Published in final edited form as:
Nat Genet. 2021 October 01; 53(10): 1443–1455. doi:10.1038/s41588-021-00925-9.

Mutational synergy during leukemia induction remodels chromatin accessibility, modification and 3-Dimensional DNA topology to alter gene expression

Haiyang Yun[1,2,3], **Nisha Narayan**[1,2], **Shabana Vohra**[1,2], **George Giotopoulos**[1,2], **Annalisa Mupo**[1,2,4,8], **Pedro Madrigal**[1,2], **Daniel Sasca**[1,2,5], **David Lara-Astiaso**[1,2], **Sarah J. Horton**[1,2], **Shuchi Agrawal-Singh**[1,2], **Eshwar Meduri**[1,2], **Faisal Basheer**[1,2], **Ludovica Marando**[1,2], **Malgorzata Gozdecka**[1,2,4], **Oliver M. Dovey**[1,2,4], **Aracely Castillo-Venzor**[1], **Xiaonan Wang**[1,2], **Paolo Gallipoli**[1,2,6], **Carsten Müller-Tidow**[3], **Cameron S. Osborne**[7], **George S. Vassiliou**[1,2,4], **Brian J. P. Huntly**[1,2,*]

[1]Wellcome - MRC Cambridge Stem Cell Institute, Cambridge, United Kingdom

[2]Department of Haematology, University of Cambridge, Cambridge, United Kingdom

[3]Department of Medicine V, Hematology, Oncology and Rheumatology, University of Heidelberg, Heidelberg, Germany

[4]Haematological Cancer Genetics, Wellcome Sanger Institute, Cambridge, United Kingdom

[5]Department of Hematology, Oncology and Pneumology, University Medical Center Mainz, Mainz, Germany

[6]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

[7]Department of Medical and Molecular Genetics, King's College London, London, United Kingdom

Abstract

Altered transcription is a cardinal feature of acute myeloid leukemia (AML), however, exactly how mutations synergize to remodel the epigenetic landscape and rewire 3-Dimensional (3D) DNA topology is unknown. Here we apply an integrated genomic approach to a murine allelic series that

*Corresponding author bjph2@cam.ac.uk.
[8]Present address: Epigenetics Programme, The Babraham Institute, Cambridge, United Kingdom

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 analysed 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.

Europe PMC Funders Author Manuscripts

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 *Npm1c* mutation to alter gene expression and drive leukemia induction. These analyses also allow the identification of long-range *cis*-regulatory circuits, including a novel super-enhancer 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.

Introduction

The functional specification of tissues in metazoans occurs through the generation of cell-specific proteomes, in turn, controlled by diverse transcriptional programs[1,2]. These transcriptional programs are tightly spatiotemporally regulated through the utilization of tissue-specific *cis*-regulatory elements such as enhancers, that are licensed and brought into direct communication with their cognate promoters through the functions of transcription factors (TFs), chromatin regulators (CRs) and genome structural proteins[3,4]. A classical model of this complexity is the hematopoietic system, where multiple mature effector cells derive from a single cell type; the hematopoietic stem cell (HSC)[5]. Hematopoiesis, the dynamic and reactive process of blood production is exquisitely regulated by a complicated interplay between multiple pioneer, lineage-specific and signal-dependent TFs, CRs and genome structural proteins[6–8]. These players are known to remodel the epigenetic landscape and reshape genome topology to allow communication between regulatory elements (e.g. promoter-enhancer).

In malignancies such as acute myeloid leukemia (AML), recurrent mutations in TFs, CRs and genome structural proteins, including components of the Cohesin complex and CTCF have been observed at high frequencies[9–11]. 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%[12]. However, although aberrant transcription is a cardinal feature of AML[13], not all cases carry mutations in the above classes of proteins, suggesting 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, 3D DNA topology and communication between enhancers and promoters to generate leukemia-specific transcriptional programs. To do so we have experimentally “deconstructed” AML utilizing an allelic series of mice that model different “transition states” during AML induction; normal, pre-malignant and overt leukemia. In addition, analysis of the single mutant (SM) pre-malignant 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.

Europe PMC Funders Author Manuscripts

Results

Murine models show transcriptional synergy of AML mutations

FLT3-ITD and *NPM1c* mutations are the most common in AML, occurring individually in ~25-30% and co-occuring in ~15% of all cases[9–11]. Mouse models carrying knock-in mutations of *Npm1* (conditional *Npm1* $^{\text{flox-cA/+}}$;*Mx1-Cre*, hereafter referred to as "*Npm1c*") or *Flt3* (constitutive *Flt3* $^{\text{ITD/+}}$, hereafter "*Flt3-ITD*") and referred to as single mutant (SM) mice, are associated with individual subtle, non-fatal but obvious pre-malignant phenotypes[14,15]. However, when combined (*Npm1* $^{\text{flox-cA/+}}$;*Mx1-Cre*;*Flt3* $^{\text{ITD/+}}$, "*Npm1c/Flt3-ITD*" or double mutant, DM) mice develop an aggressive AML with short latency[16]. To prospectively assess dynamic remodeling of the *cis*-regulatory landscape and 3D genome topology during leukemia development, we utilized an allelic series of wild-type (WT), *Npm1c*, *Flt3-ITD* and DM mice to model AML induction (Figure 1a, upper panel), analyzing the same population enriched for hematopoietic stem and progenitor cells (HSPC, negative for lineage markers, Lin-) from WT and mutant mice. Assessing gene expression (RNA-seq), multiple chromatin activation 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 (pCHiC)[18] Figure 1a, lower panel) across our allelic series, we derived a high-quality dataset with strong reproducibility between sample replicates (Supplementary Figure 1).

We first analysed differential gene expression between WT, individual pre-malignant 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 (Figure 1b,c, Extended Data Figure 1a and Supplementary Table 1). In stark contrast, when combined, *Npm1c* and *Flt3-ITD* strongly synergised to induce marked differential gene expression. Utilizing gene set enrichment analysis (GSEA), *Flt3-ITD* HSPC displayed signatures related to immune activation that were shared with DM Leukemic HSPC (Figure 1d,e and Extended Data Figure 1b). Conversely, genes downregulated in either *Npm1c* or DM HSPC were significantly enriched for leukemia-associated repressive gene signatures (Extended Data Figure 1c-e). Of note, many of the genes altered in the single mutant HSPC were also dysregulated in the same direction upon leukemia development (Figure 1f and Supplementary Figure 2a). Moreover, their expression often demonstrated an intermediate stepwise change from the single to double mutant stage, suggesting mutational synergy (Figure 1g and Supplementary Figure 2b). Examples of genes upregulated in *Npm1c* and DM cells included *Hoxa* genes (e.g. *Hoxa9*, *Hoxa10*) (Figure 1f,g). A small number of genes upregulated by both single mutations included exemplars of the immune activation programme including the AP-1 complex members *Jun, Fos* and *Fosb*, whose expression increased still further in DM HSPC. However, some genes such as *Spi1*, encoding the ETS-family TF Pu.1, *Setbp1*, and immune activation-associated Interferon regulatory factors (e.g. *Irf4, Irf8*) were upregulated only in DM HSPC (Figure 1f,g). Comparably, *Gata1*, *Gata2* and tumour suppressor genes *Gfi1* and *Gfi1b* were significantly downregulated only in DM HSPC (Figure 1g and Supplementary Figure 2a). Importantly, comparing genes differentially expressed in *NPM1c/FLT3*-ITD mutated human AML[19] with our differentially expressed genes between DM and WT HSPC, we could

Europe PMC Funders Author Manuscripts

demonstrate a highly significant overlap with 27% up- (119/441, p=1.28e-07) and 15% (74/503, p=1.27e-07) of downregulated human genes also evident in our dataset Figure 1h). Taken together, these data demonstrate marked mutational synergy not evident when only single mutations are present.

Flt3-ITD and *Npm1c* mutations 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 promoters[20]. 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 (Figure 2a,b and Supplementary Table 2). Compared with gene expression, more alterations in accessibility were preserved between the *Flt3-ITD* and DM HSPC (Figure 2c). In comparison to the degree of increased or decreased chromatin accessibility elicited by DM at differentially accessible regions, changes were intermediate in *Flt3-ITD* HSPC (Figure 2d). In contrast, less overlap was seen for the changes of accessibility evident between *Npm1c* and DM HSPC (Figure 2c and Extended Data Figure 2a). However, within this overlap were four regions located in the *Hoxa* locus (Figure 2e), 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 enhancers[21,22] (Extended Data Figure 2b). Interestingly, accessibility changes in single mutant HSPC did not significantly alter gene expression, but did correlate with gene expression upon leukemia induction (Extended Data Figure 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 (Figure 2f and Extended Data Figure 2d-f). Further Bivariate Genomic Footprinting (BaGFoot) analysis[23] simultaneously assessed differential footprint depth of TFs and altered accessibility flanking their motifs between conditions (Figure 2g and Extended Data Figure 2g,h). This confirmed that GATA factors demonstrated decreased accessibility and reduced footprint depth suggesting a loss of binding in DM HSPC (Figure 2g). Conversely, increased accessibility and footprint depth was noted for Cebpa, Cebpd 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, suggesting a potential role for prior-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 tri-methylation 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. Enhancers were considered cis-regulatory modules usually flanked by high levels of H3K4me1 and no or low levels of H3K4me3 (n=98,365, Extended Data Figure 3a and Supplementary Table 3) and were designated “primed” when lacking H3K27ac or “active”

if marked with H3K27ac[24]. To further enrich for cis-regulatory elements we overlapped this compendium with our ATAC-seq analysis to identify "accessible" enhancers (n=48,052, Extended Data Figure 3b and Supplementary Table 3). This cis-regulatory element repertoire was highly dynamic across the cellular states; ~51% demonstrated differential H3K4me1 between mutant and WT HSPC (Extended Data Figure 3c). Of note, 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 (Figure 3a,b and Supplementary Table 4) but a marked overlap between the two (Extended Data Figure 3d). Strikingly, and by marked contrast, *Npm1c* HSPC were virtually indistinguishable from WT in H3K4me1 patterns (Figure 3a,b). Moreover, among all DM dynamic elements, intermediate changes of H3K4me1 were observed in *Flt3-ITD* but not *Npm1c* HSPC Figure 3c and Extended Data Figure 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 Figure 3f,g). *Flt3-ITD* alone and DM demonstrated H3K27ac changes at hundreds of enhancer regions, whereas *Npm1c* did not significantly alter global H3K27ac signal (Figure 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 (Figure 3e and Extended Data Figure 3h).

Although no significant 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 Figure 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 and included similar available GMP maps[25] for the comparison. For DM gained enhancers, 31% (3,637/11,892) overlapped with similar changes in neutrophils and GMP (Figure 3f group Gain-1). However, the remaining 69% appeared leukemia-specific (group Gain-2, Supplementary Table 5). In contrast, of DM lost enhancers, the majority showed similar changes in both neutrophils and GMP (83%, 10,362/12,509, group Loss-2), while ~17%, although distinct from neutrophils, were highly similar to GMP (group Loss-1). Exemplar leukemia specific regulatory elements are shown for the *Socs2* gene (Figure 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 Figure 3g).

Leukemia programs utilize novel and existing 3D contacts

To capture genome-scale alterations in 3D DNA topology, we employed pCHiC[18] 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 (Figure 4a, Extended Data Figure 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 (Figure 4b,c and Supplementary Table 7). A total of 290 regions altered their compartment assignment from “inactive” B to “active” A (Figure 4c). These regions contained 345 expressed genes, including known oncogenes such as *Setbp1*, *Igf1* and a contiguous region involving an interferon-inducible *Ifi* gene cluster (Figure 4d and Extended Data Figure 5d,e). Alteration from an active to an inactive compartment, A to B, occurred at 274 regions (Figure 4c) and covered 453 genes including the tumour suppressor gene *Dach1* and the *Fancf* gene that is silenced by methylation in AML[28].

Assessing total interactions at individual promoters, only a few promoters were altered for *Npm1c* HSPC in comparison to WT. In contrast, in the presence of the *Flt3-ITD*, several hundred promoters showed lost- or gained-interactions, with this number roughly doubling in DM HSPC (Figure 4e). Moreover, nearly half of those promoters altered by *Flt3-ITD* overlapped with the change in DM HSPC (Figure 4f). At the level of individual interactions, of note, 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[29,30] (Figure 4g), with more interactions lost than newly formed in the mutant states (Figure 4h and Supplementary Table 8). We named these “rewired” interactions, and these interactions could be either lost or gained (e.g. *Gfi1b* or *Irf8*, respectively; Figure 4i,j and Extended Data Figure 5f,g). In contrast, the remainder of interactions were already pre-set in WT HSPC and were termed “hard-wired” DNA contacts Figure 4g).

Integrated analysis identifies critical regulatory networks

We first applied a multi-layered approach to integrate the multi-omics chromatin analysis[31] at all cis-regulatory regions across wildtype and mutant HSPC. Briefly, focusing on 2 kilobase (kb) bins of accessible regions (±1kb from ATAC-seq peak summit, n=75,457), normalised reads for chromatin modifications and accessibility of each HSPC sample were integrated. Treating these bins as a separate data point across all 16 conditions (4 analyses in 4 cellular states), we were then able to use dimensionality reduction tools developed for scRNA-Seq (Seurat[32]) to classify and visualise clusters of regions showing similar patterns (Figure 5a). This classified 10 individual clusters Figure 5a,b). Cis-regulatory regions with a loss of enhancer signatures during leukemia induction, characterised by concurrent loss of H3K4me1 and accessibility, with or without evident loss of H3K27ac (Cluster-9 or -10, respectively; Figure 5b,c, Extended Data Figure 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 (Figure 5d and Extended Data Figure 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 localised genes or with spatially-contacted genes using chromatin interaction profiles. Genes upregulated (n=978) in Cluster-6 demonstrated significant enrichment for immune process (Figure 5e). These included multiple genes of interest for leukemogenesis, including *Jun*, *Fos, Hoxa10* and *Irf8* (Figure 5f and Supplementary Table 10). By contrast, genes downregulated in Cluster-10 were enriched for DNA replication and contained hematopoietic TFs including *Gata1*, *Gata2*, *Gfi1b* and *Myb* (Extended Data Figure 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[19]. We could demonstrate that over a quarter of the human genes (275/1031, 26%, p=2.58e-06) were present within our Cluster-6 genes, demonstrated enriched signatures for immune regulatory processes and the majority showing increased gene expression in our system (Figure 5g-i and Extended Data Figure 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 Figure 7a). Chromatin interaction frequencies reflected by the flipping of genome compartments generally correlated with gene expression (Figure 6a and Extended Data Figure 7b). Integrating chromatin modifications with differential interactions, we demonstrated a strong correlation of H3K4me1 alteration with “rewired” interactions (Figure 6b and Extended Data Figure 7c,d). Similarly, flipping of genome compartments positively associated with alteration of enhancer marks including H3K4me1 and H3K27ac, as well as chromatin accessibility (Figure 6c).

