Acute myeloid leukemia (AML) is a heterogeneous disease caused by a variety of alterations in transcription factors, epigenetic regulators and signaling molecules. To determine how different mutant regulators establish AML subtype-specific transcriptional networks, we performed a comprehensive global analysis of cis-regulatory element activity and interaction, transcription factor occupancy and gene expression patterns in purified leukemic blast cells. Here, we focused on specific subgroups of subjects carrying mutations in genes encoding transcription factors (RUNX1, CEBPα), signaling molecules (FLT3-ITD, RAS) and the nuclear protein NPM1). Integrated analysis of these data demonstrates that each mutant regulator establishes a specific transcriptional and signaling network unrelated to that seen in normal cells, sustaining the expression of unique sets of genes required for AML growth and maintenance.

Acute myeloid leukemia (AML) is characterized by blocked myeloid lineage differentiation and accumulation of leukemic blast cells. It is a highly heterogeneous disease caused by different types of genetic alterations that affect signaling pathways as well as transcriptional and epigenetic regulators1–3. Recurrent alterations include loss-of-function substitutions in transcription factors controlling hematopoietic development, such as RUNX1, GATA2 or C/EBPα (ref. 4), and gain-of-function substitutions in signaling molecules such as FLT3, KIT, JAK2 and NRAS, which regulate inductive transcription factors such as NF-κB, STAT or AP-1 family members5–7. The most common FLT3 alterations are internal tandem duplications (FLT3-ITD), which give rise to a constitutively active growth factor receptor8,9 and often occur together with nucleophosmin 1 (NPM1) alterations. Another major group of mutations alters genes encoding epigenetic and chromatin regulators10,11 that have widespread roles in development and differentiation by controlling establishment, maintenance and extinction of lineage-specific gene expression programs. These include regulators of histone and DNA methylation such as MLL, EZH2, TET2, DNMT3A, IDH1 and IDH2 (refs 11–17). In normal cells, all common mutation targets cooperate to control the finely balanced gene expression changes that are essential for cell differentiation and lineage commitment.

Transcription factors interact with defined target gene sequences and recruit epigenetic regulators to program specific chromatin states and mediate the coordinated activation or deactivation of cis-regulatory elements driving gene expression18,19. Distal cis-regulatory elements interact directly with promoter elements, an arrangement that is both dynamic and robust20,21. From global studies examining a few selected types of AML, we know that their gene expression patterns and epigenetic landscapes differ from those in normal cells22–27. However, how the disruption of specific transcription factor activity leads to a specific pattern of aberrant chromatin programming and changes in gene expression in AML has remained unclear. It has not been established at the global level which cis-regulatory elements are affected in activity in different types of AML, how their activity is altered in subjects carrying specific transcription factor and signaling alterations, or which factors maintain their transcriptional networks.

Here we addressed these questions by collecting transcriptome, digital footprinting and chromatin conformation capture data from purified leukemic blasts from AML subjects with defined transcription factor and signaling molecule alterations and defined the components of AML subtype-specific regulatory circuitries. Our study provides a comprehensive resource on the transcriptional networks of different AML subtypes, highlighting pathways required for tumor maintenance.

Results

AML subtypes adopt unique chromatin landscapes. To examine how specific transcription factor and signaling substitutions alter the epigenome of AML, we purified leukemic blast cells from bone marrow or peripheral blood samples from AML subjects (Fig. 1a). After determining the mutation status (Supplementary Table 1), we

1Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. 2Northern Institute for Cancer Research, University of Newcastle, Newcastle, UK. 3Section of Experimental Haematology, Leeds Institute for Molecular Medicine, University of Leeds, Leeds, UK. 4West Midlands Regional Genetics Laboratory, Birmingham Women’s NHS Foundation Trust, Birmingham, UK. 5CMT Laboratory NHS Blood & Transplant, Edgbaston, Birmingham, UK. 6Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands. 7Oncode Institute, Erasmus MC, Erasmus University Medical Center, Rotterdam, The Netherlands. 8Haematological Malignancy Diagnostic Service, St. James’s University Hospital, Leeds, UK. 9Centre for Clinical Haematology, Queen Elizabeth Hospital, Birmingham, UK. 10Princess Maxima Centrum for Pediatric Oncology, Utrecht, The Netherlands.

These authors contributed equally: S. A. Assi, M. R. Imperato, D. J. L. Coleman. *e-mail: p.n.cockerill@bham.ac.uk; c.bonifer@bham.ac.uk
selected a cohort of subjects with defined mutations, which included RUNX1 mutations affecting RUNX1 DNA binding (D-type) or lacking the trans-activating domain (T-type); t(8;21) translocations fusing the DNA-binding domain of RUNX1 to the co-repressor ETO; inv(16), which fuses CBFβ to smooth muscle myosin heavy chain 11 protein; mutations of both alleles of the CEBPA gene (CEBPA (×2)
whereby one mutation leads to loss of DNA-binding activity \(^{29}\) and FLT3-ITD with or without NPM1 mutations. We performed RNA-seq (Supplementary Fig. 1a) and high-read-depth DNase I–seq (Fig. 1b) to map DNase I hypersensitive sites (DHHS) on 29 samples comprising seven major groups, and at least one analysis on 12 additional samples, carrying mutations of genes such as NDRG, CBL, JAK2, SRSE2 or the inv(3) translocation (Supplementary Table 1).

One subject with non-Hodgkin lymphoma carried a RUNX1 mutation (RUNX1-T-7). Samples were compared to CD34\(^+\) mobilized peripheral blood stem cells (PBSCs) from two healthy individuals and to cord blood CD34\(^+\) cells. To provide the community with a data resource, we established an online database containing multiple data sets including a genome browser (see Data Availability).

Unsupervised clustering revealed that distal DHHS formed different groups according to their alteration class (Fig. 1c). Samples with FLT3-ITD and/or NPM1 alterations represented one major group with subclusters for subjects with NPM1 alterations or carrying two FLT3-ITD alleles, but excluding a FLT3-ITD subject carrying a RUNX1 mutation. DHHSs from the t(8;21), inv(16) and CEBPA double mutant subjects clustered as discrete groups within a larger group, indicating that these mutations affect similar pathways.

Examples of these patterns can be seen in Supplementary Fig. 1b. DHHSs from subjects with RUNX1 alterations were more heterogeneous and clustered with the PBSCs and the inv(3) subjects. The non-Hodgkin lymphoma (RUNX1-T-7) and NPM1/RAS-3 patterns were unrelated to any of the others. We further validated our findings by analyzing an independently derived published data set based on the assay for transposase-accessible chromatin using sequencing (ATAC-seq)\(^{25}\), confirming that alterations in FLT3 underpin one major component of the clustering (Supplementary Fig. 2a).

In contrast, the presence or absence of mutations of genes encoding epigenetic regulators such as DNMT3A did not influence chromatin accessibility levels (Supplementary Fig. 2b) or gene expression (data not shown). Our mutation analyses showed no indication of the presence of confounding major subclones in purified undifferentiated AML cell populations, as mutations were present at close to either 50% or 100% allele frequency (Supplementary Data 1).

Unsupervised clustering analysis of RNA-seq data from the same subjects as well as direct comparisons between individual samples (Supplementary Figs. 1c and 2c) revealed strong correlations between mutation-specific chromatin landscapes and mutation-specific differential gene expression, as exemplified by RUNX1 and FOXC1, where mRNA patterns correlated well with chromatin profiles (Supplementary Fig. 1d). We identified distinct patterns of expression for specific transcription factor–encoding genes in different AML types (Supplementary Fig. 2d). For example, various homeodomain gene family members (HOX, NKX, IRX and PBX families) were upregulated in the FLT3-ITD- and NPM1-altered subjects. Our comparative analyses show that aberrant transcription factors and chronic signaling impose distinct alteration-specific patterns of chromatin accessibility and gene expression, irrespective of the presence of additional mutations.

AML-specific clustering of distal cis-regulatory elements. We next examined the active cis-regulatory elements specific for each AML subtype by defining the union of all AML-specific DHHS as compared to CD34\(^+\) PBSCs and performing k-mean clustering to identify unique and common DHHS shared among subjects; this identified 20 distinct DHS clusters (Fig. 2a). Less than half of these DHHS were found in any of the Corces et al. progenitor ATAC-seq data sets\(^{25}\) and the percentage overlap varied substantially between clusters (range of 2–40%). We verified mutation-specific clustering behavior of our samples by comparing them with a recently published AML histone H3K27 acetylation data set\(^{26}\) showing similar patterns for FLT3-ITD, RUNX1 and CEBPA double mutations (Supplementary Fig. 3d). We also defined mutation-specific groups of deregulated DHHS that were shared between the specific members of each of the seven major mutation groups defined in Supplementary Table 1 (Supplementary Fig. 4a,b) which were distributed between both the mutation-specific clusters and the shared clusters (Fig. 2b) and were associated with differentially expressed genes (Supplementary Data 2). Again, the t(8;21), inv(16) and CEBPA groups showed similar patterns whereby 914 upregulated DHHS were shared among the three groups (Supplementary Fig. 4a).

The FLT3-ITD, FLT3-ITD/NPM1 and NPM1 alteration groups showed substantial overlap with 942 shared DHHS, and with only 19% of these DHHS included in the 914 FLT3-ITD/NPM1-specific group. These AML-specific patterns showed little similarity to normal myeloid differentiation as the majority of these specific sites were not upregulated in granulocyte-monocyte progenitor cells relative to PBSCs (Supplementary Note Fig. SN2b).

The presence of specific DHHS was strongly correlated with the upregulation of their associated genes (Supplementary Data 3 and Supplementary Fig. 4c), as exemplified by two DHHS at POU4F1 (Supplementary Fig. 4d). Supplementary Fig. 4e shows examples of AML type–specific upregulated transcription factor genes, and growth factor or receptor genes, that were associated with AML type–specific DHHS (Supplementary Data 4). The gene expression patterns of such genes were validated using publicly available data sets (Supplementary Fig. 5).

