Mutational synergy during leukemia induction remodels chromatin accessibility, histone modifications and three-dimensional DNA topology to alter gene expression

Haiyang Yun1,2,3, Nisha Narayan1,2, Shabana Vohra1,2, George Giotopoulos1,2, Annalisa Mupo1,2,4,8, Pedro Madrigal1,2, Daniel Sasca1,2,5, David Lara-Astiaso1,2, Sarah J. Horton1,2, Shuchi Agrawal-Singh1,2, Eshwar Meduri1,2, Faisal Basheer1,2, Ludovica Marando1,2, Malgorzata Gozdecka1,2,4, Oliver M. Dovey1,2,4, Aracely Castillo-Venzor1, Xiaonan Wang1,2, Paolo Gallipoli1,2,6, Carsten Müller-Tidow3, Cameron S. Osborne7, George S. Vassiliou1,2,4 and Brian J. P. Huntly1,2

Altered transcription is a cardinal feature of acute myeloid leukemia (AML); however, exactly how mutations synergize to remodel the epigenetic landscape and rewrite three-dimensional DNA topology is unknown. Here, we apply an integrated genomic approach to a murine allelic series that models the two most common mutations in AML: Flt3-ITD and Npm1c. We then deconvolute the contribution of each mutation to alterations of the epigenetic landscape and genome organization, and infer how mutations synergize in the induction of AML. Our studies demonstrate that Flt3-ITD signals to chromatin to alter the epigenetic environment and synergizes with mutations in Npm1c to alter gene expression and drive leukemia induction. These analyses also allow the identification of long-range cis-regulatory circuits, including a previously unknown superenhancer of Hoxa locus, as well as larger and more detailed gene-regulatory networks, driven by transcription factors including PU.1 and IRF8, whose importance we demonstrate through perturbation of network members.

The functional specification of tissues in metazoans occurs through the generation of cell-specific proteomes, in turn controlled by diverse transcriptional programs1,2. These transcriptional programs are tightly spatiotemporally regulated through the use of tissue-specific cis-regulatory elements such as enhancers, which are licensed and brought into direct communication with their cognate promoters through the functions of transcription factors (TFs), chromatin regulators (CRs) and genome structural proteins3–8. These players are known to remodel the epigenetic landscape and reshape genome topology to allow communication between regulatory elements (for example promoter-enhancer).

In malignancies such as AML, recurrent mutations in TFs, CRs and genome structural proteins, including components of the Cohesin complex and CTCF have been observed at high frequencies9–14. AML is the most common acute leukemia in adults, serves as a paradigm for aberrant hematopoietic differentiation and is an aggressive disease with a dismal overall survival rate of <30%15. However, although aberrant transcription is a cardinal feature of AML16, not all cases carry mutations in the above classes of proteins, indicating that the effects of other common mutations, such as signaling alterations, indirectly converge on the same epigenetic, transcriptional and genome regulatory machinery, although the mechanistic detail of this remains obscure.

This study addresses these fundamental questions of how AML mutations, even those that do not directly control gene expression, co-opt the transcriptional and epigenetic machinery to alter chromatin states, three-dimensional (3D) DNA topology and communication between enhancers and promoters to generate leukemia-specific transcriptional programs. To do so we have experimentally ‘deconstructed’ AML using an allelic series of mice that model different ‘transition states’ during AML induction; normal, premalignant and overt leukemia. In addition, analysis of the single mutant (SM) premalignant mice also allows deconvolution of the contribution of individual mutations to altered epigenetic regulation. These studies allow us to elucidate the interplay between epigenetic and 3D genomic states and the generation of leukemia-specific transcriptional programs.
Fig. 1 | Murine models show transcriptional synergy of AML mutations. a, Schematic of overall experimental design. b, PCA of mRNA expression data of 16,771 protein-coding genes. PC, principal component. c, Volcano plots showing differential expression between mutant and WT HSPC. The P values were attained by the Wald test from DEseq2 and were two-tailed and corrected for multiple testing using the Benjamini and Hochberg method (adjP). Up- or downregulation were defined by setting adjP < 0.05 and absolute FC ≥1.5. d, Immune-response-related Hallmark gene sets from GSEA analysis of differential gene expression in Flt3-ITD or DM HSPC. e, GSEA enrichment plots showing gene set of tumor necrosis factor alpha (TNFA) signaling via nuclear factor kappa B (NFKB) in Flt3-ITD or DM HSPC. NES, normalized enrichment score. f, Overlap of upregulated genes in each mutant. Numbers of genes in each segment and representative genes are indicated. Hypergeometric test P values (one-tailed, not multiple testing corrected) are shown. ov, overlap. g, Heatmap of normalized mRNA expression (Z-score) of representative genes in WT and mutant HSPC with replicate samples. h, Overlapping analysis between Npm1c/Flt3-ITD up-or downregulated genes (with a twofold change) in human AML and mouse DM leukemia. Hypergeometric test P values (one-tailed, not multiple testing corrected) are shown.
Results

Murine models show transcriptional synergy of AML mutations. Mutations in FLRT3-ITD and NPM1c are the most common in AML, occurring individually in ~25–30% and co-occurring in ~15% of all cases1–4. Mouse models carrying knock-in mutations of Npm1 (conditional Npm1flx-xflx;Mx1-Cre, hereafter ‘Npm1c’) or Flt3 (constitutive Flt3ITD), hereafter ‘Flt3-ITD’) and referred to as SM mice, are associated with individual subtype, nonfatal but obvious premalignant phenotypes4–11. However, combined Npm1c/ Flt3-ITD or double mutant (DM) mice develop an aggressive AML with short latency16. To prospectively assess dynamic remodeling of the cis-regulatory landscape and 3D genome topology during leukemia development, we used an allelic series of wild-type (WT), Npm1c, Flt3-ITD and DM mice to model AML induction (Fig. 1a, upper panel), analyzing the same population enriched for hematopoietic stem and progenitor cell (HSPC, negative for lineage markers, Lin−) from WT and mutant mice. Assessing gene expression (RNA-seq), multiple chromatin modification states by chromatin immunoprecipitation and sequencing (ChIP-seq), chromatin accessibility by assay for transposon accessible chromatin (ATAC-seq)17 and promoter-anchored 3D chromatin interaction by promoter capture HiC (high-throughput chromosome conformation capture) (pCHiC)18 (Fig. 1a, lower panel) across our allelic series, we derived a high-quality dataset with strong reproducibility between sample replicates (Supplementary Fig. 1).

We first analyzed differential gene expression between WT, individual premalignant SM and leukemic DM stages. Principal component analysis (PCA) and pairwise comparisons demonstrated that the two single mutations induced only very modest changes of global gene expression in isolation (Fig. 1b–c, Extended Data Fig. 1a and Supplementary Table 1). In stark contrast, when combined, Npm1c and Flt3-ITD synergized strongly to induce marked differential gene expression. Using gene set enrichment analysis (GSEA), Flt3-ITD HSPC showed signatures related to immune activation that were shared with DM Leukemic HSPC (Fig. 1d,e and Extended Data Fig. 2d–f). Further bivariate genomic footprinting (BaGFoot) analysis19 simultaneously assessed differential footprint depth of TFs and altered accessibility flanking their motifs between conditions (Fig. 2g and Extended Data Fig. 2g,h). This confirmed that GATA factors demonstrated decreased accessibility and reduced footprint depth, indicating a loss of binding in DM HSPC (Fig. 2g). Conversely, increased accessibility and footprint depth was noted for Gbp3, Cebp and for the AP-1 complex members Atf2 and Atf7. Of interest, increased accessibility but no increase in footprinting was observed for PU.1 motifs in DM cells, indicating a potential role for previously bound PU.1 in determining this accessibility upon leukemia development.

Mutations in Flt3-ITD and Npm1c alter chromatin accessibility. Transcriptional activation is facilitated by the recruitment of specific TFs leading to nucleosome depletion at cis-regulatory elements such as enhancers and promoters20. Using ATAC-seq, we therefore first assessed the dynamics of chromatin accessibility across our allelic series. In contrast to the minimal effects on gene expression conferred by individual mutations, both single mutations demonstrated marked alterations in accessibility, though to a less extent than in combination (Fig. 2a,b and Supplementary Table 2). Compared with gene expression, more alterations in accessibility were preserved between the Flt3-ITD and DM HSPC (Fig. 2e). In comparison with the degree of increased or decreased chromatin accessibility elicited by DM at differentially accessible regions, changes were intermediate in Flt3-ITD HSPC (Fig. 2d). In contrast, less overlap was seen for the changes of accessibility evident between Npm1c and DM HSPC (Fig. 2c and Extended Data Fig. 2a). However, within this overlap were four regions located in the Hoxa locus (Fig. 2c), linking accessibility with an increase in gene expression in both Npm1c and DM HSPC. Regions that became sequentially less accessible in Npm1c and DM HSPC included the Gata2 promoter, as well as its distal and proximal enhancers21–23 (Extended Data Fig. 2b). Interestingly, accessibility changes in single mutant HSPC did not alter gene expression remarkably, but did correlate with gene expression upon leukemia induction (Extended Data Fig. 2c).

Performing de novo motif analysis, we identified binding sites of TFs whose function may be modulated by these accessibility changes, and demonstrated enrichment for the binding sites of a number of TFs at regions gained, lost or static (Fig. 2f and Extended Data Fig. 2d–f). Further bivariate genomic footprinting (BaGFoot) analysis19 simultaneously assessed differential footprint depth of TFs and altered accessibility flanking their motifs between conditions (Fig. 2g and Extended Data Fig. 2g,h). This confirmed that GATA factors demonstrated decreased accessibility and reduced footprint depth, indicating a loss of binding in DM HSPC (Fig. 2g). Conversely, increased accessibility and footprint depth was noted for Gbp3, Cebp and for the AP-1 complex members Atf2 and Atf7. Of interest, increased accessibility but no increase in footprinting was observed for PU.1 motifs in DM cells, indicating a potential role for previously bound PU.1 in determining this accessibility upon leukemia development.