Our analysis also identified 801 super-enhancers (SE) across WT and mutant HSPC (Extended Data Figure 8a and Supplementary Table 11). Patterns of loss and gain were similar to standard enhancers when linked to mutations (Extended Data Figure 8b,c and Supplementary Table 11). However, for *Npm1c* one of the three gained SE demonstrated a long-range contact to the *Hoxa* locus and was associated with upregulation of most *Hoxa* genes (Extended Data Figure 8d). Linking altered SE to their regulated genes using DNA interaction profiles revealed a strong correlation to expression during DM leukemia induction (Extended Data Figure 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 Figure 8e).

Leukemic regulation of exemplar reprogrammed loci

Our integrated analysis suggested *Spi1*/Pu.1 as an important network regulator in *Flt3-ITD*/*Npm1c* leukemia (Figure 6d and Extended Data Figure 8f). All WT and single mutant HSPC demonstrated H3K4me1 priming and accessibility at the well described *Spi1* upstream regulatory element (URE)[33–36]. However, H3K27ac increased only in DM HSPC, where it associated with a “rewiring” of the URE to communicate with the *Spi1* 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 (Figure 6e and 4j).

Definitively linking promoters to cis-regulatory regions using pCHiC also allowed us to identify novel long-range regulatory interactions. An exemplar of this was seen for the *Hoxa* locus, where “hard-wired” interactions with an H3K4me1 modified region ~1MB upstream (Figure 6f and Extended Data Figure 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 super-enhancer (*Hoxa*-LRSE), demonstrated a marked increase in H3K27ac modification in both *Npm1c* and in DM HSPC, where it appears to collaborate with the increased accessibility across the promoters of *Hoxa* genes (Supplementary Figure 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 Figure 3b). Using published data (including ChIP-seq on H3K4me1[6] and CTCF[37], ATAC-seq[38], together with pCHiC[18]) of human CD34+ HSPC, we could demonstrate that the interaction was also present across species, and ChIP-seq data from ourselves[39] and others[40,41] 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 Figure 3b). The human *HOXA*-LRSE was also demonstrated to have an open chromatin confirmation in *NPM1* mutant AML patients by DNase hypersensitivity analysis[19], further demonstrating cross-species conservation of this long-range element.

Network perturbation abrogates leukemia maintenance

Utilizing our integrated analysis to identify exemplar putative network nodes, we focussed on the AP-1 complex members *c-Fos* and *c-Jun*, the TFs *Spi1*/Pu.1, *Irf8*, the oncogene *Igf1*, the *Hoxa* genes and the regulatory loci of *Spi1* 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* significantly decreased clonogenic capacity in murine DM cells *in vitro* (Extended Data Figure 9a,b). Depletion of *Spi1* could abrogate leukemia cell growth and clonogenicity (Extended Data Figure 9c-e). To validate *Spi1* regulatory elements as critical for leukemia maintenance, CRISPR-Cas9 mediated genetic excision of the *Spi1*-URE was achieved using dual guide RNAs to target its 5kb central region in DM cells carrying Cas9[42] Figure 7a and Extended Data Figure 9f,g). Removal of the URE resulted in ~30% reduction of *Spi1* expression in bulk cultures (Figure 7b), significantly decreased cell growth and clonogenicity (Figure 7c,d and Extended Data Figure 9h).

The requirement of sustained expression of *Hoxa9* or *Hoxa10* for the maintenance of leukemia cell growth and clonogenicity was also demonstrated (Extended Data Figure 9c-e). Furthermore, we used a similar experimental strategy to examine the role of the *Hoxa*-LRSE by removing its ~2.4kb central region in DM-Cas9 cells (Figure 7e and Extended Data

Figure 9i,j). Genetic disruption of *Hoxa*-LRSE resulted in significantly reduced expression of all *Hoxa* genes tested (*Hoxa3, 5, 6, 7, 9, 10*, Figure 7f). In addition, excision of the *Hoxa*-LRSE also induced a significant decline in cell proliferation and reduced colony-forming capacity (Figure 7g,h and Extended Data Figure 9k).

To further correlate the relevance of these networks to human AML we also validated *SPI.1*/PU.1 as well as extending the analysis to other targets, including *IRF8* and *IGF1*. These targets were edited via CRISPR-Cas9 in the human *HOXA*-dependent AML cells lines OCI-AML3 (mutated for *NPM1*, but also *DNMT3A*) and MOLM-13 (that carries a *FLT3-ITD* mutation, but also an *MLL-AF9* rearrangement). Importantly, these experiments corroborated our murine findings, with editing decreasing proliferation and clonogenic function (Figure 7i-l and Extended Data Figure a-f). 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 only between 3 and 5 mutations on average[11]. Our study provides mechanistic evidence why multiple mutations are necessary for the full malignant phenotype and suggests that significant 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 prospectively and categorically confirms 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 differentiation[25]. 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[43,44], interferon response factors (particularly IRF4 and -8)[43] and signal responsive TFs such as PU.1[45] and RUNX1[46] in this regulation. Of note, these findings accord with and extend similar genomic analysis in patient samples from *FLT3-ITD* mutated AML[19,47]. By contrast, the major alteration that occurred in HSPC expressing *Npm1c* was an alteration of chromatin accessibility, with an obvious exemplar the changes at the *Hoxa* locus demonstrated to drive leukemogenesis. The nature of this

alteration in chromatin accessibility is not immediately apparent, however, may relate to the known histone chaperone function of NPM1[48], to its interaction with ATP-dependent chromatin remodeling complexes[20] or to its role as the major protein component of the nucleolus, a structure that regulates genome 3D topology[49,50] and further work is warranted to investigate these possibilities.

Our study has prospectively modelled the alterations in 3D genome organization that accompany the induction of hematological malignancy. Of note, 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[51], were dynamically rewired in leukemic HSPC. The majority of interactions remain stable, with the gene expression alterations presumably reflecting alterations of chromatin modification and accessibility at these loci. However, utilization of other cis-regulatory elements, such as those upregulating *Socs2* and *Irf8* (Figure 3e and 6e), was novel and specific for leukemia. These data demonstrate that malignant transformation is promiscuous in its enhancer usage, as has also been previously suggested[19,47]; it can generate novel DNA-topologies and enhancer states but may also utilize 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 models[52,53], we see that they constitute similar vulnerabilities across multiple AML genotypes. PU.1 is classically regarded as a tumour suppressor in AML. Mice with a genetically engineered decrease in PU.1 expression develop AML[35,36,54,55]. Furthermore, heterozygous deletions of the locus or rare loss-of-function mutations of PU.1 have been described in patients with AML[9,11,56]. Moreover, oncogenic events abrogate the activity of PU.1[57,58] and therapeutically, the anti-leukemic activity of LSD1 inhibitors has been associated with re-activation of a PU.1 driven transcriptional programme[59,60]. However, in contrast, our study interestingly identifies *Spi1*/PU.1 as an oncogenic TF in *Npm1c*/*Flt3*-ITD AML, and is in accord with our previous work and that of others that have demonstrated PU.1 to be a vulnerability in certain AML subtypes[52,53,61]. 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, PU.1's oncogenic effect appears to occur in combination with its known co-activators, AP-1[62,63], IRFs and Hoxa9[64,65], 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 an entirely novel *Hoxa*-long range super-enhancer (*Hoxa*-LRSE) that appears to regulate the entire *Hoxa* cluster following *Npm1* mutation. Importantly, we demonstrate that the *Hoxa*-LRSE is not species-specific but is conserved within human cells. Further study of the signals integration at the *Hoxa*-LRSE will inform its critical roles for hematopoiesis and leukemogenesis[66,67].

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.

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 strain mice carrying single mutations (conditional *Npm1* $^{\text{flox-cA/+}}$;*Mx1-Cre+*, or constitutive *Flt3* $^{\text{ITD/+}}$) or the combined mutations (*Npm1* $^{\text{flox-cA/+}}$;*Flt3* $^{\text{ITD/+}}$ *;Mx1-Cre+*) have been described previously[14,15,16]. The activation of the *Npm1* $^{\text{cA}}$ allele was achieved by induction of Cre-mediated recombination in 5- to 6-week-old *Npm1* $^{\text{flox-cA/+}}$;*Mx1-Cre+* mice which were administered six doses of polyinosinic-polycytidylic acid (pIpC). The activation of *Npm1* $^{\text{cA}}$ allele in mice carrying *Npm1* $^{\text{flox-cA/+}}$;*Flt3* $^{\text{ITD/+}}$ *;Mx1-Cre+* does not require administration of pIpC to induce Mx1-Cre for recombination[16]. For tissue harvest, mice of each genotype were utilized regardless of their sex. Wildtype (WT) C57/BL6 mice and *Flt3* $^{\text{ITD/+}}$ mice were sacrificed at 10-12 weeks of age; *Npm1* $^{\text{flox-cA/+}}$ mice were sacrificed immediately after 6x pIpC treatment (aged 11-12 weeks of age); 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) and were checked on a daily base to identify signs of heavy leukemia burden, including ruffled fur, hunchback, heavy shivering, weight loss, extreme weakness, or significant loss of activity. No mice were allowed to exceed the limit of leukemia burden, under the ethical regulations of University of Cambridge Animal Welfare and Ethical Review Body (AWERB).

Flow cytometry analysis

Flow cytometry Measurement was conducted in 7-9 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 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). Hematopoietic stem and progenitor populations (HSPC) were determined using the combined immunophenotyping markers as below: Lin-, lineage negative; HPC, hematopoietic progenitors (containing all myeloid progenitors), Lin$^-$Sca1$^-$cKit$^+$; LSK, enriched for HSC, Lin$^-$Sca1$^+$cKit$^+$; CMP, common myeloid progenitors, Lin$^-$Sca1$^-$cKit$^+$CD34$^+$CD16/32$^{\text{int}}$; GMP, granulocyte/macrophage progenitors, Lin$^-$Sca1$^-$cKit$^+$CD34$^+$CD16/32$^{\text{hi}}$; MEP, megakaryocyte/erythroid progenitors, Lin$^-$Sca1$^-$cKit$^+$CD34$^-$CD16/32$^{\text{lo}}$. Assays were performed on a BD LSRFortessa cell analyzer and all data were analyzed with FlowJo software (Tree Star, version 10.5.0).

Gene expression analysis

RNA-seq reads were quality filtered and mapped using STAR[68] (version 2.4.0) against the mouse genome (mm10). Uniquely mapped reads were quantified with HTSeq (version 0.6.0) vand protein-coding genes with non-zero read count in wildtype or mutant HSPC (n = 16,771) were included for downstream analysis. Reads Per Kilobase Million (RPKM) mapped reads for each protein-coding genes were calculated using Bioconductor package edgeR (version 3.28.1)[69]. Principal component analysis (PCA) was performed on RPKM values of 16,771 protein-coding genes to assess reproducibility[70]. Differential expression of protein-coding genes was analysed with these counts using Bioconductor package DESeq2 (version 1.12.4)[71]. Significantly differential expression was considered by setting adjusted *P* value (adjP) < 0.05 and fold change (FC) ≥ 1.5 between mutants and wildtype HSPC. Unsupervised clustering was used to generate the heatmap.

ChIP-seq

Chromatin immunoprecipitation was performed with iDeal ChIP-seq kit for Histones (Diagenode) following manufacturer's recommendations. In brief, 1 x 10^6 Lin- HSPC or neutrophils isolated from wildtype 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 minutes and then quenched by addition of Glycine (Thermo Fisher Scientific) for 5 minutes incubation. Lysis buffers iL1 and iL2 were used to prepare the nuclei pellets. Chromatin was sheared in Shearing buffer iS1 using the Bioruptor Plus (Diagenode) for 20 cycles (each cycle for 30 seconds "ON" and 45 seconds "OFF") at high power setting. Immunoprecipitation, wash, decrosslinking and elution were carried out as per the manufacturer's protocol, using 1 μg antibody (anti-H3K4me1, #ab8895, Abcam; anti-H3K27ac, #ab4729, Abcam; anti-H3K4me3, #c15410003, Diagenode). 10% of the sheared chromatins were kept aside as input samples. ChIP-seq library preparation of ChIP DNA or input DNA was performed using TruSeq ChIP Sample Prep Kit (Illumina) following standard procedures from manufacturer. The library DNA was quantified using KAPA Library Quantification Kit (Kapa Biosystems) and library average size was determined by Agilent DNA 1000 Kit (Agilent Technologies). Libraries were pooled for single-read (1 x 50 bp) sequencing on an Illumina HiSeq 2500 or HiSeq 4000 platform at CRUK Cambridge Institute Genomics Core. Experiments were performed in duplicate on biologically independent samples as recommended[72].