AML-specific transcription factor binding patterns. To identify transcription factors associated with different chromatin patterns, we analyzed high-read-depth DNase I–seq data using our Wellington digital footprinting algorithm\(^{27}\). Because closely related factors recognize identical sequences, we compiled a nonredundant database of motifs (Supplementary Table 2) and selected representative motifs encompassing each transcription factor family, as defined in more detail on our web server (see URLs). Examples of footprints are depicted for NFI and ETS motifs at the MDFI locus in FLT3-ITD/NPM1-altered AML (Fig. 3a) and for RUNX, NFAT and C/EBP motifs at the C3AR1 locus in t(8;21) and CEBPA-altered AML (Supplementary Fig. 6a). The majority of AML type–specific DHHS within the 20 AML–specific DHS clusters contained footprints (Supplementary Fig. 6b). For validation, we compared RUNX motif footprints with publicly available RUNX1 ChIP data from our studies (FLT3-ITD/NPM1, ref. 21; t(8;21); ref. 22) and others (inv(16), ref. 23) (Supplementary Fig. 6c). Some 60–85% of footprinted RUNX motifs occurred in regions shown to bind RUNX1.

We next evaluated occupied motif enrichment in the 20 AML subtype–specific DHS clusters (Fig. 3b), and found that motif occupancy patterns were highly AML type–specific. For example, the FLT3-ITD/NPM1-specific clusters 5 and 19 are enriched for occupied HOX, Fox/E-box and NFI motifs and correlate with upregulation of FOXC1, NFIX and multiple homeodomain genes (Supplementary Fig. 2d). Occupied AP-1 motifs are enriched in multiple clusters (01, 05, 07, 12, 13, 18 and 19), many of which were found in AML with signaling mutations. Because AP-1 factors mediate MAP kinase (MAPK) signaling, this enrichment indicates widespread activation of this signaling pathway in FLT3-ITD AML\(^{24}\) and other AML types. Finally, we observed significant POU4F1 motif occupancy in clusters 02 and 20, which contained samples from t(8;21) and CEBPA double mutant subjects, but nowhere else (Fig. 3b). POU4F1 is aberrantly expressed in t(8;21) cells\(^{26}\), but has so far not been linked to CEBPA double mutations. We observed a similar differential occupancy pattern when footprints were clustered according to mutation-specific groups of DHHS (Supplementary Fig. 6d). C/EBP motifs in AMLs with CEBPA double mutations showed no reduction in overall motif occupancy, suggesting compensation by other C/EBP family members.

To examine the position of transcription factor occupancy patterns within the hematopoietic hierarchy, we determined the
correlation of the presence of footprints specific for AML subtypes with accessible chromatin regions present in precursor cells (Fig. 4)25. This analysis revealed unique factor occupancy patterns in AML cells compared to normal progenitors. For example, HOX motifs within open chromatin regions observed in hematopoietic stem cells (HSCs), multipotent progenitor (MPP) cells and megakaryocyte-erythroblast progenitor (MEP) cells are occupied in the FLT3-ITD/NPM1 and RUNX1 groups, but not in the t(8;21) group, indicating an early block in differentiation (see also Supplementary Note Fig. SN1b). Many of the samples, including NPM1, FLT3-ITD/NPM1 and t(8;21) cells, display high AP-1 motif occupancy which is normally only seen in monocytes. POU4F1 is expressed in HSCs, multipotent progenitor cells (MPPs), megakaryocyte-erythroblast progenitor cells (MEPs) and common lymphoid progenitor cells (CLPs)25 and its binding motifs are occupied in t(8;21) and CEBPA double mutant cells, yet these AML cells also show strong occupancy of C/EBP motifs, which is normally a hallmark of granulocyte-monocyte progenitor cells and monocytes. In summary, our digital footprinting analysis shows that (1) each AML subtype uses a different combination of factors to bind to elements shared with different types of precursor cells and (2) lineage-unrelated expressed transcription factors such as FOXC1, NFIX and POU4F1 participate in such cooperation.

AML subtype–specific cis-element interactions. The construction of gene regulatory networks relies on linking cis-regulatory elements to their respective promoters33. We therefore examined (1) whether the differential activity of cis-regulatory elements in AML subtypes led to the formation of alternate cis-element interactions, and (2) which transcription factor families were involved in such interactions, by using promoter-capture chromosomal structure analysis (CHi-C)34. We analyzed cells from relapse subject t(8;21)-1R (Supplementary Table 1), which maintained a gene regulation network similar to the presentation sample t(8;21)-1 (Fig. 1c and Supplementary Figs. 1c and 7a), and from a subject carrying a FLT3-ITD/NPM1 alteration (ITD/NPM1-2, Supplementary Table 1).
We compared these data to a data set derived from human CD34+ cells\(^{14}\). Intrachromosomal interactions did not differ at the global level (Supplementary Fig. 7b) and the organization into topologically associated domains containing the DHSs was unaffected by the type of AML (Supplementary Fig. 7c).

The proportion of DHSs involved in AML subtype–specific interactions varied between DHS clusters (Supplementary Fig. 7d), and ~40% of all promoters showing differential interactions were associated with expressed genes (Supplementary Fig. 7e,f). A direct comparison between the CHi-C data from the two subjects (Supplementary Fig. 7g) showed that (1) differential interactions were correlated with differential DHS patterns and (2) a differential set of genes was expressed with different gene ontology (GO) terms (Supplementary Data 5) as exemplified by the KLF2 gene, which is differentially expressed in FLT3-ITD, t(8;21) and CD34+ cells (Supplementary Fig. 7h).

On average 80% of all DHSs mapped in t(8;21) AML, FLT3-ITD AML and CD34+ cells participated in interactions (Supplementary Fig. 8a). An average of 17% of interactions were AML type–specific and not present in CD34+ PBSCs (Supplementary Fig. 8b). To identify the transcription factor families involved in regulating differential interactions we determined the proportions of enriched occupied motifs in DHSs underlying these interactions (Supplementary Fig. 8c). These analyses suggest that hematopoietic transcription factors such as RUNX, ETS and C/EBP family members and transcription factors specifically expressed in AML subtypes participate in...
Motifs

compared to CD34

DHSs participating in interactions were enriched within related groups, but not within unrelated groups (Supplementary Data 5). Because ~83% of the DHSs involved in differential interactions in both AML types, together with the AP-1 factor family. In FLT3-ITD AMLs this included HOX proteins and factors occupying FOX/E-box motifs. In the (t(8;21)) AMLs this included proteins binding to FOX and POU4F1 motifs.

Differential interactions drive AML-specific gene expression.

The majority of DHSs underlying interactions among the three data sets and those of individual subjects were shared with an average of 80% overlap (Supplementary Fig. 8d), confirming that the global transcriptional network of related cells is highly related35,36. Subtype-specific DHSs participating in interactions were enriched within related groups, but not within unrelated groups (Supplementary Fig. 8e), confirming that the two subjects were representative of those groups. For both the FLT3-ITD/NPM1 and the (t(8;21)) sample the nearest promoter accounted for 65–74% of AML type–specific interactions associated with genes that are upregulated in AML compared to CD34+ cells (Fig. 5a). We observed similar results for each of the 20 DHS clusters (Supplementary Fig. 8f).

To integrate differential interaction data, digital footprinting data and gene expression data, we assigned the respective DHSs to their interacting promoters as described in Fig. 5b. GO-term and Kyoto Encyclopedia of Genes and Genomes (KEGG)-pathway analyses of expressed genes in the two types of AML (Fig. 5c–f) revealed an AML subtype–specific core signature of genes driven by specific cis-regulatory elements (for an extended gene list see Supplementary Data 6). For both AML samples these included genes involved in regulating pro-inflammatory pathways. FLT3-ITD cells also displayed an activated MAPK signaling signature, whereas the (t(8;21)) signature also included RAP, RAS, PI3K and FOXO signaling genes. FOXO1 is part of the (t(8;21)) preleukemic maintenance program34. More than 50% of all genes within these pathways were targets of RUNX1-ETO (Fig. 5g)35, thereby linking them to the actual driver mutation. A similar percentage of the genes within the FLT3-ITD/NPM1 core pathway are bound by RUNX1 (Fig. 5g), which is upregulated in FLT3-ITD24 (Supplementary Data 5). Because ~83% of the DHSs involved in interactions in each of the three samples (Supplementary Fig. 8g) were shared, we merged all three CHI-C data sets and used these data to assign the DHSs from the 20 clusters (Fig. 2a) to their respective promoters. The remaining 17% of DHSs were assigned to the nearest promoter. GO-term and KEGG-pathway analyses of such genes (listed in Supplementary Data 6) again showed activation of genes connected with signaling processes, such as inflammatory response, regulation of MAPK activity and cytokine regulation in all types of AML.

AML subtypes display unique transcription factor networks.