Only Flt3-ITD remodels the histone modification landscape. We next assessed chromatin modifications associated with regulatory activity (mono- or trimethylation of histone 3 lysine 4, H3K4me1 or H3K4me3 and acetylation of histone 3 lysine 27, H3K27ac) at cis-regulatory elements in WT and mutant HSPC using ChIP-seq.

Fig. 2 | Both Flt3-ITD and Npm1c alter chromatin accessibility. a, PCA of ATAC-seq signals across all four cellular states. Counts per million (CPM) reads were used for the analysis. b, MA plots showing ATAC-seq peaks with significantly differential accessibility in mutant versus WT HSPC. Significant increase or decrease was determined by edgeR with setting FDR (two-tailed and multiple testing corrected) <0.05 and absolute FC ≥2.5. c, Venn diagram of ATAC-seq peaks with accessibility gain or loss induced by each mutant. Hypergeometric test P values (one-tailed, not multiple testing corrected) are shown. ov, overlap. d, Profile plots of ATAC-seq signals at regions demonstrating gain or loss of accessibility induced by DM across four cellular states. e, ATAC-seq tracks showing chromatin accessibility at the Hoxa cluster in all four HSPC and WT NE. Regions showing increased accessibility by Npm1c and DM compared with WT are highlighted. f, De novo motifs significantly enriched at genomic regions with gain or loss of accessibility in DM HSPC. HOMER outputs motifs with target coverage >10% and ranked by P values (one-tailed, not multiple testing corrected). g, BaGFoot analysis illustrates TFs with differential footprint depth and accessibility in DM versus WT HSPC. Data points within bag and fence area, including 50% and >97% of the population, respectively, are not significant (NS). Motifs outside the fence and with a P value (two-tailed, not multiple testing corrected) <0.05 are statistically significant outliers.
Enhancers were considered cis-regulatory modules usually flanked by high amounts of H3K4me1 and no, or low, amounts of H3K4me3 (n = 98,365, Extended Data Fig. 3a and Supplementary Table 3) and were designated ‘primed’ when lacking H3K27ac or ‘active’ if marked with H3K27ac\(^2\). To further enrich for cis-regulatory elements, we overlapped this compendium with our ATAC-seq analysis.

**Figure a**
ATAC-seq peaks
- Accessibility gain in DM
- Accessibility loss in DM

**Figure b**
- Npm1c versus WT
- Flt3-ITD versus WT
- DM versus WT

**Figure c**
ATAC-seq peaks
- Accessibility gain
- Accessibility loss

**Figure d**
Accessibility gain in DM
- WT
- Npm1c
- Flt3-ITD
- DM

**Figure e**
ATAC-seq
- WT
- Npm1c
- Flt3-ITD
- DM
- NPM1c

**Figure f**
Motif
- PU.1
- RUNX
- NFIL3/CEBP
- AP-1
- GATA
- CTCF
- KLF
- NF1-halfsite

**Figure g**
DM versus WT
- Gata

**Table**

| Motif | Best match | Targets | \(P\) value |
|-------|------------|---------|-------------|
| PU.1  | 50.3%      | 1 \times 10^{-276} |
| RUNX  | 39.4%      | 1 \times 10^{-1371} |
| NFIL3/CEBP | 18.6% | 1 \times 10^{-348} |
| AP-1  | 11.9%      | 1 \times 10^{-288} |
| GATA  | 50.3%      | 1 \times 10^{-2914} |
| CTCF  | 12.9%      | 1 \times 10^{-1419} |
| KLF   | 14.4%      | 1 \times 10^{-229} |
| NF1-halfsite | 22.9% | 1 \times 10^{-38} |

**Figure h**
Regions showing increased accessibility by Npm1c and DM
(Chr6:52,147,740–52,262,470)

**Figure i**
Regions showing increased accessibility by Npm1c and DM
(Chr6:52,147,740–52,262,470)
to identify ‘accessible’ enhancers \((n = 48,052;\) Extended Data Fig. 3b and Supplementary Table 3). This cis-regulatory element repertoire was highly dynamic across the cellular states; \(\sim 51\%\) demonstrated differential H3K4me1 between mutant and WT HSPC (Extended Data Fig. 3c). A large number of dynamic alterations of H3K4me1 were observed for both Flt3-ITD and DM HSPC, with an increased number in the DM leukemic state (Fig. 3a, b and Supplementary Table 4) but a marked overlap between the two (Extended Data Fig. 3d). Strikingly, and in marked contrast, Npm1c HSPC were virtually indistinguishable from WT in H3K4me1 patterns (Fig. 3a, b). Moreover, among all DM dynamic elements, intermediate changes of H3K4me1 were observed in Flt3-ITD but not Npm1c HSPC (Fig. 3c and Extended Data Fig. 3e).

We next assessed activation of these elements by overlaying the H3K27ac changes on the H3K4me1 and accessibility landscape. H3K27ac alterations were relatively modest (Extended Data Fig. 3f, g). Flt3-ITD alone and DM demonstrated H3K27ac changes at hundreds of enhancer regions, whereas Npm1c did not notably alter global H3K27ac signal (Fig. 3d and Supplementary Table 4). Once again, the effects of Flt3-ITD alone were often intermediate when compared with DM, as exemplified at the Socs2 locus (Fig. 3e and Extended Data Fig. 3h).

Although no major alterations in the cellular composition of the broader Lin- HSPC compartment are evident between WT and the SM mice, this compartment is larger in DM mice and demonstrates a relative increase in LSK (Lin-‘Sca1/cKit’, enriched for HSC) and particularly the granulocyte/macrophage progenitors (GMP) compartments, as well as a modest reduction in the common myeloid progenitors (CMP) and megakaryocyte/erythroid progenitors (MEP) compartments (Extended Data Fig. 4a, b).

Therefore, to address whether altered cis-regulatory elements were truly leukemia-specific or were associated with other hematopoietic states, such as normal myeloid differentiation, we replicated our integrated genomic analysis in normal neutrophils (NE) and included similar available GMP maps\(^{[10]}\) for comparison. For DM gained enhancers, \(31\%\) (3,637/11,892) overlapped with similar changes in NE and GMP (Fig. 3f, group Gain-1). However, the remaining \(69\%\) seemed leukemia-specific (group Gain-2, Supplementary Table 5). In contrast, most DM lost enhancers showed similar changes in both NE and GMP (83%, 10,362/12,509, group Loss-2), while \(\sim 17\%\), although distinct from NE, were highly similar to GMP (group Loss-1). Exemplar leukemia-specific regulatory elements are shown for the Socs2 gene (Fig. 3e). Of note, subtle specific enrichment for motifs were observed for leukemia-specific elements, in particular, the presence of an ETS-IRF composite element for PU.1/IRF in the leukemia-specific enhancer regions (Fig. 3g).

Leukemia programs use both new and existing 3D contacts. To capture genome-scale alterations in 3D DNA topology, we used pCHiC \(^{[26]}\) to generate a compendium of promoter-associated DNA interactions and demonstrate how they differed between WT and mutant HSPC. A total of 88,624 high-confidence interactions were captured across our HSPC series (Fig. 4a, Extended Data Fig. 5a–c and Supplementary Table 6).

pCHiC data can subdivide the genome into A (active) and B (inactive) compartments using PCA\(^{[26,27]}\). Only the combination of mutations altered the compartment structure, with several hundred compartments changing their assignment (Fig. 4b, c and Supplementary Table 7). A total of 290 regions altered their compartment assignment from ‘inactive’ B to ‘active’ A (Fig. 4c).

These regions contained 345 expressed genes, including known oncogenes such as Setbp1, Ifi1 and a contiguous region involving an interferon-inducible Ifi gene cluster (Fig. 4d and Extended Data Fig. 5d, e). Alteration from an active to an inactive compartment, A to B, occurred at 274 regions (Fig. 4c) and covered 453 genes including the tumor suppressor gene Dach1 and the Fanconi gene that is silenced by methylation in AML\(^{[19]}\).

Assessing total interactions at individual promoters, only a few promoters were altered for Npm1c HSPC in comparison with WT. In contrast, in the presence of Flt3-ITD, several hundred promoters showed lost or gained interactions, with this number roughly doubling in DM HSPC (Fig. 4e). Moreover, nearly half of those promoters altered by Flt3-ITD overlapped with the change in DM HSPC (Fig. 4f). At the level of individual interactions, only \(11\%\) of all high-confidence interactions were significantly differential (gained or lost) in SM or DM HSPC, based on stringent alterations of the ranking of CHiCAGO scores\(^{[26]}\) (Fig. 4g), with more interactions lost than newly formed in the mutant states (Fig. 4h and Supplementary Table 8). We named these ‘rewired’ interactions, and these interactions could be either lost or gained (for example Gfi1b or Ifi8, respectively; Fig. 4i and Extended Data Fig. 5f, g).

In contrast, the remainder of interactions were already preset in WT HSPC and were termed ‘hard-wired’ DNA contacts (Fig. 4g).

