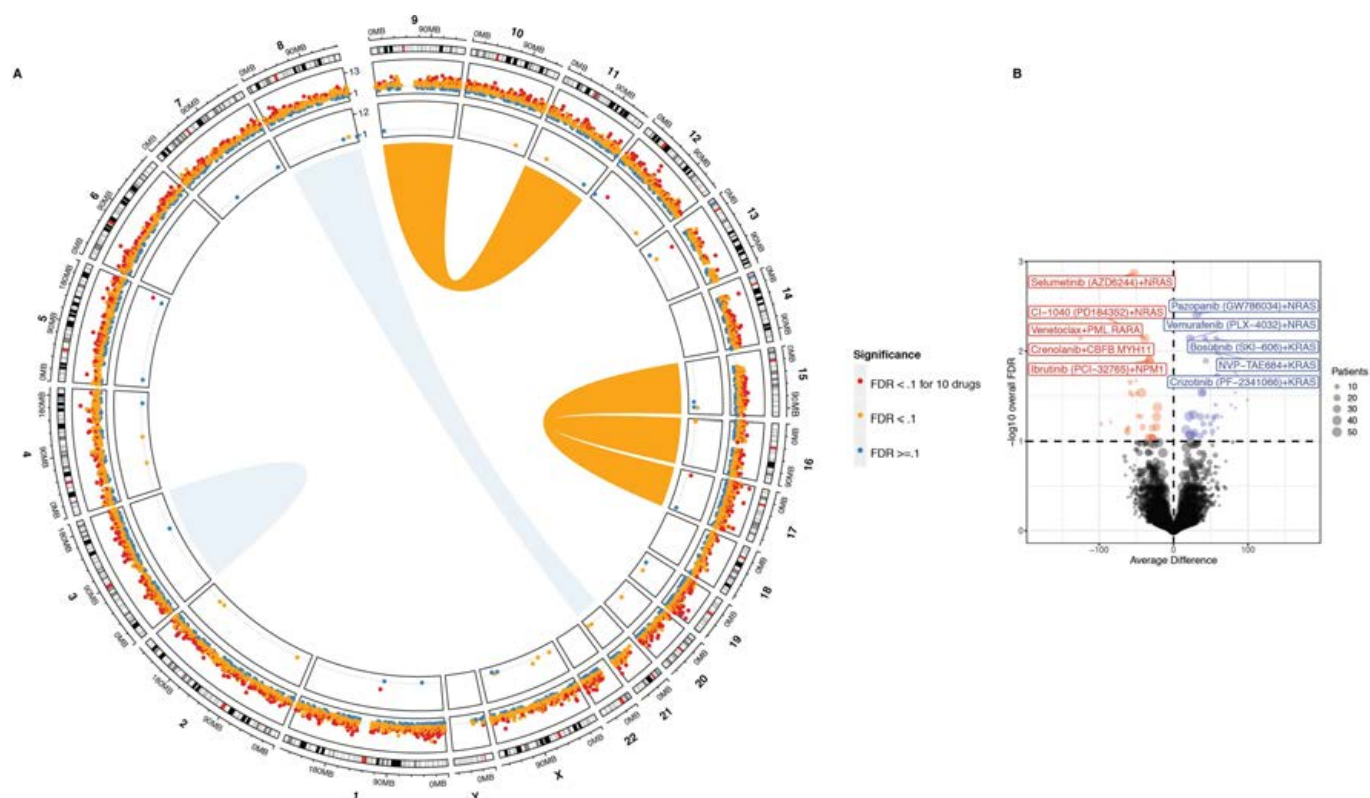
Extended Data Fig. 5 | Pairwise drug sensitivity correlations and association with drug family. To understand patterns of small-molecule sensitivity against prior annotations of the gene and pathway targets of each drug, drugs were placed into drug families according to target genes and/or pathways and the Pearson's correlation value of each drug was plotted onto a clustered heat map, showing drugs with similar or dissimilar

patterns of sensitivity across the patient cohort. Annotations based on prior knowledge of the drug families to which each drug could be assigned are shown to the right of the heat map with alternating black and grey boxes and labels used to aid in tracking. Descriptions of each drug family as well as the number of samples used to calculate each pairwise drug correlation are found in Supplementary Tables 11, 21.