Analysis of histone modifications

ChIP-seq reads were aligned to the mouse reference genome (mm10) using Bowtie (version 2.1.0)[73]. Duplicates were removed using PICARD tools (version 2.2.1). As for H3K4me1 or H3K4me3, enriched regions (peaks) were identified using findPeaks program from HOMER software (version 4.10.4)[26], with the setting of "-size 1000 -minDist 1000". Peaks overlapped in both replicates of individual cell type were kept. If centers of two consecutive peaks were less than 500 bp, only the peak with higher enrichment signal was counted. H3K4me1 or H3K4me3 peak sets were generated by combining total peaks in all four HSPC samples and extending +/- 500 bp. Frequency distribution of H3K4me3 peaks were then plotted based on H3K4me3 enrichment, and peaks with high levels of

H4me3 (CPM > 16, Extended Data Figure 3a) were considered to be active promoters. H3K4me1-based total enhancer catalog was created by excluding H3K4me1 peaks that intersect with active promoters, and then merging the overlapped ones. Accessible enhancers were defined as ATAC-seq consensus peaks overlapping with total enhancers. H3K27ac ChIP-seq peaks were called using MACS2 (version 2.0.1)[26] against input sample of each cell type with a *p*-value cutoff of 1e-9 and with setting "nomodel". A list of H3K27ac consensus peak set was made using DiffBind (version 2.0.1)[74]. Accessible enhancers that overlap with any H3K27ac peaks were defined as active enhancers. Differential enrichment of chromatin marks (H3K4me1 and H3K27ac) was analysed using edgeR (version 3.28.1), with significant changes being defined by FDR value < 0.05 and FC ≥ 1.5 (gain or loss) in the presence of any mutations. Accessible enhancers with gain or loss of H3K4me1 levels in any mutant vs WT HSPC were defined as dynamic enhancers, likewise, with gain or loss of H3K27ac enrichment were considered with 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 neutrophils (NEU) and GMP. Histone profiles and accessibility for GMP cells were previously reported[25] and available in Gene Expression Omnibus (H3K4me1, GSM1441289; H3K27ac, 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 (version 3.1.3)[75]. Enhancers showing unique changes in DM HSPC as compared to WT HSPC and neutrophils were considered "leukemia-specific" changes.

Chromatin accessibility analysis

Chromatin accessibility probed by ATAC-seq was analysed in like manner as 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 *p*-values less than 1e-20 were considered significant. A list of ATAC-seq consensus peak set was made using DiffBind. Differential enrichment was analysed 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

Scanning *de novo* motifs was performed at ±100 bp from ATAC-seq peak summit within the investigated chromatin regions, using findMotifsGenome.pl with the setting "-mask -p 12 -S 10" from HOMER package. In most cases, motifs occurred with significant *p*-values and with a coverage of > 10% of target regions were displayed. Combined analysis of TFs' occurrence and chromatin accessibility at regions of interest was assessed by Bivariate Genomic Footprinting (BaGFoot, version 0.9.7) algorithm following the procedures as described[23]. Analysis was performed 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 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 normalised to total sequencing counts in each sample. Differential accessibility was then determined as altered read counts flanking each motif in mutant vs 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 less than 0.05 were considered significant.

Promoter-anchored interaction analysis

Paired-end sequences of pCHiC were processed using the HiCUP (version 0.5.8)[76] pipeline with default parameters for the following steps: quality control, identification of reads containing HiC 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 CHiCAGO (version 1.1.1)[77], to call significant promoter-based interactions. CHiCAGO considers distance effect on interaction frequencies by virtue of a convolution background model and a distance-weighted p-value. CHiCAGO scores represent -log weighted p-values, the higher the more likelihood of interaction formed. CHiCAGO scores were calculated for each pCHiC sample, and significant interactions were called when CHiCAGO scores ≥ 5. Significant interactions were computed per HSPC by merging their replicates, and were combined to form a matrix of total unique interactions. While, 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 vs 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 ≥ 5), with scores ranked in the bottom quartile in WT but in the top quartile in mutant, were considered mutation-associated gained interactions. And vice versa, high-confidence interactions that were present in WT but absent in mutant, ranked in the top quartile in WT but in the bottom quartile in mutant, were considered as lost interactions by mutations. In addition, total unique interaction profiles facilitated the annotation of distal cis-regulatory elements (e.g. SEs) to their target genes in HSPC.

Chromatin compartment analysis

Sub-nuclear compartmentation represented by self-associating chromatin domains were analysed by means of principal component analysis (PCA) on capture HiC data in a similar way as described on HiC data[27], mainly using HOMER software. To do so, uniquely mapped reads from HiCUP analysis were used to create "Tag Directory" using makeTagDirectory from HOMER. Principal component (PC1) values were calculated by

running runHiCpca.pl with default setting (resolution at 50 kb). 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 values were stitched to be identified as A or B compartments, respectively. Genome-wide correlation of compartment PC1 values between mutant and WT cells was performed by running getHiCcorrDiff.pl from HOMER. Flipped compartments were identified using HOMER findHiCCompartments.pl. Genes involved in the flipped compartments were selected to analyse their mRNA expression changes by mutations.

Seurat-guided clustering analysis to identify differential patterns of accessible enhancers

A multi-layered multiomic approach to integrate chromatin analysis[31] at all accessible regions across wildtype and mutant HSPC was applied. Firstly, within 2-kb bins at ATAC-seq consensus peaks (±1 kb from peak summit, n=75,457), normalised sequencing read counts for each chromatin modifications (H3K4me1, H3K4me3 and H3K27ac) and chromatin accessibility of each HSPC population (WT, *Npm1c*, *Flt3-ITD* and DM) were extracted and computed as input data matrix. This matrix was composed of 16 rows (with 4x chromatin marks and 4x cell populations) and 75,457 columns (with bins of accessible regions). Then processing the data in a similar way as for single cell RNA-sequencing with Seurat (version 3.2.3) package[32], all 75,457 accessible regions (as columns) were treated as a separate data point across all 16 conditions (as rows). This allowed dimensionality reduction to classify and visualise clusters of regions whose pattern of alterations were similar. Data were plotted on a tSNE plot to visualize all separated clusters (n=10) and a heatmap was shown to illustrate the specific patterns of chromatin changes in each cell type (no row-wise ranking). These clusters were further loaded as regions of interest into deepTools to probe average profiles of chromatin modifications and accessibility across WT and mutant conditions, further supporting the separation of tSNE clusters and their unique changes of chromatin states.

Gene perturbation through shRNA-mediated knock-down

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 mIL-3, 10 ng/mL hIL-6, and 50 ng/mL mSCF (Peprotech). Lentiviral shRNA plasmids were obtained from Sigma-Aldrich in the form of MISSION pLKO.1-puro vectors and shRNA sequences were provided in Supplementary Table 12. Lentiviral particles were produced by co-transfection of shRNA plasmids with psPAX and pMDG.2 in 293T cells using the Trans-IT LT-1 transfection reagent (Mirus). 293T cells were cultured in DMEM (Gibco) plus 10% fetal bovine serum (Gibco). DM cells were infected by shRNA lentivirus twice via spinoculation in the presence of 5 μg/mL polybrene (Sigma-Aldrich). 72 hours post transduction, cells were selected with 2 μg/mL puromycin (Sigma-Aldrich) for 72 hours.

Disruption of cis-regulatory elements through CRISPR-mediated deletion

Mouse leukemia cells carrying *Npm1c*/*Flt3-ITD*/*Cas9* (termed DM-Cas9 cells)[42] were cultured in X-Vivo medium (Lonza) plus 10% fetal bovine serum (Gibco) supplemented with cytokines as described above. In the presence of Cas9, using dual guide RNAs

(gRNAs) to specifically target two genomic loci has been demonstrated previously[52]. Paired dual gRNAs targeting specific cis-regulatory elements or scramble control dual gRNAs were designed with IDT web tools (https://eu.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM) and their sequences were available in Supplementary Table 12. Oligos of dual gRNAs were cloned into lentivirual vector pKLV2.2-h7SKgRNA5(SapI)-hU6gRNA5(BbsI)-PGKpuroBFP-W (#72666, Addgene). Production of gRNA particles and transduction of DM-Cas9 cells were carried out same as for shRNA. To confirm the target deletion in the bulk transduced cells, PCR primers were designed to amplify the wildtype or modified allele (Supplementary Table 12). Size of PCR products was checked by agarose gel electrophoresis. Furthermore, PCR products were subject to TA cloning (Promega). Plasmid DNA was extracted from individual colonies and was confirmed for deletion 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 were generated by lentiviral transduction using pKLV2-EF1aBsd2ACas9-W vector (Addgene #67978)[52]. 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 guide RNAs were designed and subcloned into the pKLV2-U6gRNA5(BbsI)-PKGpuro2ABFP-W expression vector (Addgene #67974)[52]. A non-target control gRNA was also included. Target sequences of all gRNAs are detailed in Supplemental Table 12. 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 in total cell numbers relative to corresponding non-targeting control. For colony forming assays, 500-1000 cells transduced with specific gRNAs were seeded in duplicate in Methocult semi-solid medium (H4531, STEMCELL Technologies). Total colony forming units were enumerated 7 days post-culture. Three replicates for each target or control were carried out. Data were presented as a fold change in colony forming units or colony cellularity, relative to non-targeting counterparts.

Statistics & Reproducibility

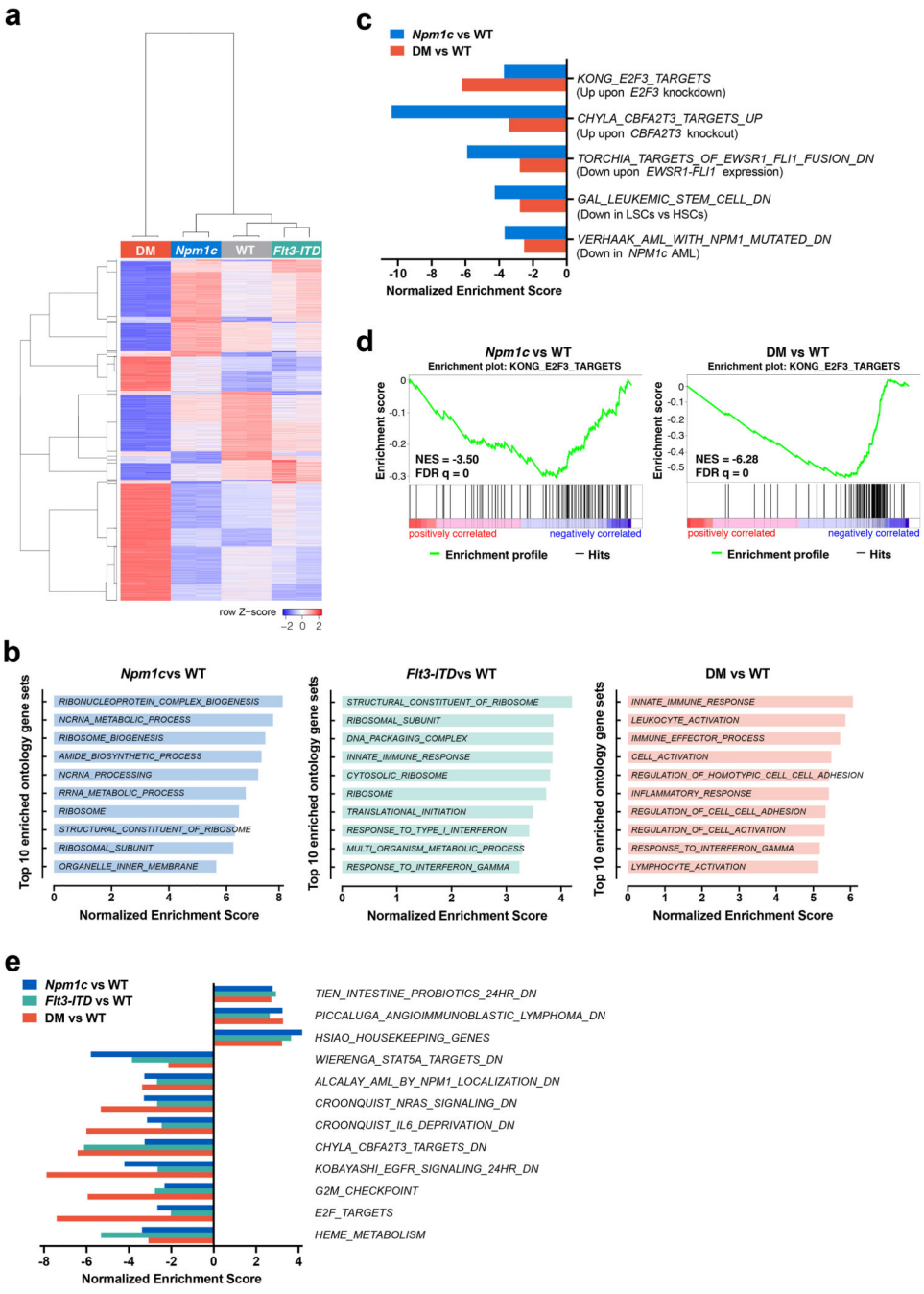
Statistical analyses in the studies were specified in details in figure legends. Sample sizes were not predetermined using any statistical method. Randomization was applied to mouse samples preparation and cell perturbation experiments. The Investigators who collected the mouse samples were not informed 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 non-mitochondrial genome. Student's unpaired or paired t-tests were performed as two-tailed with GraphPad Prism (version 8.2.1); One-tailed hypergeometric distribution was analysed in R (version 3.6.1); p values in HOMER *de novo* motif analysis were calculated as one-tailed or in BaGFoot 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.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