Constitutive and inducible transcription factors form regulatory networks by interacting with their own and/or other regulatory genes32. Cancer cells maintain a stable regulatory network, indicating that the expression of each network member is tightly controlled and remains in balance. Consequently, perturbation of the network components maintaining this balance may destabilize leukemic cells, thus offering new therapeutic options. We therefore combined footprinting, transcription factor gene expression and, where possible, CHI-C data to construct transcription factor networks in normal CD34+ cells and the different AML subtypes by linking occupied binding motifs within transcription factor genes to the specific transcription factor families recognizing these motifs. The transcription factor families involved and the full network structure for each cell type without filtering can be studied in detail on our webserver (see URLs). By comparing the different AML subtypes with normal CD34+ cells, we identified interactions among transcription factor subsets and their genes that were either shared between AMLs and CD34+ cells (Supplementary Fig. 9) or were subtype specific (Fig. 6). These results suggest that the AP-1 family network is of high regulatory relevance for all AML subtypes (Fig. 6b–g). POU4F1 and HLF family factors recognising MYC/MAX-type E-boxes formed prominent nodes in (t(8;21)) AML only, whereas HOX proteins, FOXC1, NFIX and the MAF family formed specific nodes in FLT3-ITD and NPM1 AML. Specific nodes and edges were also part of the normal precursor program.
Fig. 5 | Capture HiC shows differences in locus-specific cis-regulatory interactions between different types of AML and normal cells. a, Percentage of upregulated and downregulated genes with differential interactions from FLT3-ITD and t(8;21) compared to CD34+ PBSCs. Bar figure also shows percentage of common genes for FLT3-ITD and t(8;21); number of differentially expressed genes is on top of each bar. b, Flow diagram showing steps for identification of differential interactions and downstream analysis. c, Top enriched GO terms for upregulated genes of FLT3-ITD compared to CD34+ PBSCs as outlined in a. d, Network diagram of top KEGG pathways for upregulated genes of FLT3-ITD compared to CD34+ PBSCs as outlined in a. e, Top enriched GO terms for upregulated genes of t(8;21) compared to CD34+ PBSCs as outlined in a. f, Network diagram of top KEGG pathways for upregulated genes of t(8;21) compared to CD34+ PBSCs as outlined in b. g, Percentage of RUNX1-ETO and RUNX1 targets among upregulated genes with differential interactions; exact number of genes is above bars.

( Supplementary Fig. 9). For example, the link between the C/EBP family and NFI3 was shared between the FLT3-ITD/NPM1 cells (Fig. 6f) and CD34+ PBSCs (Supplementary Fig. 9f). A detailed discussion of the different network structures and the role of different transcription factor families with more examples can be found in the Supplementary Note.
Fig. 6 | Identification of transcription factor networks driving expression of AML type-specific upregulated transcription factor genes. a, Outline of analysis strategy. b, t(8;21)-specific transcription factor (TF) network. c, CEBPA(x2), double mutant)-specific transcription factor network. d, Inv(16)-specific transcription factor network. e, Mutant RUNX1-specific transcription factor network. f, FLT3-ITD/NPM1-specific transcription factor network. g, NPM1-specific transcription factor network. Factor families binding to the same motif as shown in Supplementary Table 2 form a node contained within a circle. Arrows pointing outward from entire node highlight footprinted motifs in individual genes generated by any member of this factor family whereby the footprint was annotated to the gene using the CHiC data where possible, or otherwise to the nearest gene. For selected nodes, the name of the underlying motif is highlighted in large letters. The expression level (FPKM) of the individual genes is depicted in white (low) or red (high) color. A smooth orange ring around the circle indicates a gene that is specifically upregulated in this type of AML compared to CD34+ PBSCs and/or other AML types; a dotted circle indicates a gene that is upregulated as compared to CD34+ cells.
Transcription factors contributing to AML-specific growth. We next validated the importance of specific transcription factors forming network nodes. To this end we transduced three different AML cell lines and primary FLT3-ITD AML cells with lentiviral vectors encoding short hairpin RNAs (shRNAs) targeting POU4F1 (specific for t(8;21)) or NFIX and FOXC1 (specific for FLT3-ITD). NFIX has a role in myeloid lineage specification but has not been linked to specific mutation types. FOXC1 is an oncogene in its own right and it is overexpressed in AMLs with FLT3-ITD alterations. However, NFIX and FOXC1 have not been linked to the maintenance of the FLT3-ITD AML. We transduced two distinct shRNA constructs per gene into FLT3-ITD and t(8;21) cell lines, which significantly reduced the corresponding transcription factor transcript and protein levels as compared to the control (Supplementary Fig. 10a–f). Knockdown of POU4F1 (Supplementary Fig. 10a,d) inhibited the proliferation of t(8;21)+ Kasumi-1 cells (Fig. 7a and Supplementary Fig. 10g), confirming our previous findings. Expression of NFIX shRNAs (Supplementary Fig. 10b,e) impaired the proliferation of FLT3-ITD+ MV4-11, but not FLT3-ITD+ Kasumi-1 cells (Fig. 7b,c and Supplementary Fig. 10h,i). We next tested the effect of shRNA knockdown of these genes on the colony-forming ability of CD34+ FLT3-ITD/NPM1 mutant subject cells compared to CD34+ PBSCs. Both NFIX and FOXC1 shRNA constructs led to lower colony-forming ability of AML cells compared to a mismatch shRNA control, but no effect was observed in normal CD34+ HSP cells (Fig. 7d,e).

Our network analysis suggests that the AP-1 family is of general importance for all AML subtypes examined, in addition to subtype-specific transcription factors such as POU4F1 or NFIX. AP-1 is a heterodimer formed by members of the FOS, JUN, ATF, CREB and JDP families of transcription factors, thus it is challenging to target by defined RNA interference approaches. To this end, we transduced AML cells with an inducible version of a dominant negative FOS (dnFOS) protein. Doxycycline-mediated induction of dnFOS significantly inhibited proliferation of both t(8;21)+ Kasumi-1 cells and FLT3-ITD-expressing MV4-11 cell lines (Fig. 7g,h and Supplementary Fig. 10j); it also inhibited the colony-forming ability of primary CD34+ FLT3-ITD cells but not that of CD34+ hematopoietic stem and progenitor (HSP) cells (Fig. 7i,j and Supplementary Fig. 10i).

Finally, we examined the importance of AP-1 for leukemia propagation in vivo by transplanting either Kasumi-1 or MV4-11 cells expressing a doxycycline-inducible dnFOS into immunodeficient mice, followed by randomization into doxycycline-treated and untreated groups. In the case of Kasumi-1 transplantation, six of seven animals in the control group, but only two animals in the doxycycline-treated group, developed granulosarcomas (Fig. 7j). Neither of the latter two tumors expressed dnFOS after doxycycline treatment (data not shown), indicating that the transgene was silenced. Doxycycline treatment of mice transplanted with FLT3-ITD MV4-11 cells harboring the dnFOS transgene inhibited the development of leukemia, whereas all untreated mice rapidly developed tumors and were killed (Fig. 7k). Taken together, these findings demonstrate the importance of AP-1 for several AML subtypes and emphasize the potential of transcriptional network analyses to predict transcription factors crucial for malignant propagation.

Discussion

In this study we defined how aberrantly expressed transcription factors and signaling molecules shape the epigenetic landscape of different subtypes of human AML. We show (1) that it is possible to use high-quality DNase I footprinting analyses of purified AML blast cells to identify AML subtype–specific transcription factor networks, (2) that such transcription factor networks highlight a dependence on specific factors for leukemic growth and (3) that the global activation of signaling pathways parallels a growth dependence on AP-1 activity in multiple types of AML. Our comprehensive integration of multi-omics data reveals a strong connection among leukemic classifier mutations, networks of transcription factors and signaling components in primary AML. Moreover, by mapping cis-element promoter interactions by CHIC, we assigned the majority of genes of all analyzed subtypes to their correct promoter. Different types of AML have been classified by their gene expression and DNA-methylation patterns, revealing the existence of specific gene regulatory networks. Our work now defines these networks in detail, and shows that leukemic drivers determine the regulatory phenotype by establishing and maintaining specific gene regulatory and signaling networks that are distinct from those in normal cells. Induced and aberrantly expressed transcription factors are not bystanders, but are important for network maintenance and leukemic growth; this observation highlights new therapeutic opportunities for targeted treatment.

The full set of target genes of RUNX1-ETO in t(8;21) is known and the t(8;21)-specific epigenome and transcription factor binding pattern has been extensively characterized. Multiple target genes relevant for the maintenance of the leukemogenic state have been identified, including FOXO1, UBASH3B, POU4F1 and LAT2, together with the members of the RUNX1–ETO complex, and our current comparative study has identified multiple new network components. However, for the other types of AML, in particular for FLT3-ITD, such knowledge was not available. Here, we identified a number of signaling and transcriptional components that distinguish FLT3-ITD from normal blasts and from other types of AML, providing a rich resource for combinatorial therapy approaches. We examined the contribution to leukemic growth for two genes with AML subtype–specific activity (NFIX and FOXC1) and showed that their elimination led to reduced growth and colony-forming ability of AML but not normal cells, confirming that AML maintains an aberrant transcriptional network required for survival.

The AP-1 factor family has an important role in many types of tumors and our study shows that it is of major importance for AML. FLT3-ITD MV4-11 cells have abundant levels of nuclear AP-1, and FLT3-ITD target genes such as CCNA1 are suppressed by MAPK inhibitors in these cells. We have recently shown that JUN scores highly in a siRNA dropout screen examining the requirements for tumor development in t(8;21) AML. FOS has an important role in resistance to the inhibition of another leukemic oncprotein, the BCR-ABL kinase in chronic myeloid leukemia by activating compensatory signaling pathways. Because several growth factor and stress signal cascades feed into AP-1, targeted inhibition of all AP-1 network components may prevent resistance via rewiring of signaling pathways.