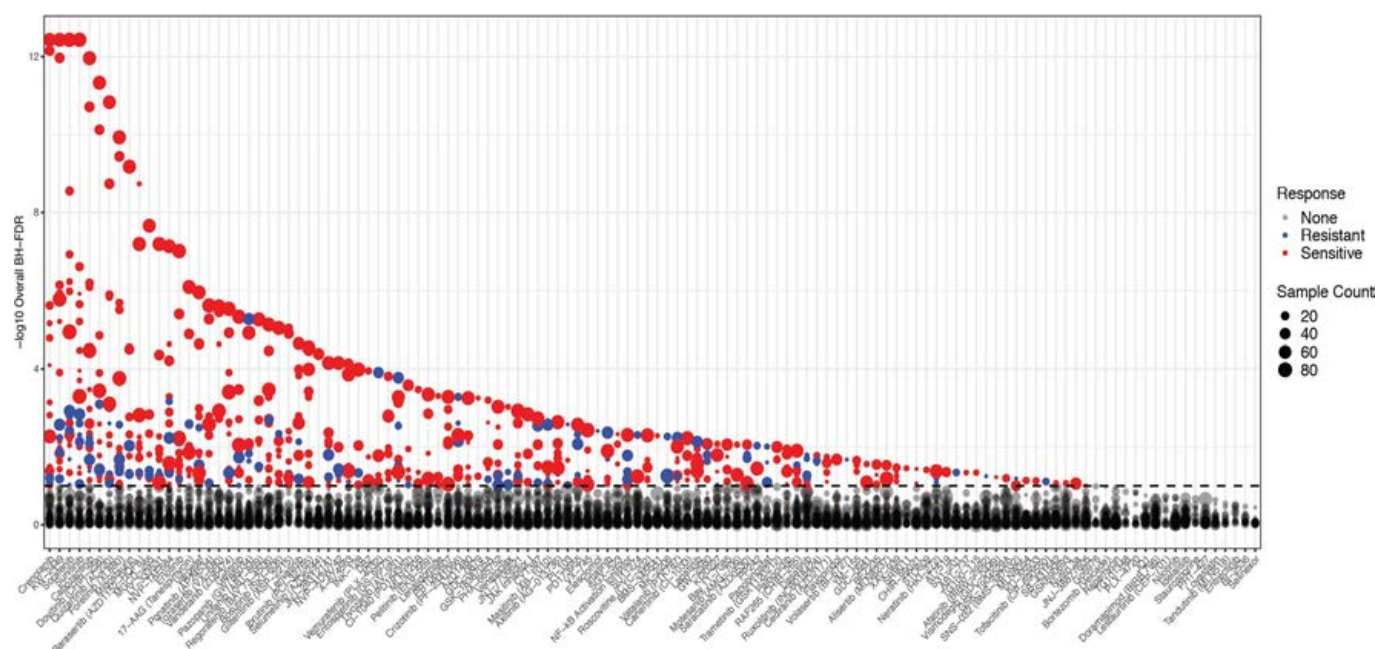
Extended Data Fig. 6 | Binary drug response calls and correlation with clinical subsets. **a**, For the intersect of every specimen with evaluable response data for each inhibitor, we created a threshold for binary sensitive or resistant calls based on whether the individual specimen response fell within the most sensitive 20% of all specimens tested against that drug. A matrix plot showing the unsupervised clustering of the binary calls can be found at http://vizome.org/additional_figures_BeatAML.html. The binary drug-resistance calls for each specimen were combined into a single value, representing the proportion of drugs to which an individual specimen was sensitive (left) or resistant (right). Specimens were divided into 'Favourable' and 'Adverse' groups based on ELN 2017 criteria to