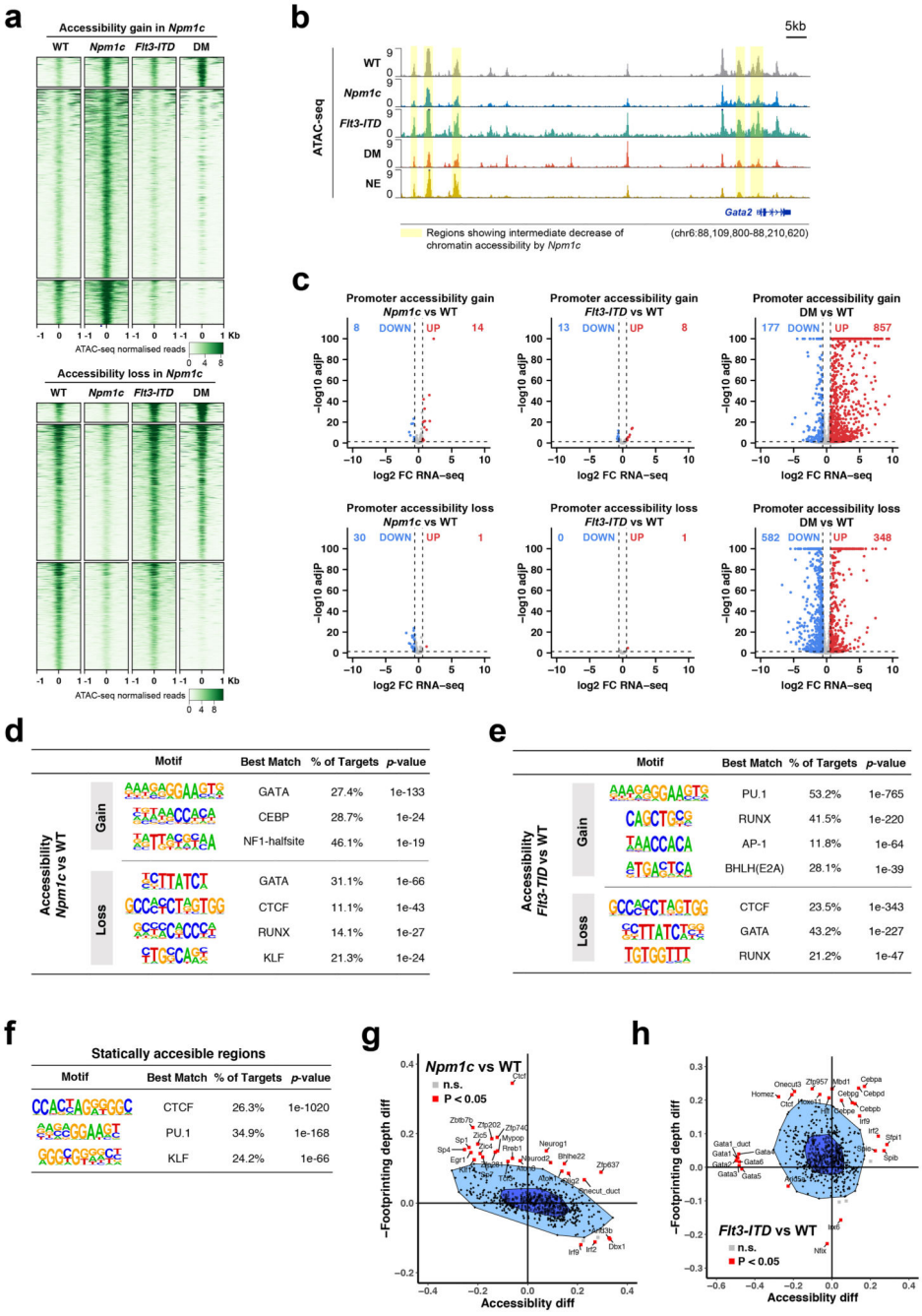
Extended Data



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.



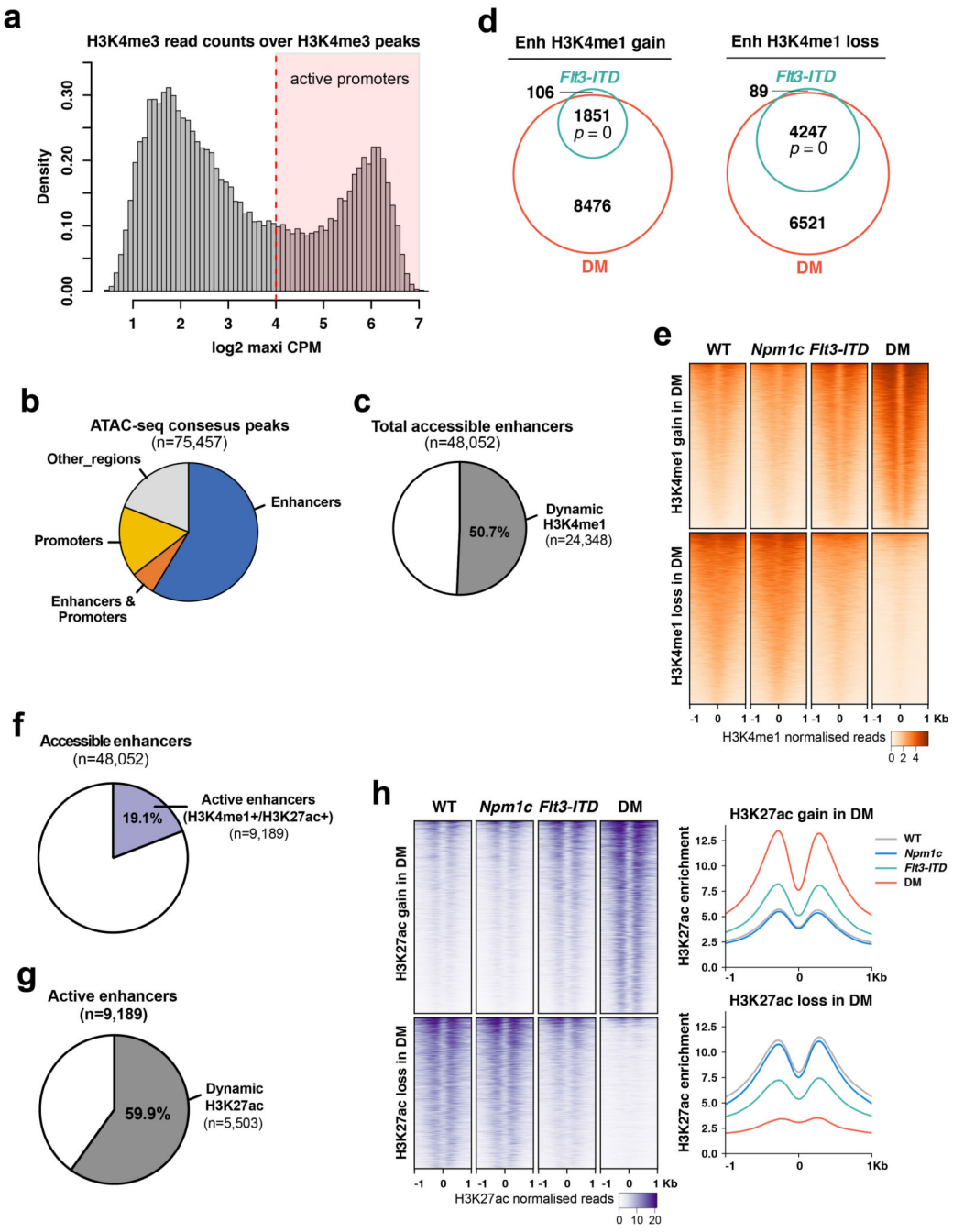
Extended Data Fig. 2. Global chromatin accessibility across WT and mutant HSPC.

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

Europe PMC Funders Author Manuscripts

accessibility in *Npmc1* 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.

Europe PMC Funders Author Manuscripts



Extended Data Fig. 3. Chromatin modifications at accessible enhancers across WT and mutant HSPC.

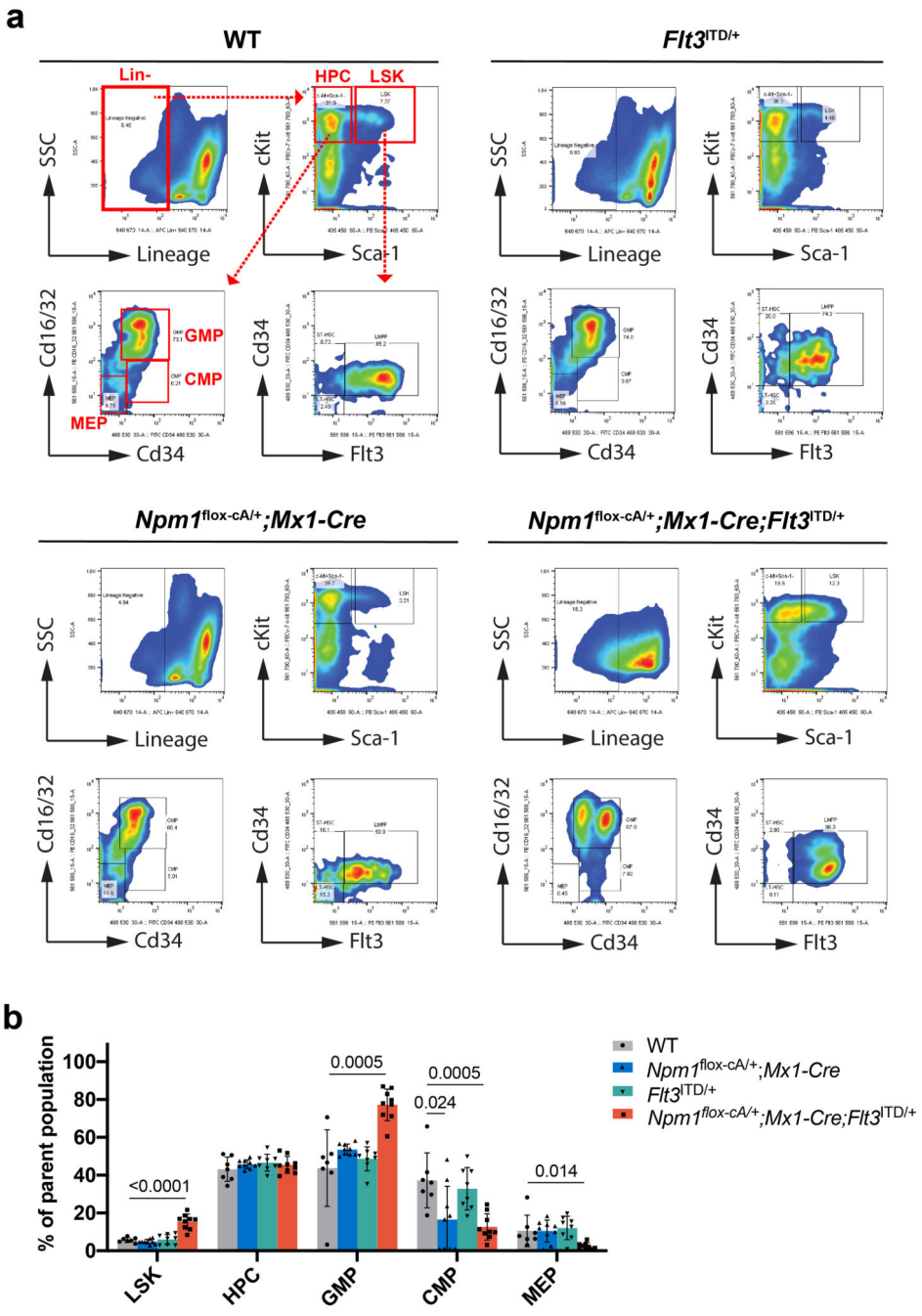
a, Distribution of all H3K4me3 peaks based on read counts across all four cell types. Active promoters were marked by high H3K4me3 with log2 CPM > 4. **b**, Genome distribution of ATAC-seq consensus peaks. **c**, Number of total accessible enhancer peaks defined in all four HSPC samples and percentage 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.

Europe PMC Funders Author Manuscripts

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.

Europe PMC Funders Author Manuscripts

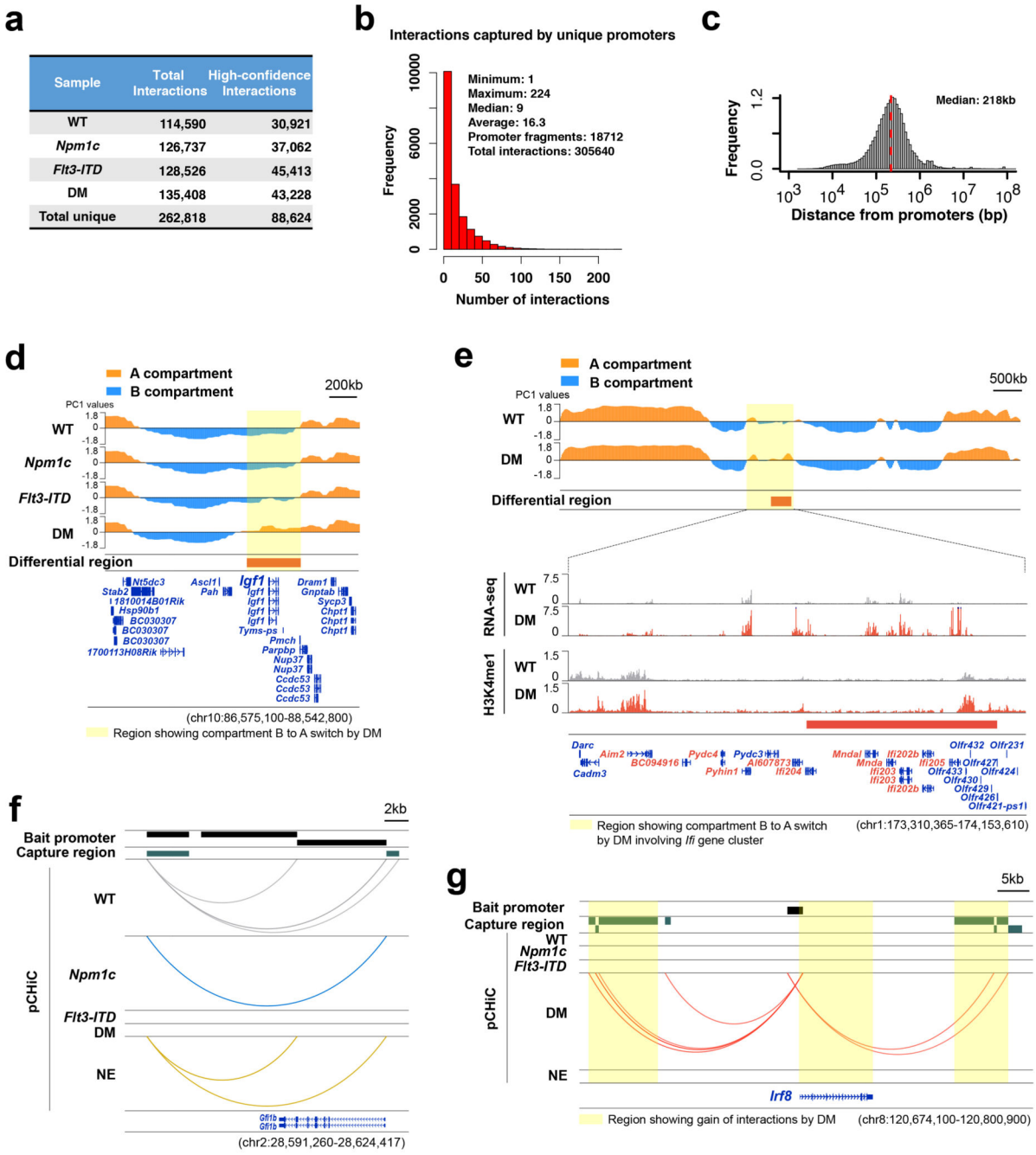
Europe PMC Funders Author Manuscripts



Extended Data Fig. 4. Immunophenotypic characterization of HSPCs isolated from bone marrow of wildtype and mutant mice.