In the classical mechanism of two-step leukemogenesis, a substitution altering a differentiation trajectory cooperates with signaling alterations directing leukemic growth. According to this definition, alterations in transcription factors, which program chromatin directly, and epigenetic regulators, which define the epigenetic landscape upon which transcription factors act, fall into the first category and the FLT3-ITD alteration falls into the second. However, these distinctions are now becoming blurred because from the viewpoint of gene regulation, growth factor receptors strongly influence transcriptional activity via inducible transcription factors, as exemplified by AP-1 family members. These factors have a dominant role in driving the differentiation trajectory as they display an AML subtype–specific occupancy signature that is not influenced by the presence or absence of epigenetic regulator alterations (in this case DNMT3A). This is not to say that alterations in epigenetic regulators do not influence the developmental trajectory of AML and clinical outcomes, as shown in CBF AML, because AML cells with such alterations acquire an altered DNA methylation landscape that probably influences transcription factor binding. However, our data show that leukemic phenotype
Fig. 7 | Identification of AML type–specific transcription factors required for maintaining leukemic growth and colony-forming ability. a–c, Scatter plot showing growth curves of Kasumi-1 cells after transduction with shPOU4F1 (a) or shNFIX (c) and of MV4-11 cells after transduction with shNFIX (b). d,e, Dot plot showing number of colonies formed by FLT3-ITD+ primary AML cell samples (d) or PBSCs (e) after transduction with shRNA targeting FOXC1, NFIX or a mismatch control. CFU, colony-forming units. f, Scatter plot showing growth curve of Kasumi-1 cells transduced with either a doxycycline-inducible dnFOS or an empty vector control with or without 1.5 μg/ml doxycycline (dox). g, Dot plots showing growth curve of MV4-11 cells transduced with either a doxycycline-inducible dnFOS or an empty vector control (right) with or without 1.5 μg/ml doxycycline. h, Expression of a dnFOS causes reduction in the colony-forming ability of CD34+ FLT3-ITD+ primary AML cells (h) but not CD34+ PBSCs (i). j, Granulosaoma formation in RG mice by Kasumi-1 expressing a doxycycline-inducible dnFOS. dnFOS was induced by intraperitoneal injection of doxycycline. k, Survival curve for RG mice transplanted with MV4-11 cells expressing doxycycline-inducible dnFOS. Induction of dnFOS significantly increased survival time of mice. Significance in all experiments was tested using two-tailed Student’s t-test (n = 3) with *P < 0.05, **P < 0.01 in both samples compared to the mismatch control where indicated. Error bars show s.e.m. HR: hazard ratio. Further detail on statistical analysis is in Online Methods.
and self-renewal of different types of AML are defined by differentially activating a multitude of different and often lineage-unrelated signaling pathways and by expressing lineage-unrelated transcription factors. From the viewpoint of finding therapeutic targets, identifying such alteration-specific pathways will offer an opportunity to eliminate their specific maintenance program by targeting multiple pathways simultaneously. Our study takes a first step toward this goal.

**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/s41588-018-0270-1](https://doi.org/10.1038/s41588-018-0270-1).

Received: 2 February 2018; Accepted: 2 October 2018; Published online: 12 November 2018

**References**

1. Cancer Genome Atlas Research, N. Genomic and epigenetic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).

2. Pappasmanull, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).

3. Bonifer, C. & Cockrell, P. N. Chromatin structure profiling identifies crucial regulators of tumor maintenance. *N. Engl. J. Med.* **374**, 157–160 (2015).

4. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* **7**, 105–117 (2007).

5. Cockrell, P. N. Receptor signaling directs global recruitment of pre-existing transcription factors to inducible elements. *Yale J. Biol. Med.* **89**, 591–596 (2016).

6. Ward, A. F., Braun, B. S. & Shannon, K. M. Targeting oncogenic Ras signaling in hematologic malignancies. *Blood* **120**, 3397–3406 (2012).

7. Parikh, C., Subbrahmanyan, R. & Ren, R. Oncogenic NRAS rapidly and efficiently induces CML-like and AML-like diseases in mice. *Blood* **108**, 2349–2357 (2006).

8. Masson, K. & Ronnstrand, L. Oncogenic signaling from the hematopoietic growth factor receptors c-KIT and Flt3. *Cell Signal.* **21**, 1717–1726 (2009).

9. Gerloff, D. et al. NF-kappab/STAT5/mir-155 network targets PU.1 in FLT3-ITD-driven acute myeloid leukemia. *Leukemia* **29**, 535–547 (2015).

10. Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl Acad. Sci. USA* **111**, 2548–2553 (2014).

11. Shluss, L. J. et al. Identification of pre-leukemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).

12. Di Croce, L. & Helin, K. Transcriptional regulation by Polycistronic group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).

13. Broske, A. M. et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat. Genet.* **41**, 1207–1215 (2009).

14. Bonifer, C. & Bowen, D. T. Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine. *Expert Rev. Mol. Med.* **12**, e6 (2010).

15. Challen, G. A. et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2012).

16. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).

17. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12**, 599–612 (2012).

18. Goede, D. K. et al. Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell* **36**, 572–587 (2016).

19. Oehir, N. & Bonifer, C. Chromatin programming by developmentally regulated transcription factors: lessons from the study of haematopoietic stem cell specification and differentiation. *FEBS Lett.* **590**, 4105–4115 (2016).

20. Wilson, N. K. et al. Combinatorial transcriptional control in blood stem/ progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544 (2010).

21. Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).

22. Ptasinska, A. et al. Identification of a dynamic core transcriptional network in t(6;21) AML that regulates differentiation block and self-renewal. *Cell Rep.* **8**, 1974–1988 (2014).

23. Li, C. et al. RUNX1-ETO and RUNX1-EVI1 differentially reprogram the chromatin landscape in t(8;21) and t(5;21) AML. *Cell Rep.* **19**, 1654–1668 (2017).

24. Cauchy, P. et al. Chronic FLT3-ITD signaling in acute myeloid leukemia is connected to a specific chromatin signature. *Cell Rep.* **12**, 821–836 (2015).

25. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).

26. McKeown, M. R. et al. Supern enhancer analysis defines novel epigenomic subtypes of non-API AML, including an RARalpha dependency targetable by SY-1425, a potent and selective RARalpha agonist. *Cancer Discov.* **7**, 1136–1153 (2017).

27. Martens, J. H. et al. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood* **120**, 4038–4048 (2012).

28. Pulkkinen, J. A., Tenen, D. G. & Behre, G. C/EBPalpha deregulation as a paradigm for leukaemogenesis. *Leukemia* **31**, 2279–2285 (2017).

29. Piper, J. et al. Wellington: a novel method for the accurate identification of digital genomics footprints from Dnase-seq data. *Nucleic Acids Res.* **41**, e201 (2013).

30. Ptasinska, A. et al. Depletion of RUNX1/ETO in t(6;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **26**, 1829–1841 (2012).

31. Mandal, B. et al. CBF-BM/MYH11/RUNX1 together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. *Leukemia* **28**, 770–774 (2014).

32. Dunne, J. et al. AML1/ETO protein expression and function in t(8;21) acute myeloid leukemia. *Cancer Res.* **70**, 3985–3995 (2010).

33. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).

34. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).

35. Chasman, D. & Roy, S. Inference of cell type specific regulatory networks on mammalian lineages. *Cure Opin. Syst. Biol.* **2**, 130–139 (2017).

36. Nep, S. et al. Circuity and dynamics of human transcription factor regulatory networks. *Cell* **150**, 1274–1286 (2012).

37. Lin, S. et al. A FOXO1-induced oncogenic network defines the AML1-ETO preleukemic program. *Blood* **130**, 1213–1222 (2017).

38. O’Connor, C. et al. NFI expression critically modulates early B lymphopoiesis and myelopoiesis. *PLoS One* **10**, e0121005 (2015).

39. Somerville, T. D. et al. Frequent derepression of the mesenchymal transcription factor gene FoxC1 in acute myeloid leukemia. *Cancer Cell* **28**, 329–342 (2015).

40. Olive, M. et al. A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J. Biol. Chem.* **272**, 18586–18594 (1997).

41. Oehir, N. et al. Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development* **143**, 4324–4340 (2016).

42. Verhaak, G. R. et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94**, 131–134 (2009).

43. Figueroa, M. E. et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).

44. Lin, S., Mulloy, J. C. & Goyama, S. RUNX1-ETO leukemia. *Adv. Exp. Med. Biol.* **962**, 151–173 (2017).
Acknowledgements

This research was funded by a program grant from Bloodwise (15001) to C.B. and P.N.C., as well as studentship awards from Cancer Research UK and Bloodwise to A.Pickin and N.G., respectively, a Kay Kendall Clinical Training Fellowship for J.L. and a MRC/Leuka Clinical Training Fellowship for S.P.

Author contributions

M.R.I., D.J.L.C., S.P., A.Ptasinska, H.B., A.Pickin, L.N.G., J.C.L., P.S.C., J.Z.-C. and S.R.J. performed experiments and generated data; R.D., M.R., S.J.R., M.J.G., P.J. and A.U. provided subject samples; S.C., A.B. and P.N.C. conducted mutation analysis; S.A.A. and FC. analyzed data; O.H. supervised transplantation experiments and helped edit the manuscript; C.B. and P.N.C. conceived and directed the study and C.B. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0270-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.N.C. or C.B.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018
Methods

Subject samples and PBSC processing. Human tissue was obtained with the required ethical approval from the National Health Service (NHS) National Research Ethics Committee. AML and PBSC samples used in this study were either subject diagnostic samples or fresh samples obtained with specific consent from the subjects. AML samples were obtained from (1) the Haematological Malignancy Diagnostic Service (St James’s Hospital, Leeds, UK), (2) the Centre for Clinical Haematology, Queen Elizabeth Hospital Birmingham, Birmingham, UK, (3) the West Midlands Regional Genetics Laboratory, Birmingham Women’s NHS Foundation Trust, Birmingham, UK, or (4) Eramus University Medical Center, Rotterdam, The Netherlands. Mononuclear cells were purified on the same day they were received, and in most cases were also directly further purified using either CD34 or CD117 (KIT) magnetic antibodies as described. For some samples with >92% blast cells, column purification was not performed. Mobilized PBSCs were provided by NHS Blood & Transplant, Leeds, UK, and NHS Blood & Transplant, Birmingham, UK.

Mutation detection. Mutated genes identified in each subject are summarized in Supplementary Table 1, together with the age, gender and white blood cell count for each subject. Mutations were identified by one of two different methods. The first batch of subjects were assayed by targeted exon sequencing of 55 cancer-associated genes using 1,212 pairs of previously defined PCR primers for amplification using a RainDance Technologies platform. The mutation sequence data from this screen was analyzed using algorithms to detect either nucleotide variants using the Genome Analysis Toolkit (GATK) or insertions and deletions using Patient Indel Detection and Indel Calling (PIDIC). Sequenced samples were screened for any previously identified or novel mutations (see URLs). Subsequent samples were assayed using the Illumina TruSight myeloid panel of primers and processes by approaches similar to those used for the first batch. All identified mutations are listed in Supplementary Table 1. Some of these subjects were also included in a previous publication from our laboratory, using different subject identification codes than those used in the current study.