Integrated analysis identifies critical regulatory networks. We first applied a multilayered approach to integrate the multimicroarray chromatin analysis\(^{[15]}\) at all cis-regulatory regions across WT and mutant HSPC. Briefly, focusing on 2-kb bins of accessible regions (\(\pm 1\text{ kb from ATAC-seq peak summit}, n = 75,457\)), normalized reads for chromatin modifications and accessibility of each HSPC sample were integrated. Treating these bins as a separate data points across all 16 conditions (four analyses in four cellular states), we were then able to use dimensionality reduction tools developed for scRNA-Seq (Seurat\(^{[32]}\)) to classify and visualize clusters of regions showing similar patterns (Fig. 5a). This classified ten individual clusters (Fig. 5a, b). Cis-regulatory regions with a loss of enhancer signatures during leukemia induction, characterized by concurrent loss of H3K4me1 and accessibility, with or without evident loss of H3K27ac (Cluster-9 or -10, respectively; Fig. 5b, c, Extended Data Fig. 6a and Supplementary Table 9). Conversely, cis-regulatory regions were also identified with variable gains of enhancer marks and accessibility, which were separated by marked gain of H3K27ac (Cluster-6) and accessibility (Cluster-7), as well as pre-established H3K4me1 in WT (Cluster-8). Regions with patterns reflecting a loss of marks associated with enhancer function (Clusters 9, 10) were enriched for GATA and KLF factors (Fig. 5d and Extended Data Fig. 6b). For regions that gained marks associated with enhancer function (Clusters 6–8), enrichment for PU.1, CEBP, RUNX and AP-1 motifs were seen.
The chromatin regions with gain or loss of enhancer signatures (Cluster-6 or -10) were further linked to their linearly localized genes or with spatially contacted genes using chromatin interaction profiles. Genes upregulated (n=978) in Cluster-6 demonstrated significant enrichment for immune process (Fig. 5e). These included several genes of interest for leukemogenesis, including Jun, Fos, Hoxa10 and Irf8 (Fig. 5f and Supplementary Table 10). By contrast, genes downregulated in Cluster-10 were enriched for DNA
Fig. 4 | Leukemia programs use new and existing 3D contacts. a. PCA of CHiCAGO scores of high-confidence interactions across all four HSPC states. High-confidence interactions were defined as significant interactions (CHiCAGO score ≥5) present in both replicates. b. Pan-chromosome correlation of chromatin compartments between mutant and WT HSPC. Dots represent individual promoters and the correlation of their PC1 values. c. Numbers of gained or lost interactions in the presence of mutations (compared with WT HSPC). h, Illustration of chromatin compartments A/B levels at a DM-induced ‘B to A’ flipped region (highlighted) containing Setbp1. e, Differential total interaction reads at individual promoters between mutant and WT HSPC. The P values were two-tailed and corrected for multiple testing (adjP). Increase or decrease were determined by setting adjP < 0.05 and absolute FC > 1. f. Scatter plot showing correlation between promoters with altered interactions induced by either Flt3-ITD (y axis) or DM (x axis). Hypergeometric test P values (one-tailed, not multiple testing corrected) are shown. g, Proportion of rewired and hard-wired interactions among all high-confidence interactions across four HSPC states. h, Number of gained or lost interactions in the presence of mutations (compared with WT HSPC). i, j, Promoter-contact plots showing the read counts of promoter bait to target pairs for Gfi1b (i) and Irf8 (j). Dots represent chromatin interaction fragments at defined distances from the bait; gray lines, expected read counts; dashed lines, the upper bound of the 95% confidence intervals. Regions highlighted yellow show loss (i) or gain (j) of interactions in DM versus WT HSPC.
replication and contained hematopoietic TFs including *Gata1*, *Gata2*, *Gfi1b* and *Myb* (Extended Data Fig. 6c,d). To further demonstrate relevance for human disease, we used a publicly available dataset of upregulated genes linked to cis-regulatory elements with increased accessibility in *NPM1c/FLT3-ITD* mutated human AML. We could demonstrate that over a quarter of the human genes (275/1031, 26%, $P = 2.58 \times 10^{-6}$) were present within our Cluster-6 genes, demonstrated enriched signatures for immune regulatory processes and many showing increased gene expression in our system (Fig. 5g–i and Extended Data Fig. 6e).

We next investigated the genome-wide relationship between chromatin states, chromatin interactions and gene expression programs. As expected, alterations of enhancer-associated marks correlated globally with gene expression upon leukemia induction, albeit modestly (Extended Data Fig. 7a). Chromatin interaction frequencies reflected by the flipping of genome compartments generally correlated with gene expression (Fig. 6a and Extended Data Fig. 7b). Integrating chromatin modifications with differential interactions, we demonstrated a strong correlation of H3K4me1 alteration with ‘rewired’ interactions (Fig. 6b and Extended Data Fig. 7c,d). Similarly, flipping of genome compartments associated positively with alteration of enhancer marks including H3K4me1 and H3K27ac, as well as chromatin accessibility (Fig. 6c).

Our analysis also identified 801 superenhancers (SE) across WT and mutant HSPC (Extended Data Fig. 8a and Supplementary Table 11). Patterns of loss and gain were similar to standard enhancers when linked to mutations (Extended Data Fig. 8b,c and Supplementary Table 11). However, for *Npm1c* one of the three SE gained demonstrated a long-range contact to the *Hoxa* locus and was associated with upregulation of most *Hoxa* genes (Extended Data Fig. 8d). Linking altered SE to their regulated genes using DNA interaction profiles showed a strong correlation with expression during DM leukemia induction (Extended Data Fig. 8d). Genes upregulated in DM HSPC were again found to be enriched for inflammatory response, as well as myeloid differentiation, while downregulated genes were enriched for erythroid differentiation (Extended Data Fig. 8e).

**Leukemic regulation of exemplar reprogrammed loci.** Our integrated analysis suggested *Sp1/Pu.1* as an important network regulator in *Flt3-ITD/Npm1c* leukemia (Fig. 6d and Extended Data Fig. 8f). All WT and single mutant HSPC demonstrated H3K4me1 priming and accessibility at the well-described *Sp1* upstream regulatory element (URE) but increased accessibility of the URE to communicate with the *Sp1* promoter and upregulate its expression. Irf8 was similarly identified as a likely network player. A slightly different chromatin pattern was observed to be associated with ‘rewiring’ and increased gene expression at the DM stage, with increased H3K4me1 and accessibility observed in both *Flt3-ITD* and DM HSPC, but no alteration of H3K27ac until the DM stage (Figs. 6e and 4i).

Definitively linking promoters to cis-regulatory regions using pCHiC also allowed us to identify new long-range regulatory interactions. An exemplar of this was seen for the *Hoxa* locus, where ‘hard-wired’ interactions with an H3K4me1 modified region ~1 Mb upstream (Fig. 6f and Extended Data Fig. 8g) could be demonstrated for all WT and mutant HSPC. However, as previously described, this same region, which we have named the *Hoxa*-long-range SE (*Hoxa*-LRSE), demonstrated a marked increase in H3K27ac modification in both *Npm1c* and in DM HSPC, where it seems to collaboratively interact with the increased accessibility across the promoters of *Hoxa* genes (Supplementary Fig. 3a). In addition, and as a direct link to human AML, this region is syntenic with a homologous region on human chromosome 7, where the same interaction can be demonstrated (Supplementary Fig. 3b). Using published data (including ChIP-seq on H3K4me1 (ref. 1) and CTCF (ref. 6)), together with pCHiC of human CD34+ HSPC, we could demonstrate that the interaction was also present across species, and ChIP-seq data from ourselves and others indicated that this region also contained a SE in human AML cells that overexpress *HOXA* genes (OCI-AML3 cell line) but not in cells that do not (Kasumi-1 cell line) (Supplementary Fig. 3b). The human *HOXA*-LRSE was also demonstrated to have an open chromatin configuration in *NPM1* mutant AML patients by DNase hypersensitivity analysis (Supplementary Fig. 3b), further demonstrating cross-species conservation of this long-range element.

**Network perturbation abrogates leukemia maintenance.** Using our integrated analysis to identify exemplar putative network nodes, we focused on the AP-1 complex members c-Fos and c-Jun, the TFs *Sp1/Pu.1, Irf8*, the oncogene *Igf1*, the *Hoxa* genes and the regulatory loci of *Sp1* and the *Hoxa* cluster, validating their critical role in the maintenance of *Flt3-ITD/Npm1c* AML using RNAi and CRISPR editing. As proof of this principle, knockdown of c-Jun and c-Fos decreased clonogenic capacity significantly in murine DM cells in vitro (Extended Data Fig. 9a,b). Depletion of *Sp1* could abrogate leukemia cell growth and clonogenicity (Extended Data Fig. 9c–e). To validate *Sp1* regulatory elements as critical for leukemia maintenance, CRISPR–Cas9 mediated genetic excision of the *Sp1*-URE was achieved using dual guide RNAs (gRNAs) to target its 5 kb central region in DM cells carrying Cas9 (ref. 42) (Fig. 7a and Extended Data Fig. 9f,g). Removal of the URE resulted in ~30% reduction of *Sp1* expression in bulk cultures (Fig. 7b) and significantly decreased cell growth and clonogenicity (Fig. 7c,d and Extended Data Fig. 9h).

The requirement for sustained expression of *Hoxa9* or *Hoxa10* for the maintenance of leukemia cell growth and clonogenicity was also demonstrated (Extended Data Fig. 9c–e). We used a similar experimental strategy to examine the role of the *Hoxa*-LRSE by removing its ~2.4 kb central region in DM-Cas9 cells (Fig. 7e and Extended Data Fig. 9i,j). Genetic disruption of *Hoxa*-LRSE resulted in significantly reduced expression of all *Hoxa* genes tested (*Hoxa3, 5, 6, 7, 9 and 10; Fig. 7i). In addition, excision of the *Hoxa*-LRSE also induced a significant decline in cell proliferation and reduced colony-forming capacity (Fig. 7g,h and Extended Data Fig. 9k).