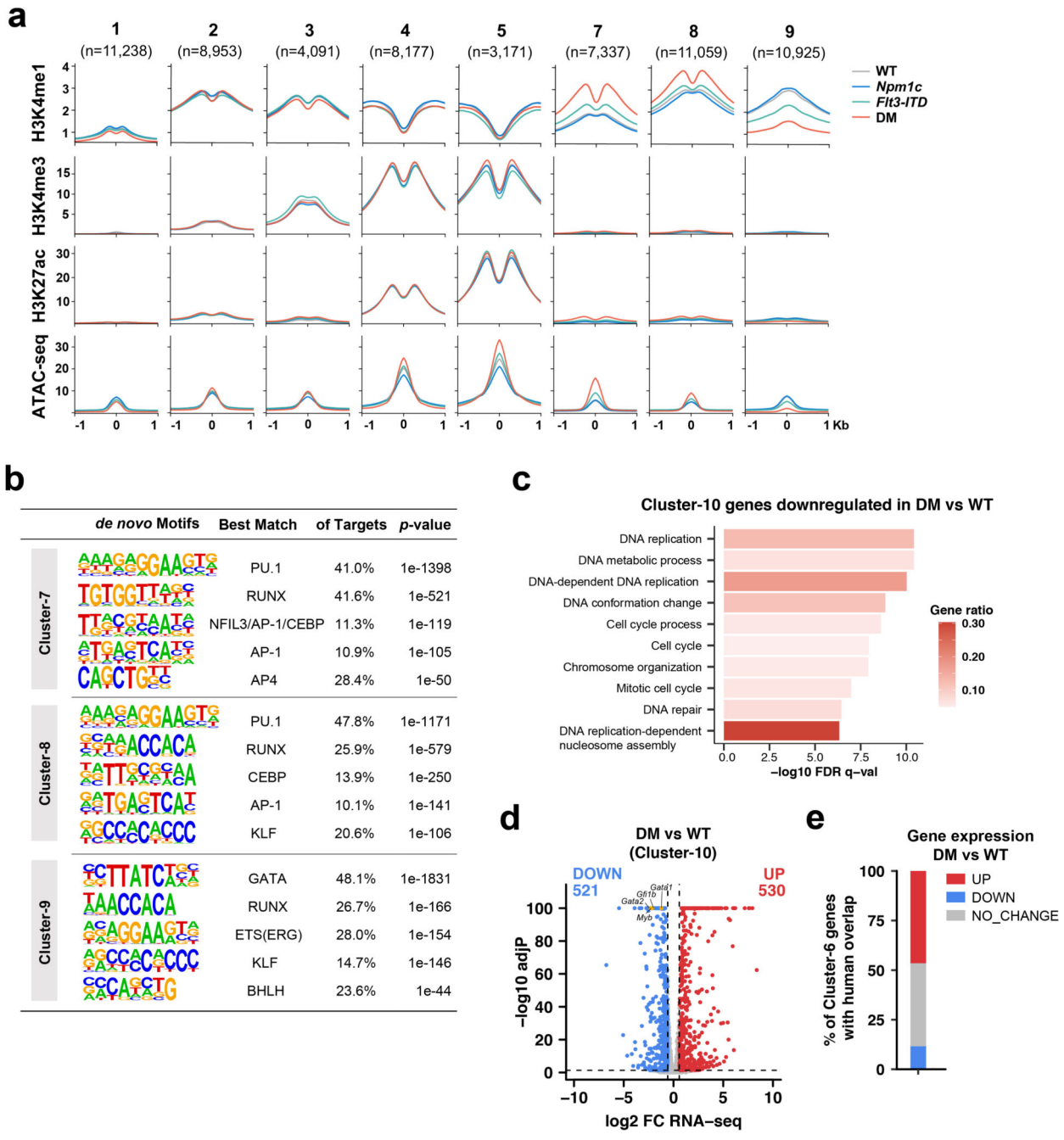
a, Representative flow cytometry plots showing proportions of hematopoietic stem and progenitors within bone marrow compartment. Lin-, lineage negative; HPC, hematopoietic progenitors ($Lin^{-}Sca1^{-}cKit^{+}$) containing all myeloid progenitors; LSK, $Lin^{-}Sca1^{+}cKit^{+}$; CMP, common myeloid progenitors; 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* $^{\text{flox-cA/+}}$;*Mx1-Cre*, n=9;

Flt3$^{ITD/+}$, n=8; *Npm1*$^{flox\text{-}cA/+}$;*Mx1-Cre*;*Flt3*$^{ITD/+}$, 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

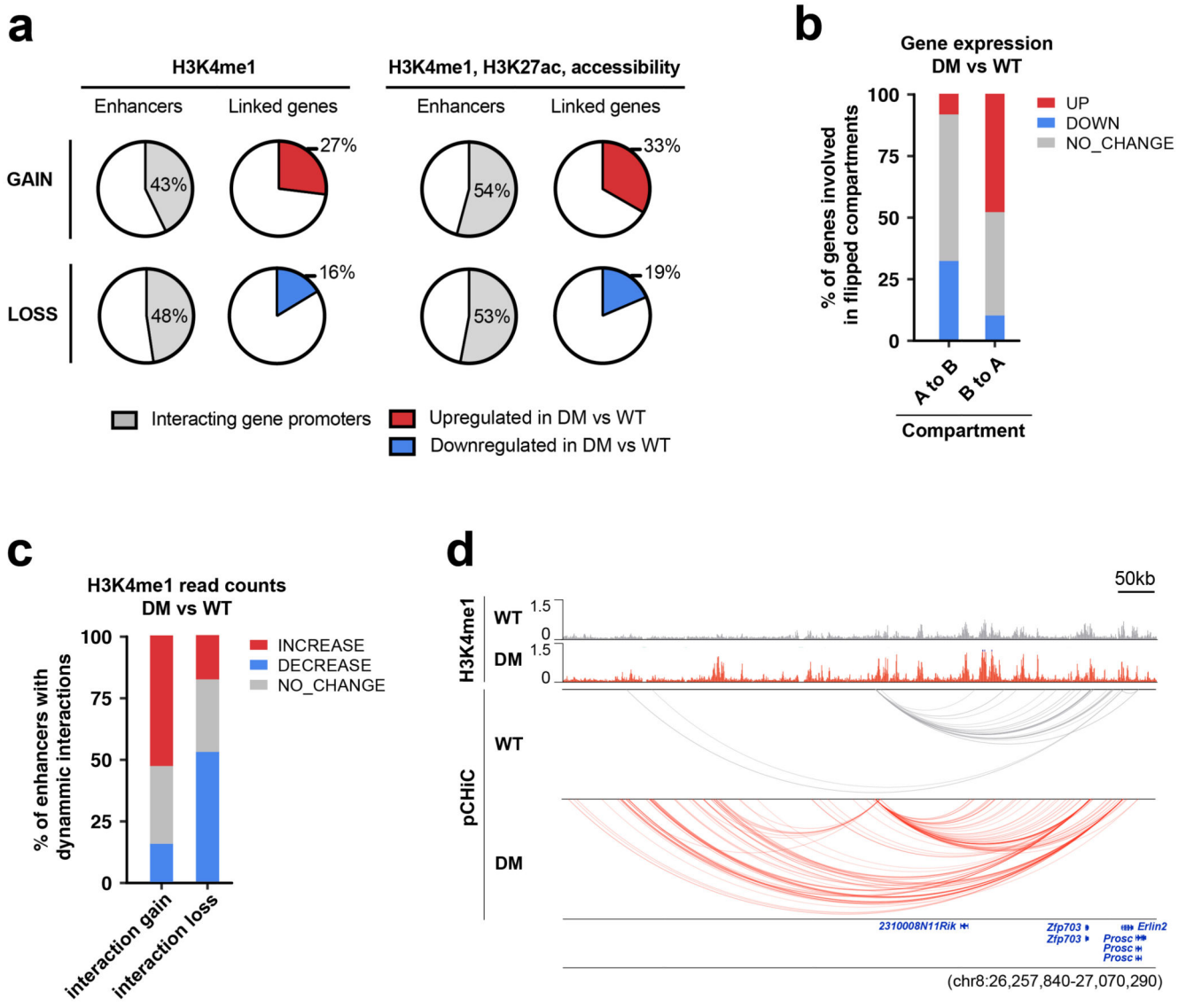
by individual promoters. **c**, Distance of interacting regions from their target promoters. Median distance was shown. **d**, Illustration of chromatin compartments A/B levels at a DM-induced “B to A” flipped region (highlighted) containing the *Igf1* oncogene. **e**. Illustration of chromatin compartments 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.

Europe PMC Funders Author Manuscripts

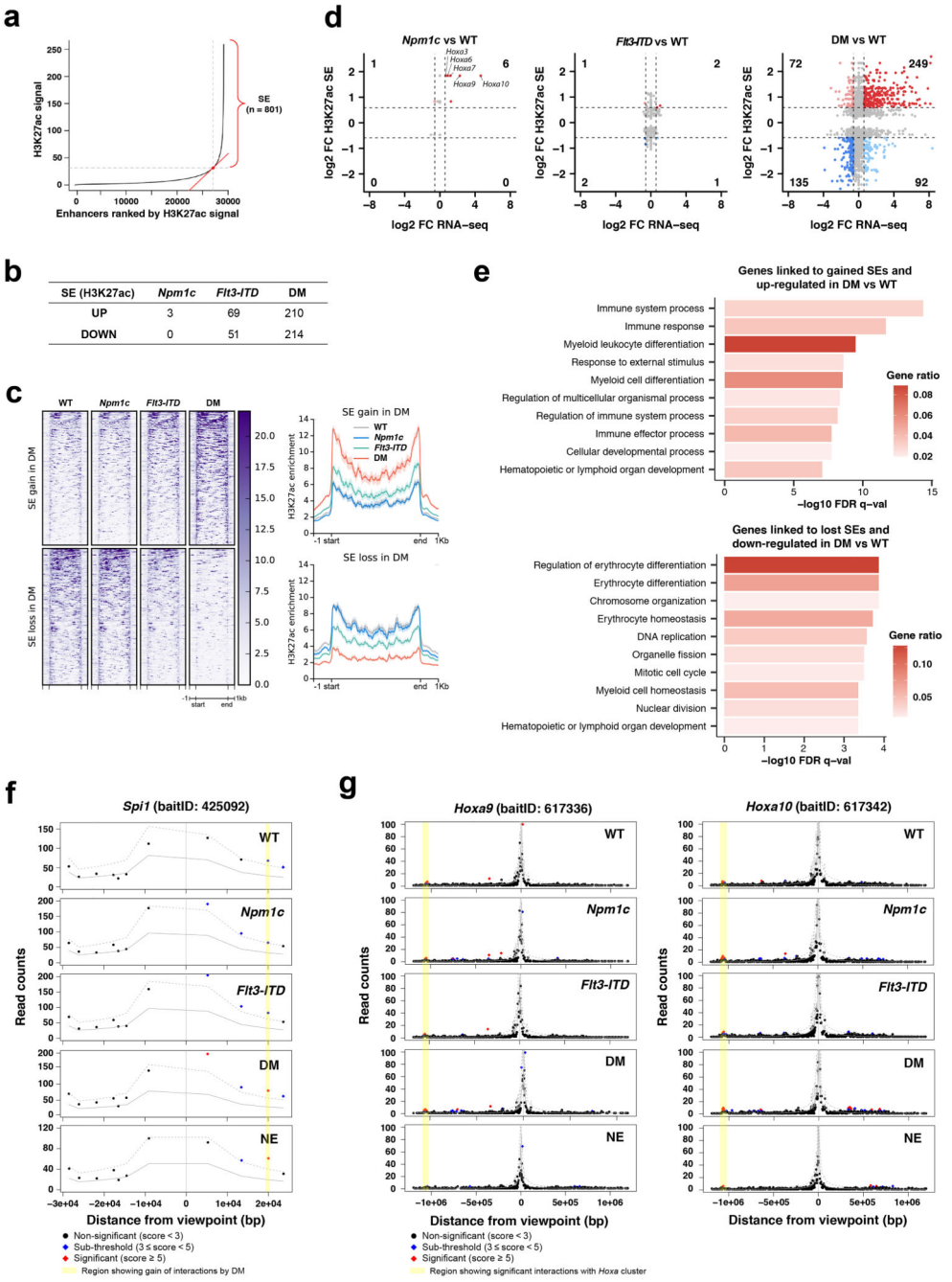
a, Profile plots of chromatin marks and accessibility at accessible regions in each of the 10 clusters (except for Cluster-6 and -10). **b**, *De novo* motifs significantly enriched at Clusters 7-9 chromatin regions. HOMER outputs motifs with target coverage > 10% and ranked by *p* values (one-tailed, 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 ≥ 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 ≥ 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 H3K4m1 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.