Cell lines. Cell lines were cultured in an incubator at 37 °C in GIBCO 1640 RPMI plus Glutamax medium supplemented with 10% heat-inactivated FCS (GIBCO), 100 μM penicillin and 100 mg/ml streptomycin.

Growth curve measurements. 250,000 MV4-11 or Kasumi-1 cells were cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% FCS, 2 mM l-glutamine, 100 μM penicillin and 100 mg/ml streptomycin. Cells were counted with Trypan blue exclusion and split every 3 d to maintain them in the log phase of growth. For the inducible dnFOS, cells were counted and split every 2 d and 1.5 μg/ml of doxycycline was added.

Coculture of primary cells with MS-5 feeders. Primary cells were maintained in coculture with MS-5 cells briefly. Cells were cultured in LCT medium (tissue medium essential medium (Lonza) supplemented with heat-inactivated 12.5% FCS (Gibco)), heat-inactivated 12.5% horse serum (Gibco), penicillin and streptomycin, 200 mM glutamine, 57.2 μM β-mercaptoethanol (Sigma) and 1 μM hydrocortisone (Sigma) supplemented with 20 ng/ml interleukin-3 (IL-3), granulocyte colony-stimulating factor and thrombopoietin in flasks precoated with MS-5 cells.

Lentiviral transduction and shRNA treatment. LEGO-iG shRNAs were generated by cloning shRNAs (Supplementary Table 3) into the LEGO-iG vector. LEGO-iG-dnFOS was generated by cloning the dnFOS insert, originally generated by Charles Vinson (National Cancer Institute, Bethesda, MD, USA), into the LEGO-iG backbone. Inducible dnFOS was cloned into a pENTR backbone and then Gateway cloning was used to insert that into the Tet-on plasmid pCW57.1 (David Root, Addgene plasmid 41393). Gateway backbone vectors LEGO-iG and inducible dnFOS were then used to generate lentiviral particles using packaging and envelope genes on four separate plasmids: TAT, REV, GAG/POL and VSV-G. For virus production, Human embryonic kidney 293T (HEK293T) cells were cultured in DMEM supplemented with 10% FCS, 2 mM l-glutamine, 100 μM penicillin, 100 mg/ml streptomycin and 0.1 mg/ml sodium pyruvate, and were seeded to a density of 5×10⁷ cells/ml in 24-well plates. Twenty-four hours after transfection, the supernatant was removed and replaced with 0.5 μg/ml polybreine by spinoculation at 1,500 g:1.2 μg:1.2 g ml–1 polybrene, and then added to the plate and transduced by spinoculation at 1,500 g for 50 min. After 12–16 h incubation at 37 °C, viral medium was exchanged for fresh medium. FACS was performed to isolate GFP+ cells 3 d after transduction.

Primary cell samples were defrosted 24 h before transduction and cocultured with MS-5 feeder cells in LCT medium. Six-well nontissue culture treated plates were coated with 24 μg/ml retinovecin (Takara Clontech) for 2 h before blocking with 2% BSA in PBS for 30 min. The blocking buffer was washed off with Hank’s buffered salt solution (Gibco) containing 2.5% HEPES. Viral concentration (1 ml) was applied to the retinovecin-coated plate by centrifugation at 2,000g for 45 min, after which the concentrated viral supernatant was refreshed and the centrifugation repeated. Primary cells were suspended to a concentration of 1×10⁶ cells/ml in the remaining viral supernatant, supplemented with 20 ng/ml granulocyte colony-stimulating factor, IL-3, thrombopoietin and 8 μg/ml polybreine, and then added to the plate and transduced by spinoculation at 1,500 g for 50 min. After 12–16 h incubation at 37 °C, viral medium was exchanged for fresh medium. FACS was performed to isolate GFP+ cells 3 d after transduction.

Colony-formation assays of primary cells. Colony-formation assays were performed on sorted cells by seeding at 2,500 cells/ml in Methocult Express Cell Stem Technologies. After 14 d colonies were counted.

Animal experiments. Immunodeficient Rag-2−/−Il2γ−/−129×Balb/c (RG) mice were housed in the Comparative Biology Centre (Newcastle University) under specific pathogen-free conditions. All animal work was conducted in accordance with Home Office Project License PPL60/4552 by researchers who had completed approved Home Office training and held current personal licenses under the Animals (Scientific Procedures) Act 1986. Kasumi-1 pCW57.1-dnFOS cells were intraperitoneally injected into 14 newborn (2-d old) RG mice at a cell dose of 2.5×10⁶ cells per mouse as described. At 12 d later, mice were randomized into two treatment groups, with one given 50 mg/kg doxycycline three times per week intraperitoneally in an unblended fashion until the experimental endpoint. MV4-11 pCW57.1-dnFOS cells were intrafemorally injected into RG mice at a cell dose of 5×10⁵ cells per mouse followed by randomization into two groups. For the doxycycline group, doxycycline was added at 2 mg/ml for the initial 3 d and at 0.2 mg/ml subsequently to drinking water containing 2% sucrose. Controls were given water containing 2% sucrose. Animals were humanely killed upon clinical signs of illness or at defined experimental endpoints.

Quantitative PCR with reverse transcription. RNA was extracted using the Machery-Nagel Nucleospin kit. RNA (1 μg) was used to make cDNA with 0.5 μg oligo(dT) primer. Murine Moloney reverse transcriptase and RNAse inhibitor (Promega) according to the manufacturer’s protocol. PCR with reverse transcription was performed using Syber Green mix (Applied Biosystems), at 2× dilution. Primers were used at 100 nM. A 7900HT system (Applied Biosystems) was used to perform quantitative PCR. Analyses were performed in technical duplicates using a standard curve derived from the untreated cell line.

Western blotting. Protein lysates from cell lines were analyzed by western blot. Relevant primary antibodies to FOXC1 (Cell Signaling Technology, 8758), NFIX (Invitrogen, PA5-31234) and POU4F1 (Santa Cruz Biotechnology, sc-8429) were used to detect target genes, and antibody to GAPDH (mouse anti-GAPDH, ab8245; rabbit anti-GAPDH, Cell Signaling Technology, 2118L) was used as a housekeeping gene. Secondary antibodies mouse anti-rabbit horseradish peroxidase (Santa Cruz Biotechnology, sc-2054) and goat anti-mouse horseradish peroxidase (Jackson ImmunoResearch, 115-035-062) enabled detection and quantification by densitometry using ImageJ lab software and a GelDoc imager.

Statistical analysis of validation experiments. Validation experiments were performed in triplicate (n=3), where each replicate represents a separately transduced cell line or primary cell sample cultured and analyzed independently. The exception are the in vitro inducible dnFOS cell line experiments, where each replicate represents a separate culture of a clone stably expressing the dnFOS peptide to avoid variations in the degree of expression. The significance of differences in colony counts, doubling time, mRNA expression and cell concentrations at individual time points on growth curves were calculated using Student’s t-test where n=3 using four degrees of freedom; t values between significantly different groups (P<0.05) were all >3.2. Standard error bars on graphs show s.e.m.

To determine the tumor mass of mice transplanted with Kasumi-1 pCW57.1-dnFOS without or with doxycycline was analyzed using a Mann-Whitney test as the distribution of data was not normal. This experiment was performed for n=7 engrafted mice in each group (P=0.0326, Mann-Whitney U=8.5). The survival curve of mice engrafted with MV4-11 pCW57.1-dnFOS with or without doxycycline was analyzed with a Kaplan-Meier plot (n=5 engrafted mice) using a Mantel-Cox chi-squared test (P=0.897, degrees of freedom=1, P=0.00044) and yielding a Mantel-Haenszel hazard ratio of 14.08.

RNA-seq library preparation. RNA was extracted and analyzed from purified AML cells as described. DNase I-seq. DNase I digestions of permeabilized cells were performed as described. Briefly, live cells were added directly to a solution of DNase I (DPFF, Worthington) in dilute Nonidet P40 and digested for 3 min at 22 °C, and the reactions were then terminated by addition of SDS to 0.5%. DNase I was typically used in the range of 2–6 μg/ml with a final 1.5×10⁶ cells/ml. DNase I–seq
libraries were then prepared and validated essentially as described 38. Libraries were run on Illumina sequencers.

Promoter capture HiC from subject AML blasts. AML cells from subject peripheral blood were first purified by density gradient centrifugation (Lymphoprep) and then used CD34 antibody–coupled beads. 5 × 10^6 (8.21) blasts (subject (8.21)-1R) and FLT3-ITD/NPM1 blasts (subject ITD/NPM1-2) were fixed in 37 ml of RPMI-1640 supplemented with 15% FBS and 2% formaldehyde for 10 min at room temperature. Glycine (6 ml of 1 M, 0.125 M formaldehyde concentration) was added to quench the reaction and cells were incubated at room temperature for 5 min, followed by 15 min on ice before the cells were pelleted at 4°C and washed in ice-cold PBS. Each sample was flash frozen in liquid nitrogen and stored at −80°C. Cells were lysed in a tight dounce homogenizer (ten cycles) with 3 ml of cold lysis buffer (10 mM Tris-HCL, pH 8, 10 mM NaCl, 0.2% Igepal CA-630, one tablet protease inhibitor cocktail (Roche complete, EDTA-free, 11873580041)). Cells were left on ice for 5 min and were then homogenized another ten times. The lysed cells, in 3 ml lysis buffer, were added to 47 ml of lysis buffer and incubated on ice for 30 min with occasional mixing. Chromatin was pelleted and resuspended in 1 ml of 1.25x NEBuffer 2 and then split into four samples. Each sample was then pelleted at 100g and resuspended in 358 μl of 1.25x NEBuffer 2. SDS (11 μl of 10%) was added and each tube was incubated at 37°C for 60 min on a rotary mixer at 950 rpm. Samples were mixed by pipetting up and down every 15 min. SDS was quenched with 75 μl 10% Triton X-100 and incubated at 37°C for 60 min. HindIII digestion, biotinylation, ligation, cross-link reversal, promoter capture and library preparation were performed exactly as described 29.