To further correlate the relevance of these networks to human AML, we also validated *Sp1/Pu.1* as well as extending the analysis to other targets, including *IRF8* and *IGF1*. These targets were
**Figure a:** Accessible chromatin regions

**Figure b:** Heatmap showing normalized log2 reads for different conditions.

**Figure c:** Bar graphs illustrating the comparison between WT and experimental groups.

**Figure d:** Table listing de novo motifs, their best matches, target genes, and p-values.

**Figure e:** Bar chart representing the upregulation of immune system process genes in DM versus WT.

**Figure f:** Volcano plot comparing DM versus WT with gene expression fold change (FC) and adjusted p-values.

**Figure g:** Venn diagram showing the overlap between mouse and human genes.

**Figure h:** Bar chart showing the enrichment of regulatory processes in Cluster-6 genes with human overlap.

**Figure i:** Additional volcano plot focusing on Cluster-6 genes with human overlap.
Fig. 6 | Integrated analyses identifies critical transcriptional networks nodes. a, Differential expression of genes involved in flipped chromatin compartments in DM versus WT HSPC. Up- or downregulation were determined by setting adjP (two-tailed and multiple testing corrected) <0.05 and absolute FC ≥1.5. b, Comparison of H3K4me1 alterations between WT and DM states at enhancers that gained (upper graph) or lost (lower graph) interactions. Increase or decrease were defined by setting adjP (two-tailed and multiple testing corrected) <0.05 and absolute FC ≥1.5. c, Enrichment of chromatin marks and accessibility involved in flipped chromatin compartments in DM versus WT HSPC. d–f, Combined profiles of chromatin states, accessibility at enhancers, target mRNA expression, as well as 3D interactions between enhancers and target promoters for Spi1 (d), Irf8 (e) and Hoxa cluster (f) in all four HSPC samples and WT NE. Significant chromatin interactions were defined by CHICAGO score ≥5 and are represented by arcs.
Fig. 7 | Network perturbation abrogates leukemia maintenance. a, Experimental strategy targeting Spi1-URE for deletion using CRISPR-Cas9 plus dual gRNAs flanking its ~5 kb central region. Primer 1 and 2 are oligonucleotides for PCR confirmation of deletion. b, Spi1 mRNA expression detected by quantitative PCR with reverse transcription (RT-qPCR) in DM-Cas9 cells expressing Spi1-URE gRNAs relative to control (Ctrl.) gRNAs (n = 3 independent experiments). c, Ex vivo proliferation (c) and CFU assays (d) of DM-Cas9 cells expressing Spi1-URE gRNAs relative to control gRNAs (Ctrl. gRNA). d, CFU assays of OCI-AML3 cells expressing human gRNAs targeting SPI1, IGF1, and IRF8 relative to control gRNAs in methylcellulose culture (n = 3 independent experiments). # indicates two independent gRNAs. Table S4 describes the number of colonies (per 10^3 cells) associated with proliferation relative to Ctrl.gRNA. Statistical analyses in b–d and f–l were performed by running Student’s unpaired t-tests and P values (two-sided, not multiple testing corrected) are shown; error bars represent mean ± s.e.m.
edited via CRISPR–Cas9 in the human HOXA-dependent AML cell lines OCI-AML3 (mutated for NPM1, but also DNMT3A) and MOLM-13 (that carries a mutation in FLT3-ITD, but also an MLL-AF9 rearrangement). Importantly, these experiments corroborated our murine findings, with editing decreasing proliferation and clonogenic function (Fig. 7i–l and Extended Data Fig. 10). Taken together, perturbation of these exemplar nodes validate our integrated analysis strategy to identify critical network members and their cis-regulatory elements and demonstrated these nodes to be critically required for the maintenance of the leukemia.

Discussion
AML is associated with a relatively uncomplicated genome by comparison with other tumors, with each case harboring, on average, only between three and five mutations. Our study provides mechanistic evidence of why several mutations are necessary for the full malignant phenotype, and suggests that marked synergism between them may explain their relatively low number. By deconstructing an experimental model of a common AML subtype, we could show that single mutations lead to only modest changes in the epigenetic landscape that do not translate into significant alterations of gene expression. However, when mutations co-occur, their combinatorial effects produce marked synergy at every epigenetic level examined. Moreover, our study confirms prospectively and categorically that common AML mutations that lack direct epigenetic or transcriptional effects can indirectly alter the epigenetic landscape and transcription to generate convergent leukemia-associated transcriptional programs. Overlap between the Flt3-ITD and DM states suggests Flt3-ITD to have a dominant role in malignant remodeling of the epigenome and 3D genome. Moreover, overt gene expression changes seen in DM cells were often preceded by alterations in chromatin modification, usually of H3K4me1, in regulatory elements associated with the same genes in Flt3-ITD HSPC, observations reminiscent of epigenetic changes preceding gene expression in normal hematopoietic differentiation41. These stepwise changes may mark the leukemogenic ‘potential’ of these elements to be subsequently modified further by the actions of a collaborating mutation. Mechanistically, GSEA and motif analysis for differential chromatin modifications, particularly H3K4me1, linked Flt3-ITD signaling to upregulated inflammatory response gene programs. Moreover, de novo motif analysis, RNAi and CRISPR editing experiments demonstrated the role of signal-inducible TFs such as AP-1 (refs. 42,44), interferon response factors (particularly IRF-4 and -8)45 and signal responsive TFs such as PU.1 (ref. 46) and RUNX1 (ref. 47) in this regulation. These findings accord with, and extend, similar genomic analysis in patient samples from FLT3-ITD mutated AML19,47. By contrast, the main alteration that occurred in HSPC expressing Npm1c was an alteration of chromatin accessibility, with an obvious exemplar the changes at the HOXA cluster demonstrated to drive leukemogenesis. The nature of this alteration in chromatin accessibility is not immediately apparent; however, it may relate to the known histone chaperone function of NPM1 (ref. 48), to its interaction with ATP-dependent chromatin remodeling complexes49 or to its role as the main protein component of the nucleolus—a structure that regulates genome 3D topology33,50 and further work is warranted to investigate these possibilities.

Our study has prospectively modeled the alterations in 3D genome organization that accompany the induction of hematological malignancy. The combined mutant state was able to alter gross genome organization at the level of genome compartments, although only 10% of the interactions, those associated with alterations of H3K4me1 (ref. 51), were rewired dynamically in leukemic HSPC. Most interactions remain stable, with the gene expression alterations presumably reflecting alterations of chromatin modification and accessibility at these loci. However, use of other cis-regulatory elements, such as those upregulating Socs2 and Irf8 (Figs. 3e and 6c), was new and specific for leukemia. These data demonstrate that malignant transformation is promiscuous in its enhancer use, as has also been suggested previously52,53; it can generate new DNA topologies and enhancer states but may also use enhancers and DNA topologies associated with other differentiation states in the tissue from which it is derived.

Our experimental framework has also allowed us to identify exemplar regulatory network nodes. Of note, when we compare these same factors with our own and other CRISPR screens performed in human AML cell line models54,55, we see that they constitute similar vulnerabilities across several AML genotypes. PU.1 is classically regarded as a tumor suppressor in AML. Mice with a genetically engineered decrease in PU.1 expression develop AML56,57,58,59. Furthermore, heterozygous deletions of the locus or rare loss-of-function mutations of PU.1 have been described in patients with AML60,61,62. Moreover, oncogenic events abrogate the activity of PU.1 (ref. 57,58) and, therapeutically, the anti-leukemic activity of LSD1 inhibitors has been associated with reactivation of a PU.1-driven transcriptional programme63,64. However, in contrast, our study identifies Spi1/PU.1 as an oncogenic TF if Npm1c/Flt3-ITD AML, and is in accord with our previous work and that of others that has demonstrated PU.1 to be a vulnerability in certain AML subtypes56,57,58,59. These findings suggest the role of PU.1 to be cell context-, genotype- and possibly stage-specific in AML, and warrant further study. Of note, the oncogenic effect of PU.1 seems to occur in combination with its known coactivators, AP-1 (ref. 42,43), IRFs and Hoxa9 (ref. 44,45), with this observation a further demonstration of the synergy between Npm1c and Flt3-ITD where coordinated upregulation of these synergistic TF occurs. Finally, our findings describe the entirely new HOXA-LRSE, which seems to regulate the entire HOXA cluster following mutation of Npm1. Importantly, we demonstrate that the HOXA-LRSE is not species-specific but is conserved in human cells. Further study of signal integration at the HOXA-LRSE will inform its critical roles for hematopoiesis and leukemogenesis51,56,57,58.

Taken together, our work demonstrates the complicated interplay that occurs between synergistic mutations to remodel the epigenetic landscape and rewire the epigenome to induce and maintain leukemogenic transcriptional programs, and identifies critical network characteristics that might be targeted for therapeutic success.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00925-9.

Received: 16 March 2020; Accepted: 28 July 2021; Published online: 23 September 2021

References
1. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet. 7, 29–59 (2006).
2. Lagha, M., Bothma, J. P. & Levine, M. Mechanisms of transcriptional precision in animal development. Trends Genet. 28, 409–416 (2012).
3. Spitze, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626 (2012).
4. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. 20, 437–455 (2019).
5. Laurenti, E. & Göttingen, B. From haematopoietic stem cells to complex differentiation landscapes. Nature 553, 418–426 (2018).
6. Huang, T. et al. Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. Dev. Cell 36, 9–23 (2016).
NATURE GENETICS

7. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132, 631–644 (2008).
8. Cedar, H. & Bergman, Y. Epigenetics of haematopoietic cell development. Nat. Rev. Immunol. 11, 478–488 (2011).
9. Cancer Genome Atlas Research, N. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. 368, 2059–2074 (2013).
10. Grimwade, D., Ivey, A. & Huntly, B. J. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood 127, 29–41 (2016).
11. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. 374, 2029–2032 (2016).
12. Dohner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute myeloid leukemia. N. Engl. J. Med. 373, 1136–1152 (2015).
13. Tenen, D. G. & Golomb, H. A. The RUNX family of transcription factors: regulators of normal and malignant hematopoiesis. Blood 115, 222–230 (2010).
14. Cowan, R. J. & Eisenman, R. N. Mammalian homeotic selector genes. Annu. Rev. Cell Biol. 9, 1–30 (1993).
15. Yanamandra, K. & Thiagarajan, R. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 15, 165–180 (2014).
16. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 17, 32–43 (2016).
17. de la Chapelle, A. & Thiagarajan, R. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 18, 26–37 (2017).
18. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 19, 21–32 (2018).
19. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 20, 33–44 (2019).
20. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 21, 45–56 (2020).
21. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 22, 57–68 (2021).
22. Wang, Y. et al. The role of histone 3 lysine 4 acetylation in regulating human leukemia. Nat. Rev. Genet. 23, 69–80 (2022).

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. 2021
Articles NATURE GENETICS

Methods
Ethical compliance. The studies were approved by the UK Medical Research Council and University of Cambridge Animal Welfare and Ethical Review Body (AWERB). All animal procedures were regulated under UK Home Office Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 under project license 80/2564. The work does not contain any experiments processing human primary samples performed by any of the authors.