Europe PMC Funders Author Manuscripts



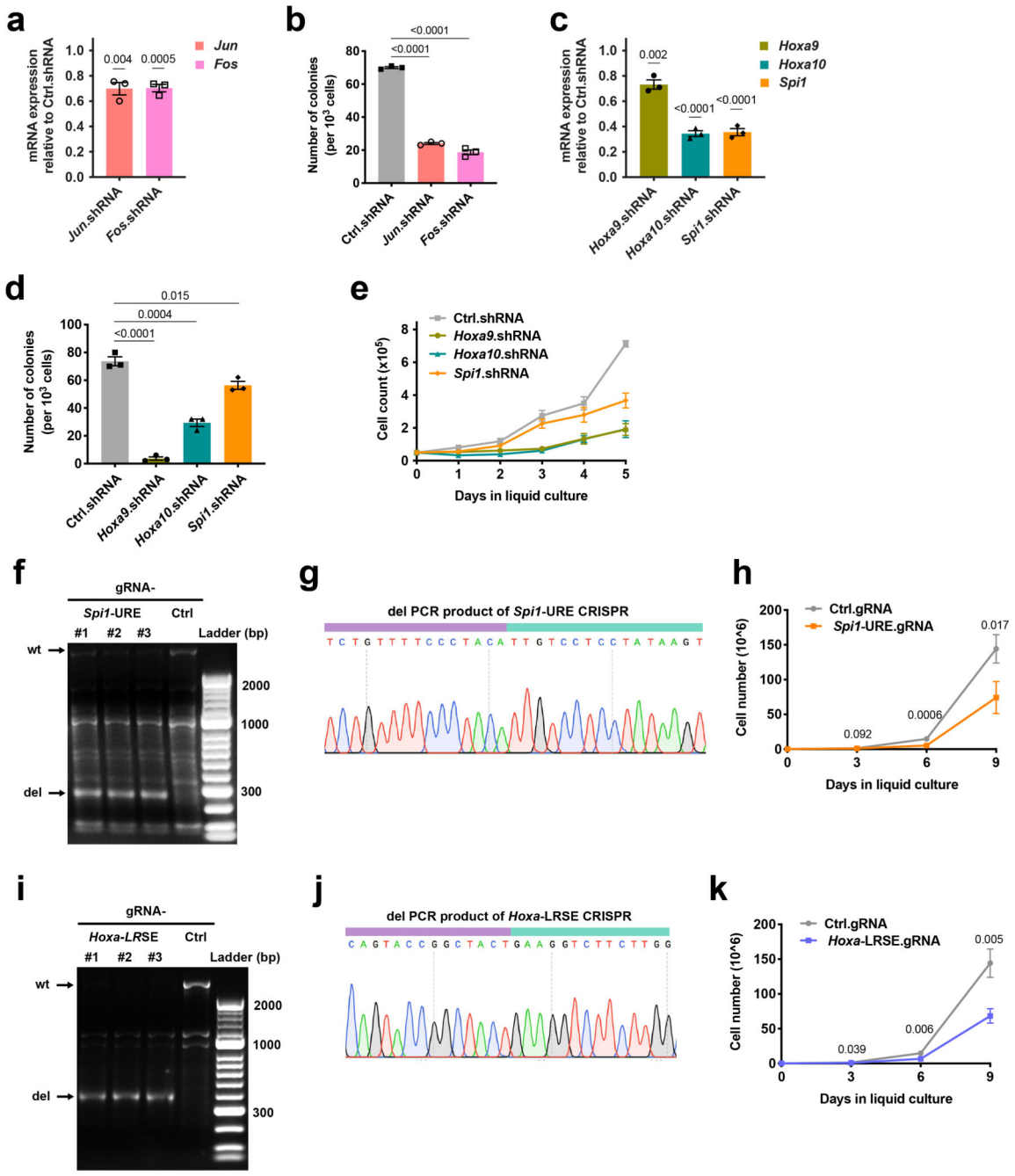
Extended Data Fig. 8. Differential Super-enhancer usage across WT and mutant HSPC.

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

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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.

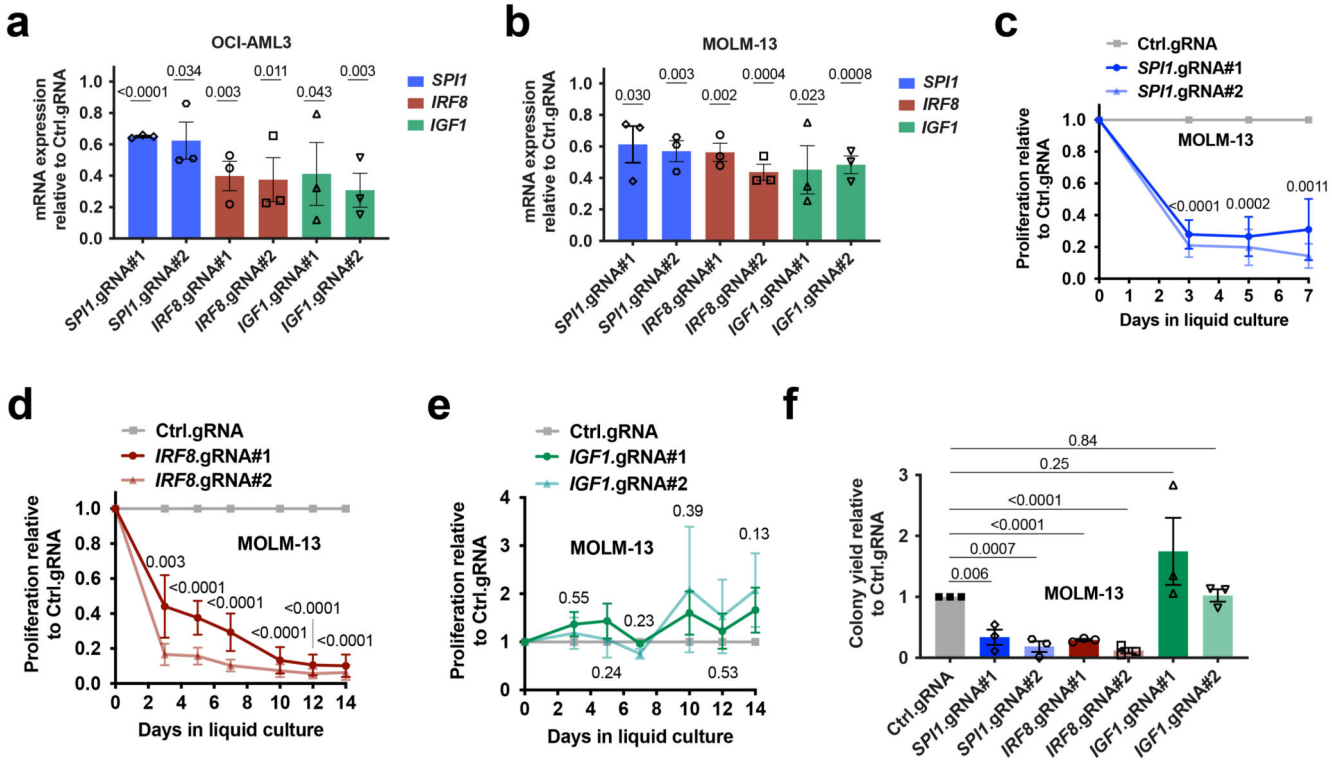


Extended Data Fig. 9. Perturbation of critical cis-regulatory hubs and their target genes abrogates leukemia maintenance in mouse DM leukemia.

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**)

Europe PMC Funders Author Manuscripts

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.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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-R009708-1), 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 (203151/Z/16/Z) and Cambridge Institute for Medical Research (100140/Z/12/Z). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. D.L.-A. is Marie-Skłodowska-Curie International Fellow (886474). D.S. was a postdoctoral fellow of the Mildred-Scheel Organization, German Cancer Aid (111875). P.G. is funded by a Cancer Research UK Advanced Clinician Scientist Fellowship (C57799/A27964) and was previously funded by a Wellcome Trust fellowship (109967/Z/15/Z) and an ASH Global Research Award. G.S.V. is a CRUK Senior Cancer Research Fellow (C22/324/A23015) and work in his laboratory was supported by Blood Cancer UK (17006). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We acknowledge Martin Dawes in the Department of Haematology for the inter-lab communications and organizational assistance. And we thank Dr Maike Paramor in the NGS library facility at 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.

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 re-analysed multiple data sets which are publicly available: the ChIP-seq on BRD4 (ArrayExpress: ERR220396) and H3K27ac (GSM2716711) in OCI-AML3 cells; H3K27ac in Kasumi-1 cells (GSM2212053); H3K4me1 (GSM1816068) 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.

Code availability

All computational analysis is described in the Methods, performed either with the software default parameters and pipelines or with custom code which is available at https://github.com/haiyang-yun/3D_chromatin_in_AML (archived also on Zenodo: https://doi.org/10.5281/zenodo.5009065)[78].

References

1. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006; 7 :29–59. [PubMed: 16719718]
2. Lagha M, Bothma JP, Levine M. Mechanisms of transcriptional precision in animal development. Trends Genet. 2012; 28 :409–416. [PubMed: 22513408]

3. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012; 13 :613–626. [PubMed: 22868264]
4. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019; 20 :437–455. [PubMed: 31086298]
5. Laurenti E, Gottgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature. 2018; 553 :418–426. [PubMed: 29364285]
6. Huang J, et al. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Dev Cell. 2016; 36 :9–23. [PubMed: 26766440]
7. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008; 132 :631–644. [PubMed: 18295580]
8. Cedar H, Bergman Y. Epigenetics of haematopoietic cell development. Nat Rev Immunol. 2011; 11 :478–488. [PubMed: 21660052]
9. Cancer Genome Atlas Research, N. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368 :2059–2074. [PubMed: 23634996]
10. Grimwade D, Ivey A, Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood. 2016; 127 :29–41. [PubMed: 26660431]
11. Papaemmanuil E, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016; 374 :2209–2221. [PubMed: 27276561]
12. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015; 373 :1136–1152. [PubMed: 26376137]
13. Tenen DG. Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer. 2003; 3 :89–101. [PubMed: 12563308]
14. Lee BH, et al. FLT3 mutations confer enhanced proliferation and survival properties to multipotent progenitors in a murine model of chronic myelomonocytic leukemia. Cancer Cell. 2007; 12 :367–380. [PubMed: 17936561]
15. Vassiliou GS, et al. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. Nat Genet. 2011; 43 :470–475. [PubMed: 21441929]
16. Mupo A, et al. A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice. Leukemia. 2013; 27 :1917–1920. [PubMed: 23478666]
17. Buenrostro JD, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10 :1213–1218. [PubMed: 24097267]
18. Mifsud B, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015; 47 :598–606. [PubMed: 25938943]
19. Assi SA, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. Nat Genet. 2019; 51 :151–162. [PubMed: 30420649]
20. Bell O, Tiwari VK, Thoma NH, Schubeler D. Determinants and dynamics of genome accessibility. Nat Rev Genet. 2011; 12 :554–564. [PubMed: 21747402]
21. Snow JW, et al. Context-dependent function of “GATA switch” sites in vivo. Blood. 2011; 117 :4769–4772. [PubMed: 21398579]
22. Yamazaki H, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer Cell. 2014; 25 :415–427. [PubMed: 24703906]
23. Baek S, Goldstein I, Hager GL. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. Cell Rep. 2017; 19 :1710–1722. [PubMed: 28538187]
24. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013; 49 :825–837. [PubMed: 23473601]
25. Lara-Astiaso D, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science. 2014; 345 :943–949. [PubMed: 25103404]
26. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38 :576–589. [PubMed: 20513432]
27. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326 :289–293. [PubMed: 19815776]

28. Tischkowitz M, et al. Bi-allelic silencing of the Fanconi anaemia gene FANCF in acute myeloid leukaemia. Br J Haematol. 2003; 123 :469–471. [PubMed: 14617007]
29. Freire-Pritchett P, et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. Elife. 2017; 6
30. Rubin AJ, et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat Genet. 2017; 49 :1522–1528. [PubMed: 28805829]
31. Ma Q, et al. Super-Enhancer Redistribution as a Mechanism of Broad Gene Dysregulation in Repeatedly Drug-Treated Cancer Cells. Cell Rep. 2020; 31 107532 [PubMed: 32320655]
32. Satija R, et al. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015; 33 :495–502. [PubMed: 25867923]
33. Li Y, et al. Regulation of the PU.1 gene by distal elements. Blood. 2001; 98 :2958–2965. [PubMed: 11698277]
34. Okuno Y, et al. Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. Mol Cell Biol. 2005; 25 :2832–2845. [PubMed: 15767686]
35. Rosenbauer F, et al. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat Genet. 2006; 38 :27–37. [PubMed: 16311598]
36. Will B, et al. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. Nat Med. 2015; 21 :1172–1181. [PubMed: 26343801]
37. Hu G, et al. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Res. 2011; 21 :1650–1658. [PubMed: 21795385]
38. Mazumdar C, et al. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell Stem Cell. 2015; 17 :675–688. [PubMed: 26607380]
39. Dawson MA, et al. Recurrent mutations, including NPM1c, activate a BRD4-dependent core transcriptional program in acute myeloid leukemia. Leukemia. 2014; 28 :311–320. [PubMed: 24220271]
40. Zhao Y, et al. High-Resolution Mapping of RNA Polymerases Identifies Mechanisms of Sensitivity and Resistance to BET Inhibitors in t(8;21) AML. Cell Rep. 2016; 16 :2003–2016. [PubMed: 27498870]
41. Gerlach D, et al. The novel BET bromodomain inhibitor BI 894999 represses super-enhancer-associated transcription and synergizes with CDK9 inhibition in AML. Oncogene. 2018; 37 :2687–2701. [PubMed: 29491412]
42. Dovey OM, et al. Molecular synergy underlies the co-occurrence patterns and phenotype of NPM1-mutant acute myeloid leukemia. Blood. 2017; 130 :1911–1922. [PubMed: 28835438]
43. Foletta VC, Segal DH, Cohen DR. Transcriptional regulation in the immune system: all roads lead to AP-1. J Leukoc Biol. 1998; 63 :139–152. [PubMed: 9468273]
44. Madrigal P, Alasoo K. AP-1 Takes Centre Stage in Enhancer Chromatin Dynamics. Trends Cell Biol. 2018; 28 :509–511. [PubMed: 29778529]
45. Carotta S, Wu L, Nutt SL. Surprising new roles for PU.1 in the adaptive immune response. Immunol Rev. 2010; 238 :63–75. [PubMed: 20969585]
46. Voon DC, Hor YT, Ito Y. The RUNX complex: reaching beyond haematopoiesis into immunity. Immunology. 2015; 146 :523–536. [PubMed: 26399680]
47. Cauchy P, et al. Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. Cell Rep. 2015; 12 :821–836. [PubMed: 26212328]
48. Gadad SS, et al. The multifunctional protein nucleophosmin (NPM1) is a human linker histone H1 chaperone. Biochemistry. 2011; 50 :2780–2789. [PubMed: 21425800]
49. Nemeth A, et al. Initial genomics of the human nucleolus. PLoS Genet. 2010; 6 e1000889 [PubMed: 20361057]
50. van Koningsbruggen S, et al. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell. 2010; 21 :3735–3748. [PubMed: 20826608]
51. Yan J, et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. Cell Res. 2018; 28 :204–220. [PubMed: 29313530]