determine whether overall drug sensitivity or resistance correlated with prognostic features of disease ($n = 233$ patients). **b**, The binary drug-resistance calls for each specimen as in **a**. Specimens were divided into diagnostic groups based on WHO 2016 categories to determine whether overall drug sensitivity or resistance correlated with cytogenetic or morphologic features of disease ($n = 340$ patients). **a**, **b**, The top and bottom points of the box plots show 1.5 times the interquartile range (IQR) from the upper and lower lines; the top, middle and bottom lines indicate the 75th, median and 25th percentile of the data with the notches extending $1.58 \times \text{IQR}/(\sqrt{n})$. Specific sample sizes of each group are reported in Supplementary Table 22.



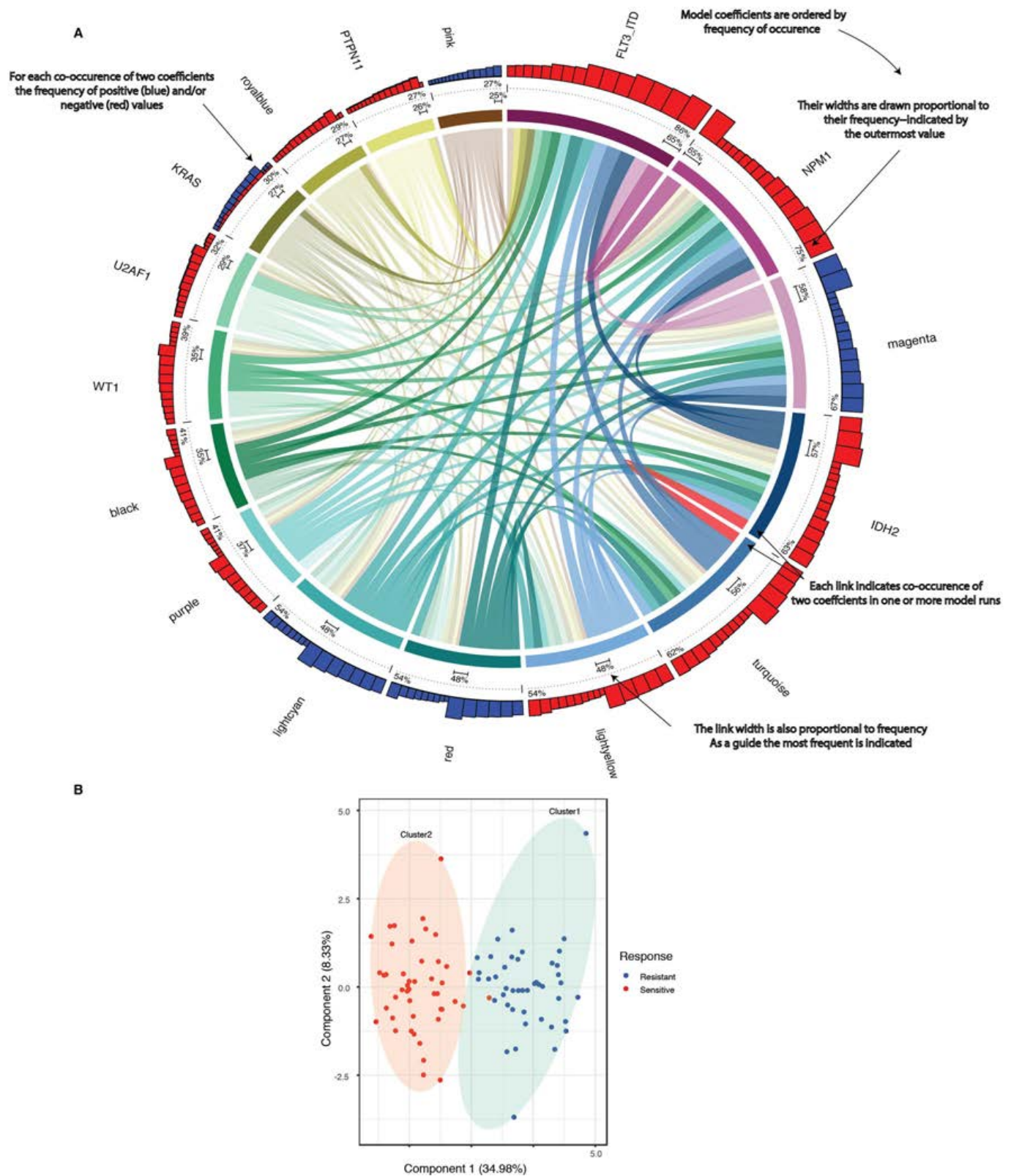
Extended Data Fig. 7 | Integration of genetic events with drug sensitivity. **a**, Circos plot showing AML rearrangements in the centre, mutational events in the next concentric ring, and gene expression events in the outer ring. The size and width indicates frequency of the event and the FDR-corrected Q value of association with drug sensitivity is colour-coded (sensitivity (red); resistance (blue)). For each gene, tests involving expression were two-sided Student's t -tests (linear model) of the differences between sensitive and resistant samples. For mutational events, the average difference in AUC between mutant and wild-type samples was determined using two-sided Student's t -tests from a linear model as shown in Fig. 2a. FDR was computed using the Benjamini–Hochberg method over all the drugs. The number of samples used to correlate each