Mice. C57/BL6 mice were housed and transferred to a temperature of approximately 21°C. Mice administered six doses of polyinosinic-polycytidylic acid (6×pIpC) or constitutive Flt3 ITD/ITD mice were maintained in a standard SPF facility (12 light/12 dark cycle, 19–23°C). Mice were killed when they exhibited signs of severe leukemia burden, including ruffled fur, hunchback, heavy shivering, weight loss, extreme weakness or evident loss of activity. No mice were allowed to exceed the limit of leukemia burden under the ethical regulations of AWERB.

Flow cytometry analysis. Flow cytometry measurements were conducted in seven to nine mice for each genotype. Briefly, bone marrow (BM) single-cell suspensions were isolated in PBS from mice femurs and tibias and stained with combinations of following anti-mouse antibodies for different populations: mouse antibodies included the cocktail containing CD3e, CD45, B220, Ly-6C/Ly-6G/Ly-6C, APC-conjugated; CD11b, 1.300 dilution, Ly-6A/E (Sca-1; Pacific Blue (PB)-conjugated); RioLegend; 1.500 dilution), CD117 (+-Kit); PE-Cy7-conjugated; BioLegend; 1.500 dilution), CD34 (FITC-conjugated; BD; 1.1000 dilution), CD16/32 (R;Fc III/II) (PE-conjugated; BioLegend; 1.500 dilution), CD135 (Flt3; PE-conjugated; BioLegend; 1.500 dilution). HSPC were determined using the combined immunophenotyping markers as below: Lin−, lineage negative; HPC, hematopoietic progenitors (containing all myeloid progenitors), Lin−Sca−1+Kit−; LSK, enriched for HSC, Lin−Sca−1+Kit−; CMP, common myeloid progenitors, Lin−Sca−1CKit+CD34−CD11b+; MEP, megakaryocyte/erythroid progenitors, Lin−Sca−1+Kit+CD34−CD16/32−; Assays were performed on a BD LSRFortessa cell analyzer and all data were analyzed with Floow software (Tree Star, v.10.0.5).

Gene expression analysis. RNA-seq reads were quality filtered and mapped using STAR (v.2.4.0) against the mouse genome (mm10). Uniformly mapped reads were quantified with HTSeq (v.0.6.0) and protein-coding genes with non-zero read count in WT or mutant HSPC (n = 16,771) were included for downstream analysis. Reads per kilobase million (RPKM) mapped reads for each protein-coding gene were calculated using Bioconductor package edgeR (v.3.2.8.1). We performed PCA on RPKM values for 16,771 protein-coding genes to assess reproducibility. Differential expression of protein-coding genes was analyzed with these counts using Bioconductor package DESeq (v.1.12.4). Significantly differential expression was considered by setting adjP < 0.05 and FC ≥ 1.5 between mutants and WT HSPC. Unsupervised clustering was used to generate the heat map.

ChIP-seq. We performed ChIP with an iDeal ChIP-seq kit for histones (Diagenode) following the manufacturer’s recommendations. In brief, 1 × 106 Lin− HSPC or NE isolated from WT or mutant mice (2–3 mice per genotype per sample replicate) were crosslinked with formaldehyde (Thermo Fisher Scientific) at a final concentration of 1% for 10 min and then quenched by addition of Glycine (Thermo Fisher Scientific) for 5 min incubation. Lysis buffers IL1 and IL2 were used to prepare the nuclei pellets. Chromatin was sheared in Shearing buffer iS1 (Thermo Fisher Scientific) for 5 min incubation. Lysis buffers iL1 and iL2 were used to quench the presence of any mutations. Accessible enhancers with gain or loss or H3K4me1 changes were defined by FDR value < 0.05 and FC ≥ 2 (gain or loss) in the presence of any mutations. Accessible enhancers with gain or loss of H3K4me1 amounts in any mutant versus WT HSPC were defined as dynamic enhancers; likewise, those with gain or loss of H3K27ac enrichment were considered as having dynamic activity.

Identification of leukemia-specific enhancer changes. Accessible enhancers with DM-induced significantly differential H3K4me1 were selected for plotting enhancer marks and chromatin accessibility in WT HSPC, DM HSPC, WT NE and GMP. Histone profiles and accessibility for GMP cells were previously reported13 and available in Gene Expression Omnibus (GEO) datasets (GSM222765, GSM2456862, GSM1441273; ATAC-seq, GSM1463173). Sequencing reads were mapped to mm10 and processed similarly as for WT HSPC. DM gained or lost enhancer groups were divided into two subgroups based on H3K4me1 enrichment patterns by k-means clustering (n = 2) using computeMatrix and plotHeatmap in deepTools (v.3.1.3). Enhancers showing unique changes in DM HSPC compared with WT HSPC and NE were considered ‘leukemia-specific’ changes.

Chromatin accessibility analysis. Chromatin accessibility probed by ATAC-seq was analyzed in a manner similar to H3K27ac ChIP-seq analysis, including the procedures of trimming, mapping, filtering and peak calling. Briefly, trimmed sequences were mapped against mm10 reference genome using Bowtie2, and only uniquely mapped reads were kept. Peaks were called using MACS2 with the setting ‘-nomodel -nolambda’ and only those with Pvalues less than 1 × 10−6 were considered significant. A list of ATAC-seq consensus peak set was made using DiffBind. Differential enrichment was analyzed using edgeR, with significant changes being defined by FDR value < 0.05 and FC ≥ 2 (gain or loss) in the presence of any mutations.

Motif enrichment and footprinting analyses. We scanned for de novo motifs at ±100 bp from ATAC-seq peak summit in the investigated chromatin regions, using motifGenome.pl with ‘setting’ ‘mask’ and ‘permute’ options for the MUMMER package. In most cases, motifs occurred with significant P values and a coverage of > 10% of target regions were displayed. Combined analysis of the occurrence of TFs and chromatin accessibility at regions of interest was assessed by BagFoot (v.0.9.7) algorithm following the procedures as described23. We performed analysis on ATAC-seq profiles at all accessible regions represented by ATAC-seq consensus peaks. In brief, known TF motifs were first scanned at ±100 bp from the ATAC-seq peak summit and their occurrence was aggregated. Next, for each motif, ATAC-seq reads were counted at aggregated motif-flanking regions (±200 bp from motif center) and normalized to total sequencing counts in each sample. Differential accessibility was then determined as altered read counts flanking each motif compared to WT. To probe TF footprinting depth, expected cuts were computed by mapping ATAC-seq reads to each motif occurred genome-wide and was set as baseline, and observed cuts only counted the reads within ±100 bp from ATAC-seq peak summit. Footprinting depth was calculated as cut bias represented by a log ratio difference of observed cuts divided by expected cuts. Together, for each comparison between mutant and WT HSPC, the differences of footprinting depth and flanking accessibility for each TF motif were plotted in a Bagplot, where the inner polygon (‘bag’) encompasses at most half of the TFs and the outer polygon (‘fence’) is formed by inflating the bag geometrically by a default factor of 2.5. The outliers with a P value < 0.05 were considered significant.

Promoter-anchored interaction analyses. Paired-end sequences of pCHC were processed using the HiCUP (v.0.5.8) pipeline with default parameters for the following steps: quality control, identification of reads containing HJC junctions, mapping to reference genome mm10 and filtering duplicated HiC di-tags. Output bam files containing valid HiC di-tags were processed by Bioconductor package BaGFoot (v.1.1.1)23, to call significant promoter-based interactions. ChiCAGO
considers distance effect on interaction frequencies by virtue of a convolution background model and a distance-weighted Pvalue. CHiCAGO scores represent –log weighted P values; the higher the Pvalue, the more likelihood of interaction formed. CHiCAGO scores were calculated for each pCHiC sample, and significant interactions were called when CHiCAGO scores were ≥ 2. Significant interactions were computed per HSPC by merging their replicates, and were combined to form a matrix of total unique interactions, whereas interactions present in both replicates per HSPC were considered with high confidence. Total pCHiC reads at individual promoters were summed to perform differential analysis in mutant versus WT HSPC using edgeR. Differential total interaction reads were defined by adjP < 0.05 and absolute FC > 1. Rewired interactions were identified by comparing and ranking their CHiCAGO scores in WT and mutant cells. Those high-confidence interactions, absent in WT (score ≤ 5) but present in mutant (score ≥ 2), with scores ranked in the bottom quartile in WT but in the top quartile in mutant, were considered as lost interactions by mutations. In addition, total unique interaction profiles facilitated the annotation of distal cis-regulatory elements (for example SEs) to their target genes in HSPC.

Chromatin compartment analysis. Subnuclear compartmentation represented by self-associating chromatin domains were analyzed by means of PCA on capture HiC data in a similar way as described in HiC data2, mainly using HOMER (available at https://homer.ucsd.edu/homer/) and being uniquely mapped from HiCUPS2.2.5 were used to create ‘tag Directory’ using maketagDirectory from HOMER. Principal component (PC) values were calculated by runningHiCpca.pl with default setting (resolution at 50kb). This led to the separation of chromatin into two compartments, with positive PC1 regions reflecting ‘active’ chromatin and negative PC1 regions indicative of ‘inactive’ chromatin. Regions of continuous positive or negative PC1 value were stitched to be identified as A or B compartments respectively. We performed genome-wide correlation of compartment PC1 values between mutant and WT cells by running getHiCcorrDiff.pl from HOMER. Flipped compartments were identified using HOMER findHCompartment.pl. Genes involved in the flipped compartments were selected to analyze their mRNA expression changes by mutations.

Seurat-guided clustering analysis to identify differential patterns of accessible enhancers. A multilayered multiomic approach to integrate chromatin analysis31 between mutant and WT cells by running getHiCcorrDiff.pl from HOMER. Flipped compartments were identified using HOMER findHCompartment.pl. Genes involved in the flipped compartments were selected to analyze their mRNA expression changes by mutations.