DNase I–seq data analysis. DNase I–seq from subject AML blasts. The CHi-C paired-end sequences reads from ITD/NPM1-2 and (8.21)-1R subjects and a publicly available CD34+ data set (accession numbers ERR436032 and ERR436025) were put through HiCUP pipeline 32. The raw sequencing reads were separated and mapped against the human genome (hg38). The reads were then filtered for experimental artifacts and duplicate reads, and then re-paired. Statistically significant interactions were called using GOTHIC package 32 and HOMER software. This uses a cis-regulatory motif to detect interactions between distal genomic loci that have significantly more reads than expected by chance, by using a background model of random interactions. This analysis assigns each interaction a P value, which represents its significance. Differential interactions were determined with HOMER 32 for (8.21) using FLT3-ITD or CD34+ as background and FLT3-ITD using (8.21) or CD34+ as background. A difference with P < 0.1 was deemed significant.

Transcription factor gene regulatory network construction. We identified a subset of 310 transcription factor genes that are expressed in one or more of our AML samples. The gene names for transcription factors in humans were obtained from AnimalTFDB 35. The 310 transcription factors were considered as nodes and the and the nodes were colored according to their expression values at each AML subtype (Fig. 6, Supplementary Fig. 9 and Supplementary Note Fig. SN5). Node border color signifies whether the gene was upregulated or downregulated in comparison to two-fold change compared to CD34+ cells. The node border type indicates whether a gene was differentially expressed in one AML subtype as compared to other subtypes. A directed edge from transcription factor to TF, indicates motif binding of TF to the locus encoding transcription factor TF, and the edge is prominently displayed if TF, bound to the locus at that stage. The edge was classified and color coded according to the significance of motif count enrichment.

Motif count enrichment for transcription factor network construction. Initially footprints for each AML subtype were identified using the Wellington algorithm 32 and were annotated to their related promoter using Chi-C data where possible. Motif searches within footprint coordinates were performed using HOMER 32. The number of motifs per transcription factor gene were counted and the significance of motif enrichment was used in bootstrapping random sampling; a random set of mapped motifs was extracted from all union footprint motif of all AML subtypes and the CD34+ cells. After 1,000 iterations, the mean, standard deviation and z-scores were computed. Motif (TF_F) was linked to gene (TF_P) with positive z-score values only.

Motif count enrichment for upregulated transcription factors. The correlations (r) between all transcription factor genes based on FPKM values from the RNA-seq analysis were identified and the correlations (r) between all transcription factor genes based on motif count binding were identified. The correlation coefficients were z-transformed using Fisher z-transformation with FDR correction in R. The percentage of the transformed values (values > 0.05) were identified using bootstrapping of random sampling; a random set of mapped motifs was extracted from all union footprint motif of all AML subtypes and the CD34+ cells. After 1,000 iterations, the mean, standard deviation and z-scores were computed. Motif (TF_F) was linked to gene (TF_P) with positive z-score values only.

Motif count enrichment for upregulated transcription factors. The correlations (r) between all transcription factor genes based on FPKM values from the RNA-seq analysis were identified and the correlations (r) between all transcription factor genes based on motif count binding were identified. The correlation coefficients were z-transformed using Fisher z-transformation with FDR correction in R. The percentage of the transformed values (values > 0.05) were identified using bootstrapping of random sampling; a random set of mapped motifs was extracted from all union footprint motif of all AML subtypes and the CD34+ cells. After 1,000 iterations, the mean, standard deviation and z-scores were computed. Motif (TF_F) was linked to gene (TF_P) with positive z-score values only.

List of used position weight matrices. Motif curvature is described the legend of Supplementary Table 2.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.
Data availability
Raw data have been deposited at GEO under accession number GSE108316. Processed data are available from our data server (http://bioinformatics-bham.co.uk/tfinaml/).

References
53. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).
54. van Gosliga, D. et al. Establishing long-term cultures with self-renewing acute myeloid leukemia stem/progenitor cells. Exp. Hematol. 35, 1538–1549 (2007).
55. Weber, K., Bartsch, U., Stocking, C. & Fehse, B. A multicolor panel of novel lentiviral “gene ontology” (LeGO) vectors for functional gene analysis. Mol. Ther. 16, 698–706 (2008).
56. Mostoslavsky, G. et al. Efficiency of transduction of highly purified murine hematopoietic stem cells by lentiviral and oncoretroviral vectors under conditions of minimal in vitro manipulation. Mol. Ther. 11, 932–940 (2005).
57. Bert, A. G., Johnson, B. V., Baxter, E. W. & Cockerill, P. N. A modular enhancer is differentially regulated by GATA and NFAT elements that direct different tissue-specific patterns of nucleosome positioning and inducible chromatin remodeling. Mol. Cell. Biol. 27, 2870–2885 (2007).
58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
59. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
61. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
62. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53 (2013).
63. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 4, 1310 (2015).
64. Mifsud, B. et al. GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. PLoS One 12, e0174744 (2017).
65. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
66. Zhang, H. M. et al. AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res. 40, D144–D149 (2012).
**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).

| Item                                                                 | 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**

All software is described in the manuscript
The list of software used to analyze DNA-seq data
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: Bowtie version 2.3.1. (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml)
3. For peak calling MACS version 2 (https://github.com/taoliu/MACS)
4. For normalizing to the total read depth among more than one samples DEseq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
5. Digital genomic footprinting was performed using the Wellington_footprints function of the Wellington algorithm (http://pythonhosted.org/pyDNase/)

List of software used to analyse RNA-seq data
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: STAR (https://github.com/alexdobin/STAR)
3. For FPKM values cufflinks (http://cole-trapnell-lab.github.io/cufflinks/)
4. for differential expression analysis (http://bioconductor.org/packages/release/bioc/html/limma.html)

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

Raw data have been deposited at GEO under the accession number GSE108316.

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

Sample size
We used a total of 42 AML patient samples to ensure that we had an average of 3 to 4 patients for each of the major mutation groups. Because we are investigating a total 128,000 DNase-Seq peaks, and over 20,000 genes in each RNA-Seq, this is sufficient for very robust statistical analyses of enriched gene sets in each population.

Data exclusions
Patients were excluded if the quality of the RNA or DNA samples failed our quality control. By definition, these patients are not included in any of the data submitted.

Replication
All manual experiments such as viral transductions, growth and colony assays were done in triplicate. The mutation-specific gene sets identified in our data were validated by comparing them with equivalent data from (1) an analysis of 461 AML patients by Veerhaak et al, and (2) chromatin data from two studies by the Majeti laboratory (Corces et al., 2016, Nature Gen., McKeown et al. 2017, Cancer Discovery). Our digital footprinting experiments were validated using publicly available ChIP-Seq data as described in the manuscript. Promoter capture HiC data were compared to publicly available data from Mifsud et al., 2015, Nature Gen.

Randomization
Not applicable

Blinding
Not relevant as this study is discovery science, not a test.

Reporting for specific materials, systems and methods
Materials & experimental systems

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

Methods

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

Antibodies

| Antibodies used | αFOXC1: Cell signaling #8758, Clone: Rabbit (D8A6) mAb, Lot# 1 |
|------------------|----------------------------------------------------------|
|                   | αNFIX: Invitrogen #PA5-31234, Clone: Rabbit polyclonal, Lot# RK2302461A |
|                   | αPOU4F1: Santa Cruz Biotechnology sc-8429, Clone: Mouse (14A6) mAb, Lot# E2318 |
|                   | Mouse αGAPDH: Abcam ab8245, Clone: mouse (6CS) mAb, Lot# GR3185172-5 |
|                   | Rabbit αGAPDH: Cell signaling 2118L, Clone: Rabbit (14C10) mAb, Lot# 10 |
|                   | Goat αMouse HRP: Jackson ImmunoResearch - 115-035-062, Clone: N/A, Lot# 118840 |
|                   | Mouse αRabbit: Santa Cruz Biotechnology sc-2054, Clone: N/A, Lot# C1315 |

Validation

| αFOXC1: Described as reactive with human, mouse rat on manufacturer's website with example western blot of FOXC1 expressing cell lines. 3 citations taken from manufacturer’s website: Fabiani, E., Falconi, G., et al. (2017), Oncotarget, 8 (48), pp. 84074-84085 Chung, S., Jin, Y., et al. (2017), Cell Commun Signal, 15 (1), pp. 22 Ou-Yang, L., Xiao, S. J., et al. (2015), Mol Med Rep, 12 (6), pp. 8003-9 |
| Western Blot in manuscript was performed including Kasumi-1 protein as a negative control. αNFIX: Described as targeting NFIX in WB applications and shows reactivity with Human samples on manufacturer's website with example western blot of NFIX in 293T cell line. Previously tested in lab using RAJI lymphoma cell line as negative control. Protein detection on western blot mirrors mRNA level when shRNA/siRNA knockdown performed in previous experiments. αPOU4F1: Recommended for detection of Brn-3a (POU4F1) of mouse, rat and human origin by WB, IP, IF, IHC(P) and ELISA by manufacturer. 34 listed citations on manufacturer’s website. Western Blot in figure conducted using MV4-11 protein as a negative control. Mouse αGAPDH: This antibody can be used as a loading control antibody for Mouse, Rat, Rabbit, Chicken, Hamster, Cat, Dog, Human, Pig, Xenopus laevis, Fish, Monkey, Zebrasfish, Baboon, African green monkey according to manufacturer. Cited 1056 times on manufacturer’s website. Rabbit αGAPDH: Detects endogenous levels of total GAPDH protein in Human, Mouse, Rat, Monkey, Bovine, Pig according to manufacturer. 1764 citations on manufacturer’s website. |

Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | MV4-11, Fujioka and Kasumi-1 were obtained from ATCC |
| Authentication | The FLT3-ITD in MV4-11 was validated by PCR of the duplicated region. The t(8;21) fusion in Kasumi-1 was validated by sequencing across it. Fujioka cell line expressed FOXC1 mRNA (qPCR) and protein (Western Blot). |
| Mycoplasma contamination | All cell lines were regularly tested |
| Commonly misidentified lines (See ICLAC register) | Not applicable |

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | Rag2-/-Il2rg-/-129xBalb/c (RG) mice, newborn, both sexes |
| Wild animals | The study did not involve wild animals |
| Field-collected samples | The study did not involve samples collected from the field |
Human research participants

Policy information about human research participants

Population characteristics

| Supplementary table 1 lists the age and gender of each patient donor of AML cells |

Recruitment

AML and PBSC samples used in this study were either surplus diagnostic samples, or were fresh samples obtained with specific consent from the patients. AML samples were obtained from either (i) the Haematological Malignancy Diagnostic Service (St James’s Hospital, Leeds, UK), (ii) the Centre for Clinical Haematology, Queen Elizabeth Hospital Birmingham, Birmingham, UK, (iii) the West Midlands Regional Genetics Laboratory, Birmingham Women’s NHS Foundation Trust, Birmingham, UK, or from (iv) Erasmus University Medical Center, Rotterdam, The Netherlands.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

Files in database submission

Note that the manuscript does not contain any new ChIP-Seq experiments but DNaseI-Seq and RNA-Seq data. ChIP-Seq data used in the analysis were from published sources as outlined in the Methods section.

DNAse-I-seq list

CEBPAx2-1_DHS.wig.gz
CEBPAx2-2_DHS.wig.gz
CEBPAx2-3_DHS.wig.gz
inv16-1_DHS.wig.gz
inv16-2_DHS.wig.gz
inv16-3_DHS.wig.gz
inv3CBL-2_DHS.wig.gz
Inv3RAS-1_DHS.wig.gz
ITD-1_DHS.wig.gz
ITD-2_DHS.wig.gz
ITDxNPM1-1_DHS.wig.gz
ITDxNPM1-2_DHS.wig.gz
ITDNPM1-1_DHS.wig.gz
ITDNPM1-2_DHS.wig.gz
ITDNPM1-3_DHS.wig.gz
ITDNPM1-4_DHS.wig.gz
ITDNPM1-5_DHS.wig.gz
NPM1-1_DHS.wig.gz
NPM1-2_DHS.wig.gz
NPM1RAS-3-DHS.wig.gz
RUNX1-D_S_DHS.wig.gz
RUNX1-D_JAK-1_DHS.wig.gz
RUNX1-DT-1_DHS.wig.gz
RUNX1-DT-3_DHS.wig.gz
RUNX1-DT_CEBPA-2_DHS.wig.gz
RUNX1-T-7_DHS.wig.gz
RUNX1-T_CEBPA-6_DHS.wig.gz
RUNX1-T_JAK-2_DHS.wig.gz
RUNX1x2-DT-4_DHS.wig.gz
t821-1_DHS.wig.gz
t821-1_rel_DHS.wig.gz
t821-3_DHS.wig.gz
t821K1T-2_DHS.wig.gz

RNA-seq list

CD34Plus_CB.bedgraph.gz
CEBPA-5.bedgraph.gz
CEBPAx2-1.bedgraph.gz
CEBPAx2-2.bedgraph.gz
CEBPAx2-3.bedgraph.gz
inv16-1.bedgraph.gz
inv16-2.bedgraph.gz
inv16-3.bedgraph.gz
inv3RAS-1.bedgraph.gz
inv3RAS-3.bedgraph.gz
Genome browser session
(e.g. UCSC)

UCSC link to DNASe-seq
https://genome-euro.ucsc.edu/cgi-bin/hgTracks?
  hgS_doOtherUser=submit&hgS_otherUserName=salam&hgS_otherUserSessionName=AML_DHS_Assi.et.al

UCSC link to RNA-seq data
https://genome-euro.ucsc.edu/cgi-bin/hgTracks?
  hgS_doOtherUser=submit&hgS_otherUserName=salam&hgS_otherUserSessionName=AML_RNAseq_Assi.et.al

Methodology

Replicates

The DNAsel-seq samples were generated from primary cells from individual patients and for each major group of mutations we examined the following number of samples:

Number of samples for each group
RUNX1-T/JAK-2
  inv(3) 2
RUNX1 6
ITD(2x)/NPM1 2
ITD/NPM1 6
ITD 3
CEBPA(2x) 3
t(6;21) 4
inv(16) 3
RUNX1-T-7 (NHL) 1
NPM1/RAS-3 1
PBSC 2

Single patients were included as controls

Sequencing depth

All DNAsel-seq samples are single-end reads and the length of reads is 50 bp

Sample code total number of aligned reads
RUNX1-T/JAK-2 100,965,405
RUNX1-D/JAK-1 169,740,092
inv(3)/CBL-2 60,810,627
inv(3)/RAS-1 "176,938,868"
RUNX1-D-S 193,586,370
RUNX1-T/CEBPA-6 138,443,303
RUNX1-DT-1 387,658,545
RUNX1-DT/CEBPA-2 28,292,155
RUNX1-DT-3 22,760,769
RUNX1(x2)-DT-4 257,468,842
ITD(2x)/NPM1-2 216,542,686
Antibodies (NA) as these are DNASe1-seq experiments

Peak calling parameters

Here is the command used for read mapping and peak calling:

```
bowtie2 -p 4 --very-sensitive-local -x hg38 -U DNASE1.fastq -S DNASeI.sam
samtools view -bt hg38.fa.fai -o DNASeI.bam DNASeI.sam
macs2 callpeak –t DNASe1.bam -f BAM --nomodel --call-summits -q 0.005 -n DNASEI_macs2_q_0.005
```

Data quality

DNase-seq sequences from all experiments were mapped onto the reference human genome version hg38, with Bowtie version 2.3.1. Low quality reads were trimmed using trimmomatic prior to the alignment and the quality control (QC) statistics for the samples were obtained using FastQC tools.

DNase I Hypersensitive Sites (DHSs) were called with MACS2 using callpeak function (nomodel, call-summits and q= 0.005 parameters).

To define a common set of coordinates covering all of the significant DHSs investigated in this study, we merged all of the individual DNase-Seq reads for all of the AML samples assayed by DNaseI-Seq. This data set was then used to define the peak summits of the union of all peaks. This approach was designed to maximize the precision and sensitivity of the peak detection. The DNA read counts were then determined within 400 bp windows centered on each peak for each sample. To account for the different number of reads in each of the samples; the read counts were normalized for total read depth using DEseq2.

Sample code number of peaks

RUNX1-T/JAK-2 37078
RUNX1-D/JAK-1 39913
inv(3)/CBL-2 50167
inv(3)/RAS-1 51146
RUNX1-D/5 45624
RUNX1-T/CEBPA-6 54020
RUNX1-DT-1 43328
RUNX1-DT/CEBPA-2 50423
RUNX1-DT-3 50770
RUNX1(x2)-DT-4 39008
ITD(2x)/NPM1-2 41083
ITD(2x)/NPM1-1 57180
NPM1-2 29151
NPM1-1 53055
ITD/NPM1-2 42176
ITD/NPM1-3 53911
ITD/NMP1-1 43248
ITD/NPM1-6 53427
ITD/NPM1-5 47032
ITD/NPM1-4 47957
ITD-2 50485
ITD-3 54989
ITD-1 63293
CEBPA(x2)-3 35666
CEBPA(x2)-2 54508
AML and PBSC samples used in this study were either surplus diagnostic samples, or were fresh samples obtained with specific consent from the patients. AML samples were obtained from either (i) the Haematological Malignancy Diagnostic Service (St James’s Hospital, Leeds, UK), (ii) the Centre for Clinical Haematology, Queen Elizabeth Hospital Birmingham, Birmingham, UK, (iii) the West Midlands Regional Genetics Laboratory, Birmingham Women’s NHS Foundation Trust, Birmingham, UK, or from (iv) Erasmus University Medical Center, Rotterdam, The Netherlands.

The list of software used to analyze DNAseI-seq data:
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: Bowtie version 2.3.1. (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml)
3. For peak calling MACS version 2 (https://github.com/taoliu/MACS)
4. For normalizing to the total read depth among more than one samples DEseq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
5. Digital genomic footprinting was performed using the Wellington_footprints function of the Wellington algorithm (http://pythonhosted.org/pyDNase/)

List of software used to analyse RNA-seq data:
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: STAR (https://github.com/alexdobin/STAR)
3. For FPKM values cufflinks (http://cole-trapnell-lab.github.io/cufflinks/)
4. For differential expression analysis (http://bioconductor.org/packages/release/bioc/html/limma.html)
ChIP-seq Reporting Summary

Data deposition

1. For all ChIP-seq data:
   a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.
   b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all relevant data deposition access links. The entry may remain private before publication.
   https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108316
   secure token: yzmbwaocjhwztsj

3. Provide a list of all files available in the database submission.

Note that the manuscript does not contain any new ChIP-Seq experiments but DNaseI-Seq and RNA-Seq data. ChIP-Seq data used in the analysis were from published sources as outlined in the Methods section.