mutational event with drug sensitivity is reported in Supplementary Table 17. **b**, As in Fig. 2a, the average difference in AUC drug response between mutant and wild-type cases was determined using a two-sided Student's t -test from a linear model fit (plotted on the x axis and the FDR-corrected Q value is plotted on the y axis). This analysis shows only FLT3-ITD-negative cases. FDR was computed using the Benjamini–Hochberg method over all the drugs. The number of samples used to correlate each mutational event with drug sensitivity is reported in Supplementary Table 17. Expanded and interactive plots are available in our online data browser (<http://www.vizome.org/> and http://vizome.org/additional_figures_BeatAML.html).



Extended Data Fig. 8 | Integration of drug sensitivity with genetic events. Correlation between drug sensitivity and mutational events. The average difference in AUC drug response between mutant and wild-type cases was determined using a two-sided Student's *t*-test from a linear model fit. FDR was computed using the Benjamini–Hochberg method

over all the drugs. The degree of significance is represented on the *y* axis (sensitivity (red); resistance (blue)). The number of samples used to correlate each mutational event with drug sensitivity is reported in Supplementary Table 17.



Extended Data Fig. 9 | Functional drug sensitivity landscape of the Beat AML cohort. a, Co-occurrences with regard to WGCNA gene expression clusters and/or mutational events (coefficients) were detected by multivariate modelling with respect to ibrutinib response (resistance (blue); sensitivity (red)) and the degree of correlation is shown in the stacked bar plot (top). All coefficients that appear in 25% of the bootstrapped sample sets are shown as segments of the circle. Segment width (the coloured ring) corresponds to the percentage of bootstrapped samples in which that coefficient appears (quantified above the dotted line). The variables appear in descending order clockwise starting at

12 o'clock. Each link indicates pairwise co-occurrence of mutational events and gene expression clusters (width represents frequency of the co-occurrence). The largest co-occurrence for each coefficient is quantified. **b,** The capacity of differential gene expression to distinguish the 20% most ibrutinib-sensitive ($n = 46$) from 20% most resistant ($n = 44$) specimens is shown on a principal component plot ($n = 239$ patient samples were tested for ibrutinib sensitivity and RNA sequencing). For the number of samples used to correlate each drug with gene expression and perform LASSO regression, see Supplementary Table 17.

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 qrc; 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 -l 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 (WGCNA); 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; corplot; 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) $\geq 8\%$ and normal VAF $\leq 5\%$ 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 (WGCNA). 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= $\geq .9$; RNA-RNA= $\geq .83$; DNA-RNA= $\geq .89$) 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:

$$(\text{normalized_viability} / 100) \sim 1 + \log_{10}(\text{concentration})$$

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-occurrence. A plot of the co-occurrences was generated using corplot 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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> 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.