Gene perturbation through shRNA-mediated knockdown. NPM1c/FLT3-ITD leukemia cells (termed DM cells) were derived by isolation of Lin− BM HSPC from double mutant mice and were cultured in X-Vivo medium (Lonza) plus 10% fetal bovine serum (Gibco) in the presence of 10 ng ml−1 IL-3, 10 ng ml−1 IL-6 and 50 μM 2-β-mercaptoethanol (Gibco). Lentiviral shRNA expression was achieved by electroporation with Lentiviral Packaging Kit (Addgene catalog no. 67978)52. Production of shRNA particles and transduction of DM-Cas9 cells were carried out in the same way as for shRNA. To confirm the target deletion in the bulk transduced cells, PCR primers were subjected to TA cloning (Promega). Plasmid DNA was extracted from individual colonies and deletions were confirmed by Sanger sequencing.

CRISPR-mediated gene loss of function in human leukemia cells. OCI-AML3 and MOLM-13 cells were obtained from the Sanger Institute Cancer Cell Line Panel, with constitutive expression of Cas9 generated by lentiviral transduction using pkLV2-ESiIb22AcA9-C9 vector (Addgene catalog no. 67978)11. OCI-AML3 and MOLM-13 cells were maintained mycoplasma-free in MEM-alpha supplemented with 20% FCS and RPMI supplemented with 10% FCS, respectively. For each of the human targets (including IRF8, SPI1 and IGF1), two gRNAs were designed and subcloned into the pkLV2-U6gRNA5(RbsI)-PKPpuro2ABFP-W expression vector (Addgene catalog no. 67974)42. A nontarget control gRNA was also included. Target sequences of all gRNAs are detailed in Table 2. For cell proliferation assays, 50,000 cells transduced with specific gRNAs were plated and cells were counted every 2–3 days for a period of 7–14 days. Three replicates for each target or control were carried out. Data were presented as a fold change (FC) in total cell numbers relative to corresponding nontargeting controls. For colony-forming assays, 500–1000 cells were transduced with specific gRNAs were seeded in duplicate in Methocult semi-solid medium (H4351, STEMCELL Technologies). Total colony-forming units (CFU) were enumerated 7 days postculture. Three replicates for each target or control were carried out. Data were presented as a FC in CFU or colony cellularity, relative to nontargeting controls.

Statistics and reproducibility. Statistical analyses in the studies were specified in details in figure legends. Sample sizes were not predetermined using any statistical method. Randomization was applied to mouse sample preparation and cell perturbation experiments. The investigators who collected the mouse samples were not informed about the sample allocation for NGS experiments. The investigators who performed other experiments were not blinded to allocation during experiments and outcome assessment. No data were excluded for analysis, except for the exclusion of reads mapped to mitochondrial DNA, as the overall study aim was to investigate only the chromatin landscape of the nonmitochondrial genome. Student’s unpaired or paired t-tests were performed as two-tailed with GraphPad Prism (v.8.2.1). One-tailed hypergeometric distribution was analyzed in R (v.3.5.1). P values in HOMER de novo motif analysis were calculated as one-tailed or in BatFoot analysis as two-tailed. These analyses were not corrected for multiple testing. Statistical calculations in all differential analyses of sequencing data were performed with DESeq2 or edgeR, generating two-tailed and multiple testing corrected P (with the Benjamini and Hochberg method, adjP) or FDR values. Output data with significance P, adjP or FDR values < 0.05 were considered statistically significant. Number of independent experiments or independent samples were specified in figure legends. Representative data or images were replicated in at least three independent experiments or three independent samples.

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

Data availability All sequencing raw data, normalized bigwig tracks for RNA-seq, ChIP-seq and ATAC-seq have been deposited in the GEO database under the series GSE146669 (subseries GSE146668 for RNA-seq, GSE146663 for ChIP-seq, GSE146613 for ATAC-seq and GSE146662 for pCHiC) and with no restrictions to access. All supporting data derived from the sequencing analysis to assist understanding of the results and discussions in the paper were provided in several supplementary tables. The studies have also reanalyzed multiple datasets that are publicly available: the ChIP-seq on BRD4 (ArrayExpress: ERR220396) and H3K27ac (GSM271671) in OCI-AML3 cells; H3K27ac in Kasumi-1 cells (GSM221083); H3K4me1 (GSM2016688) and CTCF (GSM149141) in human CD34+ HSPC (DHS-seq) in AMI patients (GSM2893610, GSM2893614, GSM2893615 and GSM2893616); the ATAC-seq in human CD34+ HSPC (GSM1888553); the pCHiC from human CD34+ cells (ArrayExpress: ERR343027); H3K4me1 (GSM1442890); H3K27ac (GSM1441273) and ATAC-seq (GSM1463173) in mouse GMP cells. Source data are provided with this paper.

Code availability All computational analysis is described in Methods with either the software default parameters and pipelines or custom code available at https://github.com/haiyang-yun/3D_chromatin_in_AMI (archived also on Zenodo: https://doi.org/10.5281/zenodo.5009067).
References
68. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
69. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
70. Conesa, A. et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13 (2016).
71. 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).
72. Bailey, T. et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput. Biol. 9, e1003326 (2013).
73. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
74. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393 (2012).
75. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016).
76. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. PLoS Comput. Biol. 4, 1310 (2015).
77. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 17, 127 (2016).
78. Yun, H. haiyang-yun/3D_chromatin_in_AML. Zenodo https://doi.org/10.5281/zenodo.5009065 (2021).

Acknowledgements
This study was carried out in the laboratory of B.J.P.H. with funding from Cancer Research UK (C18680/A25508), the European Research Council (647685), MRC (MR-R099708-1), the Kay Kendall Leukaemia Fund (KKL1243), the Wellcome Trust (205254/Z/16/Z) and the Cancer Research UK Cambridge Major Centre (C49940/A25117). This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014), and was funded in part by the Wellcome Trust, who supported the Wellcome – MRC Cambridge Stem Cell Institute for the help with preparation of RNA-seq libraries and the Cancer Research UK (CRUK) Cambridge Institute Genomics Core for providing the NGS services.

Author contributions
H.Y. and B.J.P.H. conceived the study, designed the experiments and prepared the manuscript. H.Y. designed and conducted most of the experiments and analyzed the data. N.N. performed the CRISPR experiments in human leukemic cell lines. S.V. executed the NGS data alignment and helped with bioinformatic analysis. G.G. performed the flow cytometry experiments. A.M. and G.S.V. coordinated mouse tissues. P.M. performed the BaGFoot analysis. D.S., D.L.-A., S.J.H., S.A.-S., F.B., L.M., A.C.-V. and P.G. provided technical assistance. E.M. and X.W. helped with computational analysis. M.G. and O.M.D. provided the DM-Cas9 cells. C.M.-T. participated in the discussion of the project. C.S.O. provided expertise on promoter capture HiC assays. B.J.P.H supervised all the work and directed the project.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00925-9.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00925-9.
Correspondence and requests for materials should be addressed to Brian J. P. Huntly.
Peer review information Nature Genetics thanks Charles Mullighan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Transcriptional changes across WT and mutant HSPC. a, Heatmap showing unsupervised clustering analysis of global gene expression in WT and mutant HSPC. b, Top 10 positively enriched gene ontology (GO) gene sets revealed by GSEA for each mutant (vs WT). c, Commonly enriched gene sets from GSEA analysis of differential gene expression by Npm1c and DM. d, GSEA enrichment plots showing gene set of Kong_E2F3_Targets in Npm1c or DM vs WT by GSEA analysis. e, Gene sets positively or negatively enriched in all three mutants vs WT.
Extended Data Fig. 2 | See next page for caption.
**Extended Data Fig. 2 | Global chromatin accessibility across WT and mutant HSPC.**

**a.** Heatmaps and profile plots of ATAC-seq enrichment across WT and mutant HSPC over regions with gain or loss of accessibility in the presence of Npm1c. Peaks were ranked by average enrichment across all samples.

**b.** Chromatin accessibility at the Gata2 genes and its upstream enhancers in all four HSPC and wildtype neutrophils. Regions showing loss of accessibility in Npm1c and DM HSPC are highlighted.

**c.** Linking differential accessibility at gene promoters to their mRNA expression changes by each mutant. Up- or downregulation were defined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and absolute FC ≥ 1.5.

**d-f.** De novo motifs significantly enriched at genomic regions with altered accessibility by Npm1c (d) and Flt3-ITD (e) or open chromatin sites with static accessibility (f) where FDR of differential analysis > 0.2 in any mutant vs WT HSPC. HOMER outputs motifs with target coverage > 10% and ranked by p values (one-tailed, not multiple testing corrected).

**g and h.** BaGFoot analysis illustrates TFs with differential footprint depth and accessibility in Npm1c (g) and Flt3-ITD (h) vs WT HSPC. Motifs outside the fence and with a p value (two-tailed, not multiple testing corrected) < 0.05 are statistically significant outliers; n.s., not significant.
Extended Data Fig. 3 | Chromatin modifications at accessible enhancers across WT and mutant HSPC. 

a, Distribution of all H3K4me3 peaks based on read counts across all four cell types. Active promoters were marked by high H3K4me3 with log2 CPM > 4. 
b, Genome distribution of ATAC-seq consensus peaks. 
c, Number of total accessible enhancer peaks defined in all four HSPC samples and percentage of dynamic enhancers which were defined by differential H3K4me1 in the presence of any mutations. 
d, Numbers of overlapping accessible enhancers with gain or loss of H3K4me1 in the presence of Flt3-ITD or DM. Hypergeometric test p values (one-tailed, not multiple testing corrected) are shown. 
e, Heatmaps of H3K4me1 enrichment over gained or lost enhancers during DM leukemia induction. Peaks were ranked by average enrichment across all samples. 
f, Percentage of accessible enhancers marked by both H3K4me1 and H3K27ac in total accessible enhancers across all four HSPC states. 
g, Percentage of enhancers showing dynamic H3K27ac modification in the presence of single or double mutations. 
h, Heatmaps and profile plots of H3K27ac enrichment over enhancers showing gain or loss of H3K27ac during DM leukemia induction. Peaks were ranked by average enrichment across all samples.
Extended Data Fig. 4 | Immunophenotypic characterization of HSPCs isolated from bone marrow of wildtype and mutant mice. 