DNAseq list

CEBPAx2-1_DHS.wig.gz
CEBPAx2-2_DHS.wig.gz
CEBPAx2-3_DHS.wig.gz
inv16-1_DHS.wig.gz
inv16-2_DHS.wig.gz
inv16-3_DHS.wig.gz
inv3CBL-2_DHS.wig.gz
Inv3RAS-1_DHS.wig.gz
ITD-1_DHS.wig.gz
ITD-2_DHS.wig.gz
ITDxNPM1-1_DHS.wig.gz
ITDxNPM1-2_DHS.wig.gz
ITDNPM1-1_DHS.wig.gz
ITDNPM1-2_DHS.wig.gz
ITDNPM1-3_DHS.wig.gz
ITDNPM1-4_DHS.wig.gz
ITDNPM1-5_DHS.wig.gz
NPM1-1_DHS.wig.gz
NPM1-2_DHS.wig.gz
NPM1RAS-3-DHS.wig.gz
RUNX1-D-5_DHS.wig.gz
RUNX1-D_JAK-1_DHS.wig.gz
RUNX1-DT-1_DHS.wig.gz
RUNX1-DT-3_DHS.wig.gz
RUNX1-DT_CEBPA-2_DHS.wig.gz
RUNX1-T-7_DHS.wig.gz
RUNX1-T_CEBPA-6_DHS.wig.gz
RUNX1-T_JAK-2_DHS.wig.gz
RUNX1x2-DT-4_DHS.wig.gz
t821-1_DHS.wig.gz
t821-1_rel_DHS.wig.gz
1821-3_DHS.wig.gz
t821KIT-2_DHS.wig.gz

RNAseq list

CD34Plus_CB.bedgraph.gz
CEBPA-5.bedgraph.gz
CEBPAx2-1.bedgraph.gz
CEBPAx2-2.bedgraph.gz
CEBPAx2-3.bedgraph.gz
inv16-1.bedgraph.gz
inv16-2.bedgraph.gz
inv16-3.bedgraph.gz
inv3RAS-1.bedgraph.gz
inv3RAS-3.bedgraph.gz
ITD-1.bedgraph.gz
ITD-2.bedgraph.gz
ITD2x_NPM1-1.bedgraph.gz
ITD2x_NPM1-2.bedgraph.gz
ITD-3.bedgraph.gz
ITDNPM1-1.bedgraph.gz
ITDNPM1-2.bedgraph.gz
ITDNPM1-3.bedgraph.gz
ITDNPM1-4.bedgraph.gz
ITDNPM1-5.bedgraph.gz
ITDNPM1-6.bedgraph.gz
NPM1-1.bedgraph.gz
NPM1-2.bedgraph.gz
NPM1RAS-3.bedgraph.gz
PBSC-1.bedgraph.gz
RUNX1-D-5.bedgraph.gz
RUNX1-D_JAK-1.bedgraph.gz
RUNX1-DT-1.bedgraph.gz
RUNX1-DT-3.bedgraph.gz
RUNX1-DT_CEBPA-2.bedgraph.gz
RUNX1-T-7.bedgraph.gz
RUNX1-T_CEBPA-6.bedgraph.gz
RUNX1-T_JAK-2.bedgraph.gz
RUNX1x2-DT-4.bedgraph.gz
SRSF2-1.bedgraph.gz
SRSF2-2.bedgraph.gz
t821-1.bedgraph.gz
t821-1_rel.bedgraph.gz
t821.ITDx2-5.bedgraph.gz
t821.KIT-2.bedgraph.gz

4. Provide a link to an anonymized genome browser session (e.g. UCSC), if available.

UCSC link to DNASe-seq
https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=salam&hgS_otherUserSessionName=AML_DHS_Assiet.al

UCSC link to RNA-seq data
https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=salam&hgS_otherUserSessionName=AML_RNAseq_Assiet.al

▶ Methodological details

5. Describe the experimental replicates.

The DNAseI-seq samples were generated from primary cells from individual patients and for each major group of mutations we examined the following number of samples:

Number of samples for each group
RUNX1-T/JAK 2
inv(3) 2
RUNX1 6
ITD(2x)/NPM1 2
ITD/NPM1 6
ITD 3
CEBPA(x2) 3
t(8;21) 4
inv(16) 3
RUNX1-T-7 (NHL) 1
NPM1/RAS-3 1
6. Describe the sequencing depth for each experiment.

All DNAseI-seq samples are single-end reads and the length of reads is 50 bp

Sample code total number of aligned reads
RUNX1-T/JAK-2 100,965,405
RUNX1-D/JAK-1 169,740,092
inv(3)/CBL-2 60,810,627
inv(3)/RAS-1 "176,938,868"
RUNX1-D-5 193,586,370
RUNX1-T/CEBPA-2 138,443,303
RUNX1-DT-1 387,658,545
RUNX1-DT/CEBPA-2 28,292,155
RUNX1-DT-3 22,760,769
RUNX1(x2)-D&T-4 257,468,842
ITD(2x)/NPM1-2 216,542,686
ITD(2x)/NPM1-1 190,804,956
NPM1-2 227,981,679
NPM1-1 280,521,132
ITD/NPM1-2 172,152,278
ITD/NPM1-3 311,033,458
ITD/NMP1-1 "176,091,621"
ITD/NPM1-6 "496,935,874"
ITD/NPM1-5 251,705,752
ITD/NPM1-4 196,992,053
ITD-2 129,490,450
ITD-3 "368,474,143"
ITD-1 212,807,508
CEBPA(x2)-3 237,827,166
CEBPA(x2)-2 195,788,368
CEBPA(x2)-1 188,824,613
t(8;21)-4 405,680,774
t(8;21)/KIT-2 234,428,779
t(8;21)-3 254,489
t(8;21)-1 214,032,566
t(8;21)-1 rel 29,269,736
inv(16)-3 325,349,713
inv(16)-2 209,458,324
inv(16)-1 28,000,121
RUNX1-T-7 (NHL) 168,424,124
NPM1/RAS-3 159,361,216
PBSC-1 173,261,382
PBSC-2 205,284,104

7. Describe the antibodies used for the ChIP-seq experiments.

(NA) as these are DNASeI-seq experiments

8. Describe the peak calling parameters.

Here is the command used for read mapping and peak calling

```
bowtie2 -p 4 --very-sensitive-local -x hg38 -U DNASE1.fastq -S DNASeI.sam
samtools view -bt hg38.fa.fai -o DNASeI.bam DNASeI.sam
macs2 callpeak --t DNASeI.bam -f BAM --nomodel --call-summits -q 0.005 -n DNASeI_macs2_q_0.005
```

9. Describe the methods used to ensure data quality.

DNase-seq sequences from all experiments were mapped onto the reference human genome version hg38, with Bowtie version 2.3.1. Low quality reads were trimmed using trimmomatic prior to the alignment and the quality control (QC) statistics for the samples were obtained using FastQC tools.

DNase I Hypersensitive Sites (DHSs) were called with MACS2 using callpeak function (nomodel, call-summits and q = 0.005 parameters). To define a common set of coordinates covering all of the significant DHSs investigated in this study, we merged all of the individual DNase-Seq reads for all of the AML samples assayed by DNase-seq. This data set was then used to define the peak summits of the union of all peaks. This approach was designed to maximize...
the precision and sensitivity of the peak detection. The DNA read counts were then determined within 400 bp windows centered on each peak for each sample. To account for the different number of reads in each of the samples; the read counts were normalized for total read depth using DEseq2.

Sample code number of peaks
RUNX1-T/JAK-2 37078
RUNX1-D/JAK-1 39913
inv(3)/CBL-2 50167
inv(3)/RAS-1 51146
RUNX1-D-5 45624
RUNX1-T/CEBPA-6 54020
RUNX1-DT-1 43328
RUNX1-DT/CEBPA-2 50423
RUNX1-DT-3 50770
RUNX1(x2)-D&T-4 39008
ITD(2x)/NPM1-2 41083
ITD(2x)/NPM1-1 57180
NPM1-2 29151
NPM1-1 53055
ITD/NPM1-2 42176
ITD/NPM1-3 53911
ITD/NMP1-1 43248
ITD/NPM1-6 53427
ITD/NPM1-5 47032
ITD/NPM1-4 47957
ITD-2 50485
ITD-3 54989
ITD-1 63293
CEBPA(x2)-2 35666
CEBPA(x2)-2 54508
CEBPA(x2)-1 45723
t(8;21)-4 48657
t(8;21)/KIT-2 45318
t(8;21)-3 59155
t(8;21)-1 49755
t(8;21)-1 rel 46773
inv(16)-3 52057
inv(16)-2 55820
inv(16)-1 35666
RUNX1-T-7 (NHL) 48159
NPM1/RAS-3 53479
PBSC-1 43607
PBSC-2 39347

10. Describe the software used to collect and analyze the ChIP-seq data.

AML and PBSC samples used in this study were either surplus diagnostic samples, or were fresh samples obtained with specific consent from the patients. AML samples were obtained from either (i) the Haematological Malignancy Diagnostic Service (St James’s Hospital, Leeds, UK, (ii) the Centre for Clinical Haematology, Queen Elizabeth Hospital Birmingham, Birmingham, UK, (iii) the West Midlands Regional Genetics Laboratory, Birmingham Women’s NHS Foundation Trust, Birmingham, UK, or from iv) Erasmus University Medical Center, Rotterdam, The Netherlands.

The list of software used to analyze DNAseq data
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: Bowtie version 2.3.1. (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml)
3. For peak calling MACS version 2 (https://github.com/taoliu/MACS)
4. For normalizing to the total read depth among more than one samples DEseq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
5. Digital genomic footprinting was performed using the Wellington_footprints function of the Wellington algorithm (http://pythonhosted.org/pyDNase/)

List of software used to analyse RNA-seq data
1. For quality check: fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. For read mapping: STAR (https://github.com/alexdobin/STAR)
3. For FPKM values cufflinks (http://cole-trapnell-lab.github.io/cufflinks/)
4. For differential expression analysis: (http://bioconductor.org/packages/release/bioc/html/limma.html)