52. Tzelepis K, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 2016; 17 :1193–1205. [PubMed: 27760321]
53. Wang T, et al. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell. 2017; 168 :890–903. e815 [PubMed: 28162770]
54. McKenzie MD, et al. Interconversion between Tumorigenic and Differentiated States in Acute Myeloid Leukemia. Cell Stem Cell. 2019; 25 :258–272. e259 [PubMed: 31374198]
55. Rosenbauer F, Koschmieder S, Steidl U, Tenen DG. Effect of transcription-factor concentrations on leukemic stem cells. Blood. 2005; 106 :1519–1524. [PubMed: 15914558]
56. Bonadies N, Pabst T, Mueller BU. Heterozygous deletion of the PU.1 locus in human AML. Blood. 2010; 115 :331–334. [PubMed: 19890096]
57. Huang G, et al. The ability of MLL to bind RUNX1 and methylate H3K4 at PU.1 regulatory regions is impaired by MDS/AML-associated RUNX1/AML1 mutations. Blood. 2011; 118 :6544–6552. [PubMed: 22012064]
58. Vangala RK, et al. The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. Blood. 2003; 101 :270–277. [PubMed: 12393465]
59. Bell CC, et al. Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. Nat Commun. 2019; 10 :2723. [PubMed: 31222014]
60. Cusan M, et al. LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBPalpha-dependent enhancers in AML. Blood. 2018; 131 :1730–1742. [PubMed: 29453291]
61. Gozdecka M, et al. UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat Genet. 2018; 50 :883–894. [PubMed: 29736013]
62. Behre G, et al. Meropenem monotherapy versus combination therapy with ceftazidime and amikacin for empirical treatment of febrile neutropenic patients. Ann Hematol. 1998; 76 :73–80. [PubMed: 9540761]
63. Grondin B, et al. c-Jun homodimers can function as a context-specific coactivator. Mol Cell Biol. 2007; 27 :2919–2933. [PubMed: 17283046]
64. Huang Y, et al. Identification and characterization of Hoxa9 binding sites in hematopoietic cells. Blood. 2012; 119 :388–398. [PubMed: 22072553]
65. Zhou J, et al. PU.1 is essential for MLL leukemia partially via crosstalk with the MEIS/HOX pathway. Leukemia. 2014; 28 :1436–1448. [PubMed: 24445817]
66. Neijts R, Deschamps J. At the base of colinear Hox gene expression: cis-features and trans-factors orchestrating the initial phase of Hox cluster activation. Dev Biol. 2017; 428 :293–299. [PubMed: 28728680]
67. Noordermeer D, et al. The dynamic architecture of Hox gene clusters. Science. 2011; 334 :222–225. [PubMed: 21998387]
68. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29 :15–21. [PubMed: 23104886]
69. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26 :139–140. [PubMed: 19910308]
70. Conesa A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016; 17 :13. [PubMed: 26813401]
71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15 :550. [PubMed: 25516281]
72. Bailey T, et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol. 2013; 9 e1003326 [PubMed: 24244136]
73. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10 :R25. [PubMed: 19261174]
74. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 481 :389–393. [PubMed: 22217937]
75. Ramirez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016; 44 :W160–165. [PubMed: 27079975]

76. Wingett S, et al. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 2015; 4 :1310. [PubMed: 26835000]

77. Cairns J, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 2016; 17 :127. [PubMed: 27306882]

78. Yun, Haiyang. haiyang-yun/3D_chromatin_in_AML. 2021.

Europe PMC Funders Author Manuscripts

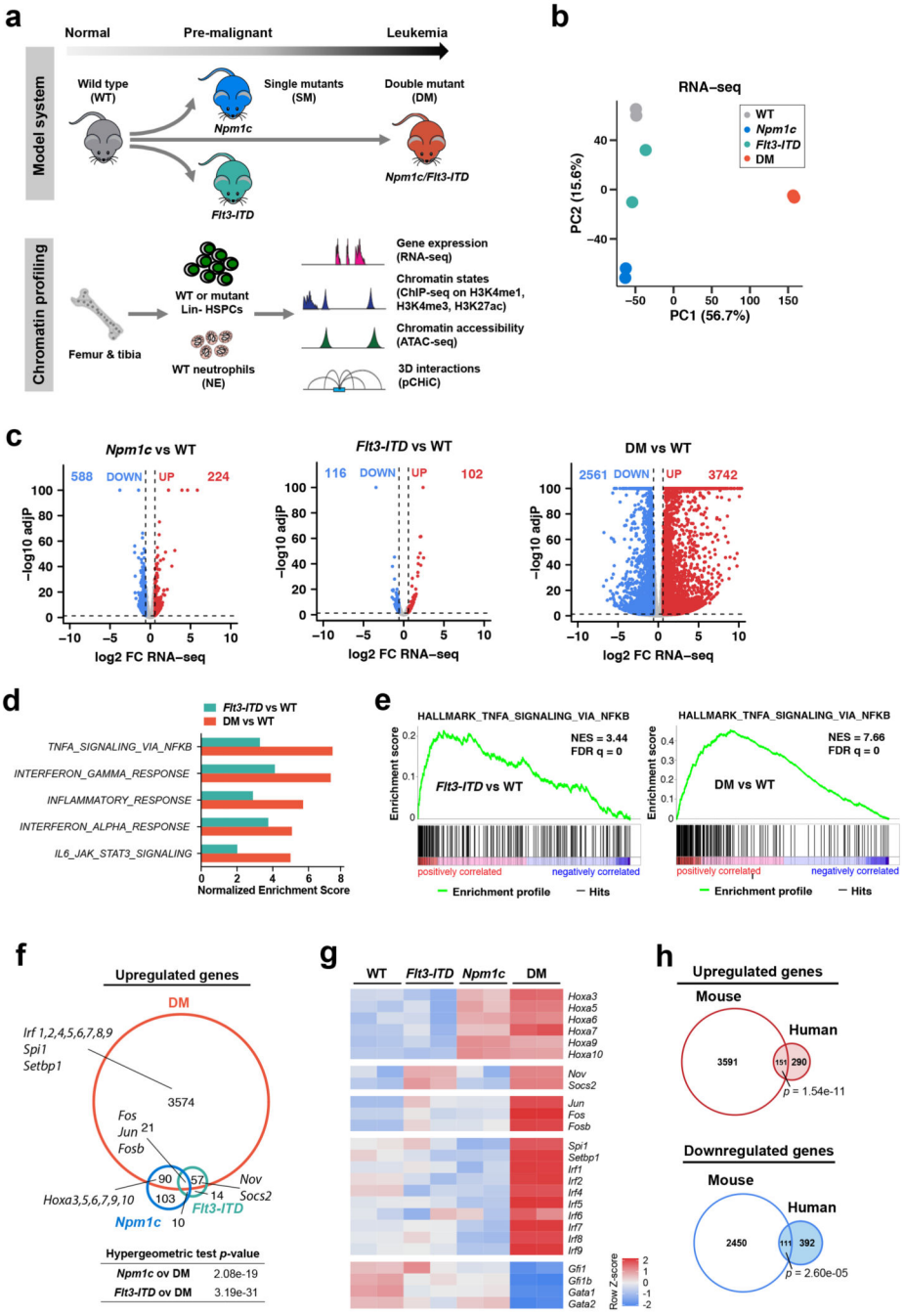


Figure 1. Murine models show transcriptional synergy of AML mutations.

a, Schematic of overall experimental design. **b**, Principal component analysis (PCA) of mRNA expression data of 16,771 protein-coding genes. **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 fold change (FC) ≥ 1.5. **d**, Immune response related Hallmark gene sets from GSEA analysis of differential gene expression in *Flt3-ITD* or DM HSPC.

Europe PMC Funders Author Manuscripts

e, GSEA enrichment plots showing gene set of TNFA signalling via NFKB in *Flt3-ITD* or DM HSPC. NES, normalised 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. **g,** Heatmap of normalised 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 2-fold change) in human AML and mouse DM leukemia. Hypergeometric test *p* values (one-tailed, not multiple testing corrected) are shown.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

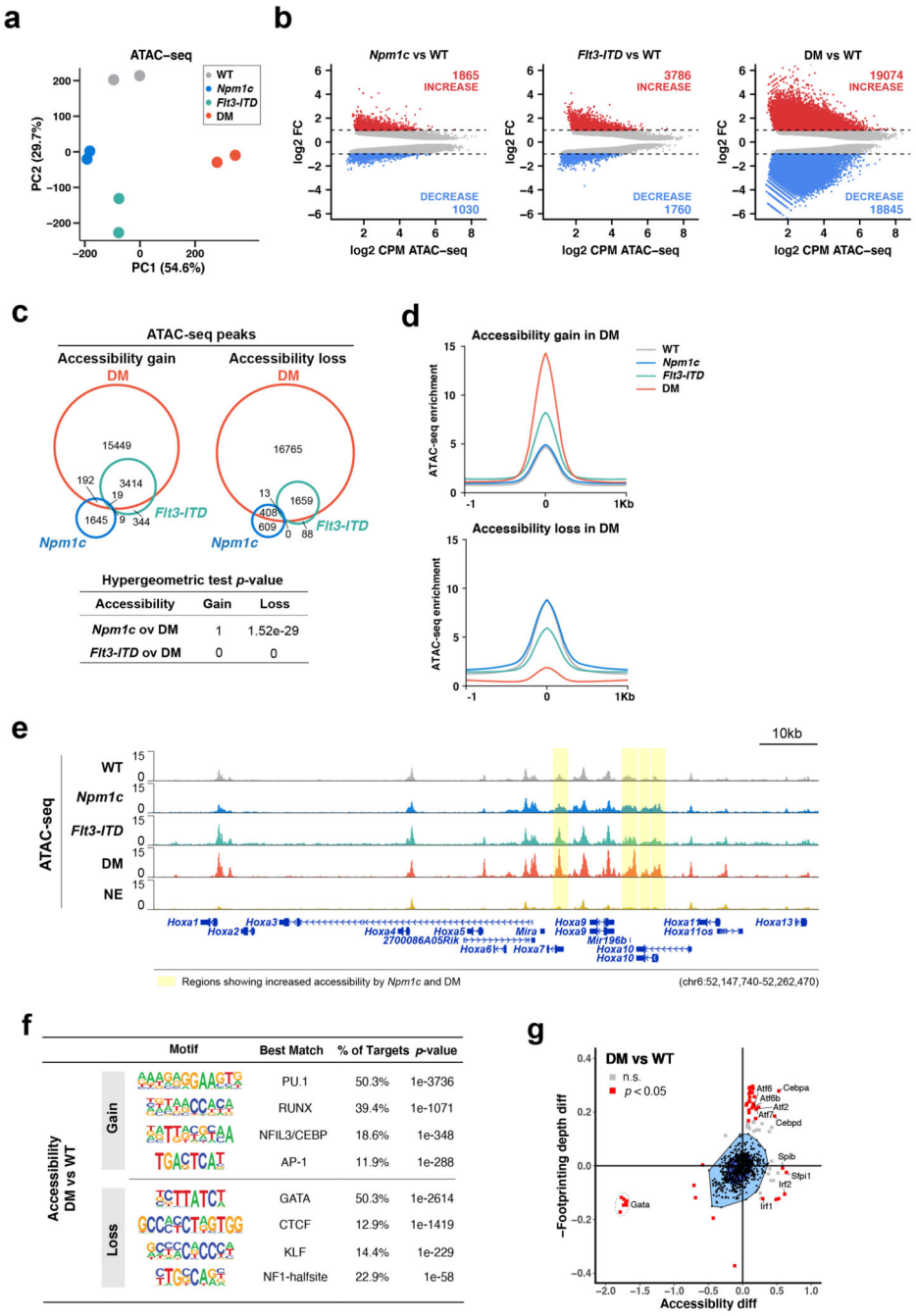


Figure 2. Both *FLT3-ITD* and *Npm1c* alter chromatin accessibility.

a, PCA of ATAC-seq signals across all four cellular states. Counts per million reads were used for the analysis. **b**, MA plots showing ATAC-seq peaks with significantly differential accessibility in mutant vs WT HSPC. Significant increase or decrease were determined by edgeR with setting FDR (two-tailed and multiple testing corrected) < 0.05 and absolute FC ≥ 1.5. **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. **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 wildtype neutrophils (NE). Regions showing increased accessibility by *Npmc1* and DM compared with WT were 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 vs WT HSPC. The data points within bag and fence area include 50% and > 97% of the population, respectively, are not significant. 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.