**a**, Representative flow cytometry plots showing proportions of hematopoietic stem and progenitors within bone marrow compartment. Lin−, lineage negative; HPC, hematopoietic progenitors (Lin−Sca1−cKit+) containing all myeloid progenitors; LSK, Lin−Sca1+cKit+; CMP, common myeloid progenitors; GMP, granulocyte/macrophage progenitors; MEP, megakaryocyte/erythroid progenitors.  

**b**, Comparison of proportions of hematopoietic stem and progenitors in their parental populations in mutant mice (Npm1^flx-cA/+;Mx1-Cre, n = 9; Flt3^T/D/+; n = 8; Npm1^flx-cA/+;Mx1-Cre;Flt3^T/D/+, n = 9) compared with wildtype mice (WT, n = 7). Student’s unpaired t-tests (two-sided) were performed for the comparisons; data are presented as mean values ± standard deviation; Only the p values < 0.05 were shown.
Extended Data Fig. 5 | 3D chromatin interaction profiles across WT and mutant HSPC. **a**, Numbers of total and high-confidence pCHiC interactions in all four HSPC states. High-confidence interactions were defined as significant interactions (with CHiCAGO score ≥ 5) overlapping in both replicates of each cell type. **b**, Numbers of pCHiC interactions captured by individual promoters. **c**, Distance of interacting regions from their target promoters. Median distance was shown. **d**, Illustration of chromatin compartments A/B levels at a DM-induced ‘B to A’ flipped region (highlighted) containing the Igf1 oncogene. **e**, Illustration of chromatin compartments A/B levels at a DM-induced ‘B to A’ flipped region (highlighted) containing an Ifi (interferon-inducible) gene cluster (in red). **f** and **g**, Significant interactions, represented by arcs, associated with the Gfi1b (f) or Irf8 (g) promoters in different cell types.
Extended Data Fig. 6 | Integrative analysis with Seurat-guided clustering of chromatin profiles across WT and mutant HSPC. 

**a**, Profile plots of chromatin marks and accessibility at accessible regions in each of the 10 clusters (except for Cluster-6 and -10). 

**b**, De novo motifs significantly enriched at Clusters 7–9 chromatin regions. HOMER outputs motifs with target coverage >10% and ranked by p values (one-tailed, not multiple testing corrected). 

**c**, GO analysis of genes that are linked to the chromatin regions in Cluster-10 using pCHiC profiles and downregulated during leukemia induction. 

**d**, Volcano plot showing differential expression of Cluster-10 linked genes in DM vs WT HSPC. Up- or downregulation were defined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and absolute FC $\geq$ 1.5. 

**e**, Percentage of Cluster-6 genes with human overlap showing their expression in DM vs WT HSPC. Differential expression was defined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and absolute FC $\geq$ 1.5.
Extended Data Fig. 7 | Correlating enhancer alterations with differential expression of target genes or DNA topology changes. a, Percentage of dynamic enhancers interacting with gene promoters and the percent of up- or downregulation of the genes they linked to. Dynamic enhancers were marked by either differential H3K4me1 only, or together with differential H3K27ac and accessibility. b, Percentage of genes involving flipped chromatin compartments and expressed in DM vs WT HSPC. Differential expression was defined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and absolute FC ≥ 1.5. c, Percentage of enhancers with dynamic interactions showing their H3K4me1 read counts in DM vs WT HSPC. Differential H3K4me1 levels were defined by setting FDR (two-tailed and multiple testing corrected) < 0.05 and FC ≥ 1.5. d, Example genomic region demonstrating correlation of H3K4me1 changes with ‘rewired’ interactions.
Extended Data Fig. 8  |  See next page for caption.
Extended Data Fig. 8 | Differential Super-enhancer usage across WT and mutant HSPC. a, Definition of super-enhancers (SEs) by ranking H3K27ac peaks that were overlapped in both replicates of each cell type based on normalised H3K27 counts. The top-ranked 801 regions were considered as SEs across four HSPC. b, Number of super-enhancers (SE) with increase (UP) or decrease (DOWN) in H3K27ac modification in mutant vs WT HSPC. c, Heatmaps and profile plots of H3K27ac enrichment in WT and mutant HSPC over SEs showing gain or loss of H3K27ac during DM leukemia induction. d, Linking differential activity of SEs to altered mRNA expression of their target genes (determined by pChIC interactions) in mutant vs WT HSPC. Upregulated genes connecting to SEs with H3K27ac gain or downregulated genes connecting to SEs with H3K27ac loss were determined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and FC ≥ 1.5 for expression, or FDR (two-tailed and multiple testing corrected) < 0.05 and FC ≥ 1.5 for H3K27ac. Several Hoxa genes were indicated in Npm1c vs WT. e, Significantly enriched GO terms for genes linked to gained SEs and upregulated in DM vs WT (upper panel) or for genes linked to lost SEs and downregulated (lower panel). f and g, Promoter-contact plots showing the read counts of promoter bait to target pairs (bait-‘other end’) for Spi1 (f) and Hoxa9/Hoxa10 (g). Regions highlighted yellow show rewired interactions.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Perturbation of critical cis-regulatory hubs and their target genes abrogates leukemia maintenance in mouse DM leukemia.

a, mRNA expression of AP-1 components (Jun, Fos) detected by RT-qPCR in DM cells expressing shRNAs targeting Jun or Fos relative to control shRNA (n = 3 independent experiments).

b, CFU assays of DM cells expressing shRNAs targeting Jun or Fos or control gRNAs (n = 3 independent experiments).

c, mRNA expression of Hoxa9, Hoxa10, and Spi1 in DM cells expressing shRNAs targeting them specifically relative to control shRNA (n = 3 independent experiments).

d and e, CFU assays (d) and ex vivo cell proliferation (e) of DM cells expressing shRNAs targeting Hoxa9, Hoxa10, Spi1 or control shRNA (n = 3 independent experiments).

f and i, Gel electrophoresis of PCR products on genomic DNA of DM-Cas9 cells expressing control gRNAs, Spi1-URE (f) or Hoxa-LRSE (i) gRNAs. PCR experiment was performed once on three independent gRNA transductions (marked with #).

g and j, Sanger sequencing to confirm the deletion of Spi1-URE (g) or Hoxa-LRSE (j) in DM-Cas9 cells expressing Spi1-URE gRNAs.

h and k, Ex vivo cell growth of DM-Cas9 cells expressing Spi1-URE gRNAs (h) or Hoxa-LRSE gRNAs (k) in comparison to cells expressing control gRNAs (n = 3 independent experiments). Statistical analyses in a-d, h and k were performed by running Student’s unpaired t-tests and p values (two-sided, not multiple testing corrected) were shown; error bars represent mean and standard error of mean.
Extended Data Fig. 10 | Perturbation of critical target genes impairs human leukemia cells. a and b, Relative mRNA expression of target gene expression upon CRISPR-mediated knockdown in OCI-AML3 (a) and MOLM-13 (b) leukemia cell lines (n = 3 independent experiments). c–e, Cell proliferation of MOLM-13 cells expressing human gRNAs targeting SPI1 (c), IRF8 (d) and IGF1 (e) relative to control gRNAs (n = 3, 4, 3 independent experiments, respectively). # indicates two independent gRNAs. f, CFU assays of MOLM-13 cells expressing human gRNAs targeting SPI1, IRF8 and IGF1 relative to control gRNAs (n = 3 independent experiments). Statistical analyses in a–f were performed by running Student’s unpaired t-tests and p values (two-sided, not multiple testing corrected) were shown; error bars represent mean and standard error of mean.
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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
- NGS sequencing reads were collected from Illumina HiSeq2500 or HiSeq4000.

Data analysis
- Bowtie (version 2.1.0), STAR (version 2.4.0), MACS2 (version 2.0.1), DiffBind (version 2.0.1), Homer (version 4.10.4), HTSeq (version 0.6.0), PICARD tools (version 2.2.1), R (version 3.6.1), edgeR (version 3.28.1), DESeq2 (version 1.12.4), GSEA (version 3.0), ShinyGO (version 0.61), deepTools (version 3.2.3), CHICAGO (version 1.1.1), HiCUP (version 0.5.8), Graphpad Prism (version 8.2.1), WashU Epigenome Browser (version 46.2), Seurat (version 3.2.3), bagFoot (version 0.9.7), FlowJo (version 10.5.0)

- Peak calling using findPeaks (Homer), parameters "size 1000 -minDist 1000 -tbp 3"
- Peak calling using MACS2, parameters "callpeak -p 1e-20 -nomodel"
- pChIC data processing using HiCUP, parameters "--longest 800 --shortest 150"
- Chromatin compartment analysis using runHiCpca.pl and getHiCcorrDiff.pl (Homer), parameters "-res 50000 -genome mm10"
- Significant interaction were identified by running runChicago.R (CHICAGO), parameters "--cutoff 5"
- Motif analysis using findMotifsGenome.pl (Homer), parameters "-mask-p 12 -s 10"
- BagFoot analysis, default parameters plus "factor = 2.5" for "gen_bagplot_chisq" function, mouse genome mm10

For manuscripts utilizing custom algorithms or software that are not yet described in published literature, software must be made available to editors/reviewers. 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 sequencing raw data, normalised bigwig tracks for RNA-seq, ChIP-seq and ATAC-seq have been deposited in the GEO database under the series GSE146669 (subseries GSE146668 for RNA-seq, GSE146663 for ChIP-seq, GSE146613 for ATAC-seq, and GSE146662 for pCHIC) and with no restrictions to access. All supporting data derived from the sequencing analysis to assist understanding of the results and discussions in the paper were provided in multiple supplementary tables. The studies have also been re-analysed multiple data sets which are publicly available: the ChIP-seq on 8BlD4 (ArrayExpress: ERR220396) and H3K7ac (GSM2716711) in OCI-AML3 cells; H3K27ac in Kasumi-1 cells (GSM2212053); H3K4me1 (GSM1815068) and CTCF (GSM651541) in human CD34+ HSPC; the DHS-seq in AML patients (GSM2893610, GSM2893614, GSM2893615 and GSM2893616); the ATAC-seq in human CD34+ HSPC (GSM1888536); the pCHIC from human CD34+ cells (ArrayExpress: ERR436027); H3K4me1 (GSM14412890), H3K27ac (GEO: GSM1441273) and ATAC-seq [GSM1463173] in mouse GMP cells.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-list.pdf

Life sciences study design

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

Sample size

Sample sizes were not predetermined using any statistical method. To perform NGS-based experiments, variable number of cells were required and therefore starting materials were collected by pooling cells isolated from multiple mice. For ex vivo perturbations assays, samples were engineered and then expanded for further analysis.

Data exclusions

No data were excluded for analysis, except for the exclusion of reads mapped to mitochondrial DNA, as the overall study aimed to investigate only the chromatin landscape of non-mitochondrial genome.

Replication

NGS-based experiments were performed on pooled cells isolated from individual mice to serve as biological replicates, and were carried out independently and sequenced separately. Two independent biological samples were sequenced for all the conditions. Ex vivo perturbation experiments were carried out on independently transduced cells and were analysed separately as biological replicates (at least three).

Randomization

Mice were randomly grouped and sacrificed at specific ages as described in the methods. Cells for sequencing studies were isolated from individual mice and pooled as randomized starting materials. Cells for flow cytometry analysis were harvested from individual mice in a randomized order.

Blinding

The investigators who collected the mouse samples were not informed about the sample allocation for NGS experiments. The investigators who performed other experiments were not blinded to allocation during experiments and outcome assessment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a

- ☒ Antibodies
- ☒ Eukaryotic cell lines
- ☒ Palaeontology
- ☒ Animals and other organisms
- ☒ Human research participants
- ☒ Clinical data

Methods

n/a

- ☒ ChiP-seq
- ☒ Flow cytometry
- ☒ MRI-based neuroimaging

Antibodies

Antibodies used

For ChiP-seq: anti-H3K4me1, #ab8895, Abcam; anti-H3K27ac, #ab4729, Abcam; anti-H3K4me3, #ab15410D03, Diagenode. 1 μg
antibody was used per ChIP-seq sample (1x10⁶ cells).

For flow cytometry: Mouse Lineage antibody cocktail (containing CD3e/CD11b/B220/Ly-76/Ly-6G/Ly-6C; APC conjugated; BD; 1:300 dilution), Ly-6A/E (Sca-1; Pacific Blue (PB) conjugated; BioLegend; 1:500 dilution), CD117 (c-Kit; PE-Cy7 conjugated; BioLegend; 1:500 dilution), CD34 (FITC conjugated; BD; 1:1000 dilution), CD16/32 (FcR III/II) (PE conjugated; BioLegend; 1:500 dilution), CD135 (Flt3; PE conjugated; BioLegend; 1:500 dilution).

Validation
Antibodies used for ChIP-seq in this study were suggested by ENCODE project or ChIP-seq experiments published previously. http://genome.ucsc.edu/ENCODE/validation/antibodies/mouse_h3K4me1_ab8895_validation_Ren.pdf
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2427862
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3020545

Antibodies used for flow cytometry were previously described (Girotopoulos G. et al. J Exp Med 2015).

Eukaryotic cell lines

Policy Information about cell lines

Cell line source(s) OCI-AML3 and MOLM-13 cell lines were obtained from the Sanger Institute Cancer Cell Collection.

Authentication
Cell lines were authenticated by morphology under microscopy and PCR and sanger sequencing for STR.

Mycoplasma contamination Cell lines were regularly tested and were negative for mycoplasma infection.

Commonly misidentified lines (See ITAC register) No the cell lines were not commonly misidentified lines.

Animals and other organisms

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

Laboratory animals All mice used in this study were C57/B16 strain, no gender selection. Wildtype mice, or mice carrying either of the single mutations were sacrificed at 10-12 weeks of age as described in the methods. Mice carrying the combined mutations were sacrificed at 5-6 weeks of age. All mice were maintained in a standard SPF facility (12 light/12 dark cycle, 19-23°C with 40-60% humidity)

Wild animals The study did not use wild animals.

Field-collected samples The study did not use any samples collected from the field.

Ethics oversight Our animal procedures were regulated under UK Home Office Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 under project license 80/2564 and ethics were approved by Wellcome Sanger Institute’s Animal Welfare and Ethics Review Body.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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
For reviewer's access to our GEO data, please use the link and token below:
To review GEO accession GSE146669:
Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146669
Enter token wzbikenezavyp into the box

Files in database submission
GSM4399714 WT.HSFC_ATAC-seq_rep1
GSM4399715 WT.HSFC_ATAC-seq_rep2
GSM4399716 Npm1c.HSFC_ATAC-seq_rep1
GSM4399717 Npm1c.HSFC_ATAC-seq_rep2
GSM4399718 Flt3_ITD.HSFC_ATAC-seq_rep1
GSM4399719 Flt3_ITD.HSFC_ATAC-seq_rep2
GSM4399720 DM.HSFC_ATAC-seq_rep1
GSM4399721 DM.HSFC_ATAC-seq_rep2
GSM4399722 WT.Neu_ATAC-seq_rep1
GSM4399723 WT.Neu_ATAC-seq_rep2
GSM4403849 WT.HSFC_pCHIC_rep1
GSM4403850 WT.HSFC_pCHIC_rep2
GSM4403851 Npm1c.HSFC_pCHIC_rep1
GSM4403852 Npm1c.HSFC_pCHIC_rep2
Methodology

Repeats

Experiments were performed independently on two biological replicates. Pearson correlation between biological replicates were > 0.9.

Sequencing depth

ChIP-seq data were produced as 50bp single-read sequences.

Samples:

Total uniquely mapped reads, Reads after removing duplicates

DM.HSPC.H3K27ac: R1,30267545,26377494
DM.HSPC.H3K27ac: R2,25939483,23768443
DM.HSPC.H3K4me1: R1,25750197,24853829
DM.HSPC.H3K4me1: R2,3465066,32791524
DM.HSPC.H3K4me3: R1,36224393,23547132
DM.HSPC.H3K4me3: R2,47884119,24859769
DM.HSPC.Input: 26090927,25579171
Flt3-TD: HSPC.H3K27ac: R1,32116367,24754887
Flt3-TD: HSPC.H3K27ac: R2,26050925,23911848
Flt3-TD: HSPC.H3K4me1: R1,42100218,38378965
Flt3-TD: HSPC.H3K4me1: R2,28889432,27427100
Flt3-TD: HSPC.H3K4me3: R1,58183394,18217105
Flt3-TD: HSPC.H3K4me3: R2,68221416,26679303
Antibodies

anti-H3K4me1, rabbit, Abcam; anti-H3K27ac, #ab4729, Abcam; anti-H3K4me3, #c15410003, Diagenode

Peak calling parameters

For H3K4me1 and H3K4me3 (from narrow to broad), peak calling using findPeaks (Homer), parameters "-size 1000 -minDist 1000 -bop 3";
For H3K27ac [narrow], peak calling using MACS2, parameters "callpeak -p 1e-9 --nomodel"

Data quality

Peaks were called for enrichment against input control of each cellular states. Cutoffs were set at FDR < 0.001 (Homer) or P-value < 1e-9.

Sample,Number of_peaks
DM.HSPC.H3K27ac_rep1,27376
DM.HSPC.H3K27ac_rep2,21910
DM.HSPC.H3K4me1_rep1,104761
DM.HSPC.H3K4me1_rep2,109371
DM.HSPC.H3K4me3_rep1,31848
DM.HSPC.H3K4me3_rep2,37050
Flt3-TD.HSPC.H3K27ac_rep1,123564
Flt3-TD.HSPC.H3K27ac_rep2,24807
Flt3-TD.HSPC.H3K4me1_rep1,128798
Flt3-TD.HSPC.H3K4me1_rep2,116150
Flt3-TD.HSPC.H3K4me3_rep1,33021
Flt3-TD.HSPC.H3K4me3_rep2,39320
Npm1c.HSPC.H3K27ac_rep1,29052
Npm1c.HSPC.H3K27ac_rep2,30150
Npm1c.HSPC.H3K4me1_rep1,102927
Npm1c.HSPC.H3K4me1_rep2,125709
Npm1c.HSPC.H3K4me3_rep1,38562
Npm1c.HSPC.H3K4me3_rep2,32463
WT.HSPC.H3K27ac_rep1,22065
WT.HSPC.H3K27ac_rep2,21709
WT.HSPC.H3K4me1_rep1,122876
WT.HSPC.H3K4me1_rep2,131205
WT.HSPC.H3K4me3_rep1,30310
WT.HSPC.H3K4me3_rep2,34333
WT.Neu_H3K27ac_rep1,25554
WT.Neu_H3K27ac_rep2,26437
WT.Neu_H3K4me1_rep1,11029
WT.Neu_H3K4me1_rep2,110503
WT.Neu_H3K4me3_rep1,34146
WT.Neu_H3K4me3_rep2,32298

Software

MACS(version 2.0.1), Homer (version 4.10.4), deepTools (version 2.3.3)
Flow Cytometry

Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

| Sample preparation | Bone marrow (BM) single-cell suspensions were isolated in PBS from mice femurs and tibias and stained with combinations of the following anti-mouse antibodies for different populations. |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Instrument         | Assays were performed on a BD LSRII Fortessa cell analyzer.                                                                                                                                              |
| Software           | All data were analyzed with FlowJo software (Tree Star).                                                                                                                                                  |
| Cell population abundance | No cell sorting procedure was carried out.                                                                                                                                                                |
| Gating strategy    | Cells were gated according to FSC and SSC in order to discard cell debris and cell clumps. Fluorescent cells were gated with a threshold capturing the 0.5% upper tail of a negative control population.                  |

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.