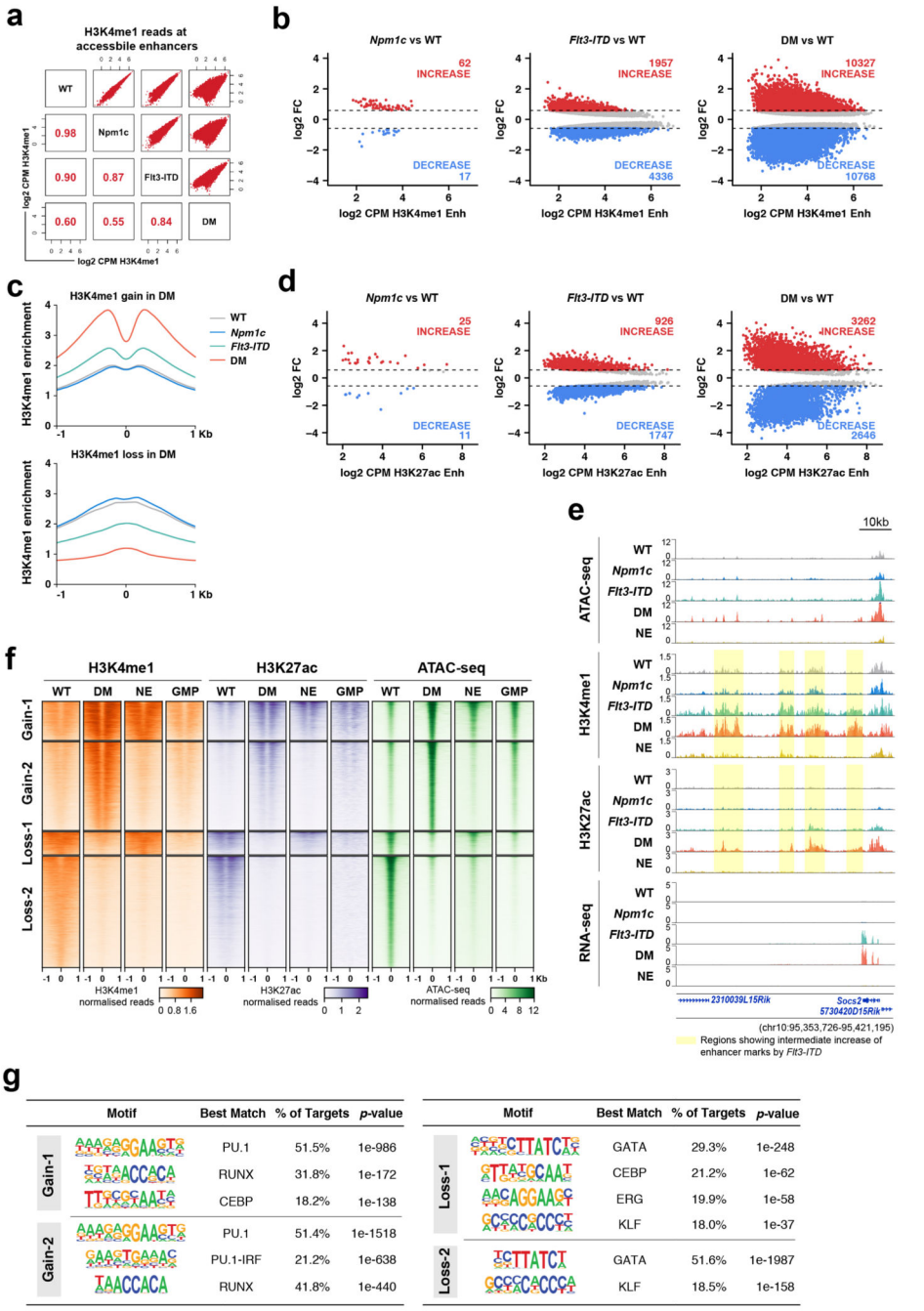


Figure 3. *Flt3-ITD* but not *Npm1c* remodels the enhancer landscape.

a, Correlation matrix of H3K4me1 reads at accessible enhancers in mutant vs WT HSPC. Counts per million (CPM) reads were log2 transformed and Pearson correlation coefficients are shown. **b** and **d**, MA plots showing accessible enhancers with significantly differential H3K4me1 (**b**) or H3K27ac (**d**) levels in mutant vs WT HSPC. Significant increase or decrease were determined by edgeR with setting FDR (two-tailed and multiple testing corrected) < 0.05 and absolute FC ≥ 1.5. **c**, Profile plots of H3K4me1 enrichment scores at accessible enhancers (1kb ± ATAC-seq peak summit) with gain (upper panel) or loss

(lower panel) of H3K4me1 by DM across four cellular states. **e**, Chromatin profiles of *Socs2* gene and its proximal region with enhancer marks, accessibility and RNA-seq reads in all four HSPC and wildtype neutrophils (NE). Regions showing intermediate increase of enhancer marks by *Flt3-ITD* were highlighted. **f**, Heatmaps of the H3K4me1, H3K27ac and ATAC-seq profiles at DM gained and lost enhancers across WT and DM HSPC, wildtype neutrophils (NE) and granulocyte/macrophage progenitors (GMP). Group Gain-2 demonstrates “leukemia-specific” changes. Peaks were ranked by average enrichment across all samples. **g**, *De novo* motifs significantly enriched at individual enhancer groups. HOMER outputs motifs with target coverage > 10% and ranked by *p* values (one-tailed, not multiple testing corrected).

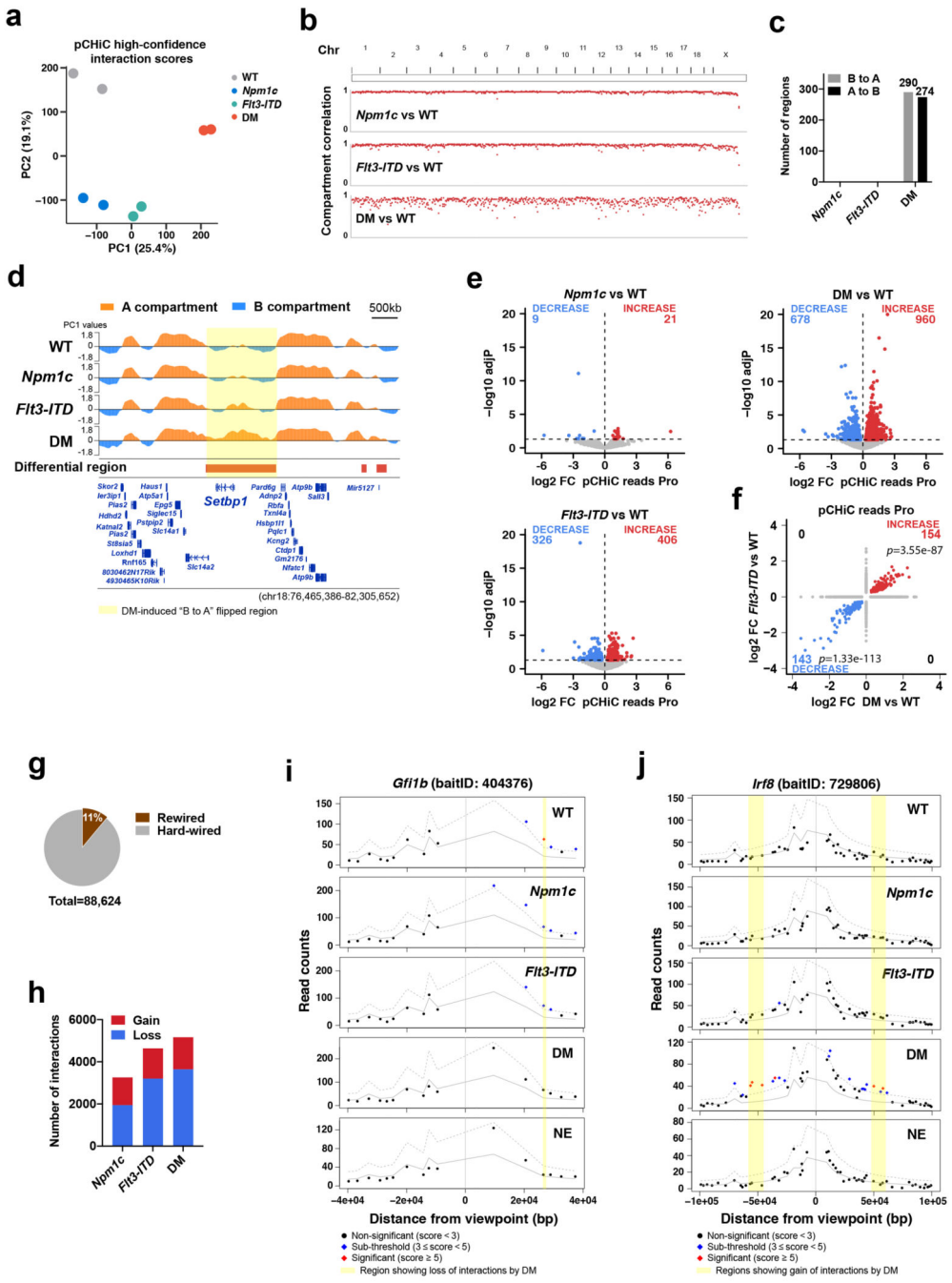


Figure 4. Leukemia programs utilize novel and existing 3D contacts.

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 compartment and the correlation of their PC1 values. **c**, Numbers of flipped chromatin compartments upon mutations. **d**, 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

Europe PMC Funders Author Manuscripts

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 to WT HSPC). **i** and **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; Grey 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 vs WT HSPC.

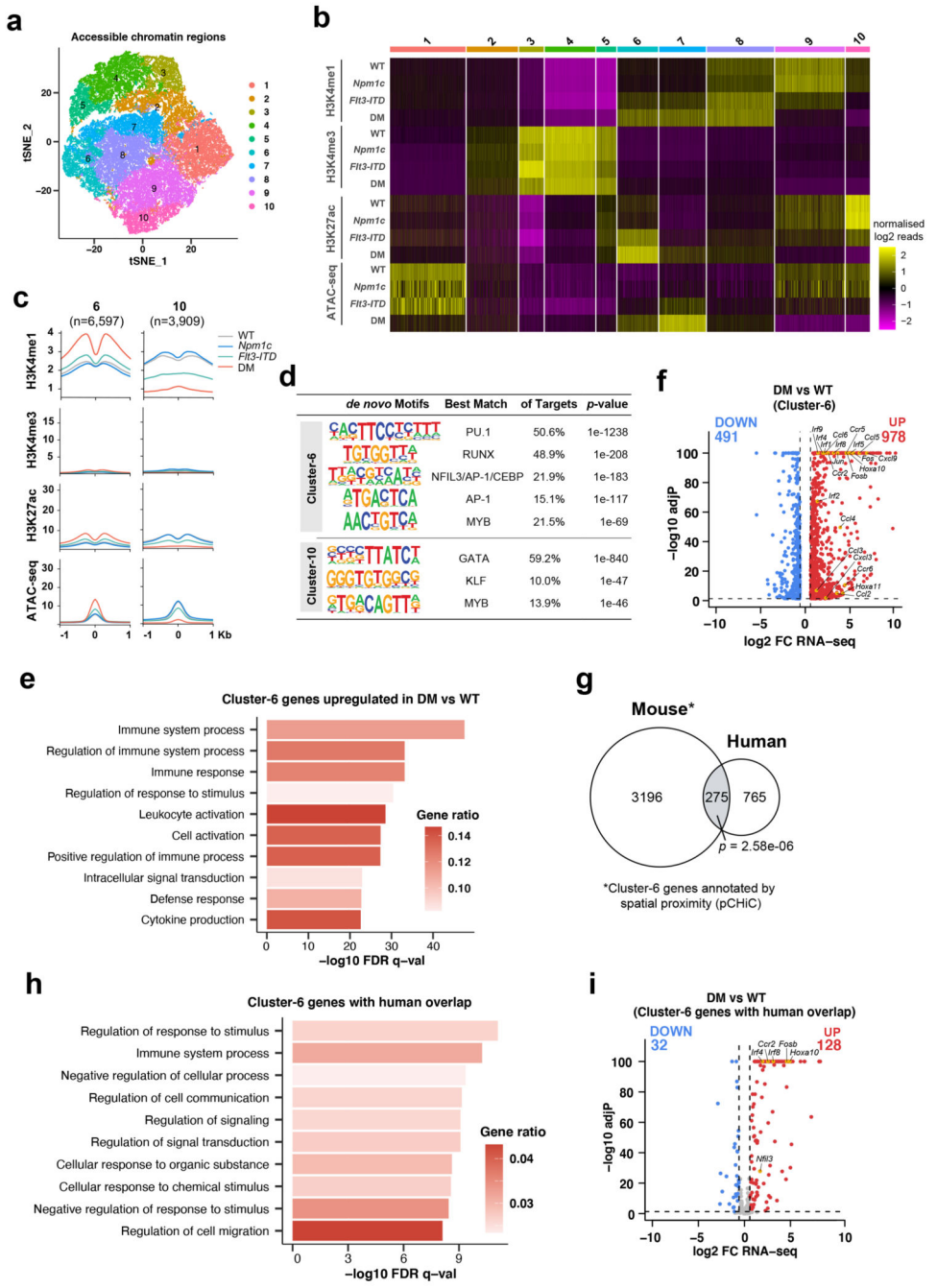


Figure 5. Integrated analyses identifies critical regulatory networks.

a, tSNE plot showing Seurat-guided clustering of all accessible regions (±1kb from ATAC-seq peak summit) based on chromatin profiles (H3K4me1, H3K4me3 and H3K27ac, normalised read counts) and accessibility (normalised ATAC-seq reads) across WT and mutant HSPC. **b**, Heatmap presenting levels of chromatin modifications and accessibility at chromatin regions from the 10 clusters. **c**, Profile plots of chromatin marks and accessibility at accessible regions in Cluster-6 and -10 regions across all four cellular states. **d**, *De novo* motifs significantly enriched at Cluster-6 and -10 chromatin regions. HOMER outputs

motifs with target coverage > 10% and ranked by *p* values (one-tailed, not multiple testing corrected). **e**, Gene ontology (GO) analysis of genes that are linked to the chromatin regions in Cluster-6 by pCHiC profiles and upregulated during leukemia induction. **f**, Volcano plot showing differential expression of Cluster-6 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 ≥ 1.5. **g**, Overlap of genes linked to chromatin regions with increased accessibility in *NPM1c*/*FLT3-ITD* leukemia in mouse (Cluster-6 genes annotated by pCHiC profiles) and human[19]. Hypergeometric test p values (one-tailed, not multiple testing corrected) are shown. **h** and **i**, GO analysis (**h**) and differential expression (**i**) of Cluster-6 genes with human overlap. Up- or downregulation were defined by setting adjP (two-tailed and multiple testing corrected) < 0.05 and absolute FC ≥ 1.5.

Europe PMC Funders Author Manuscripts

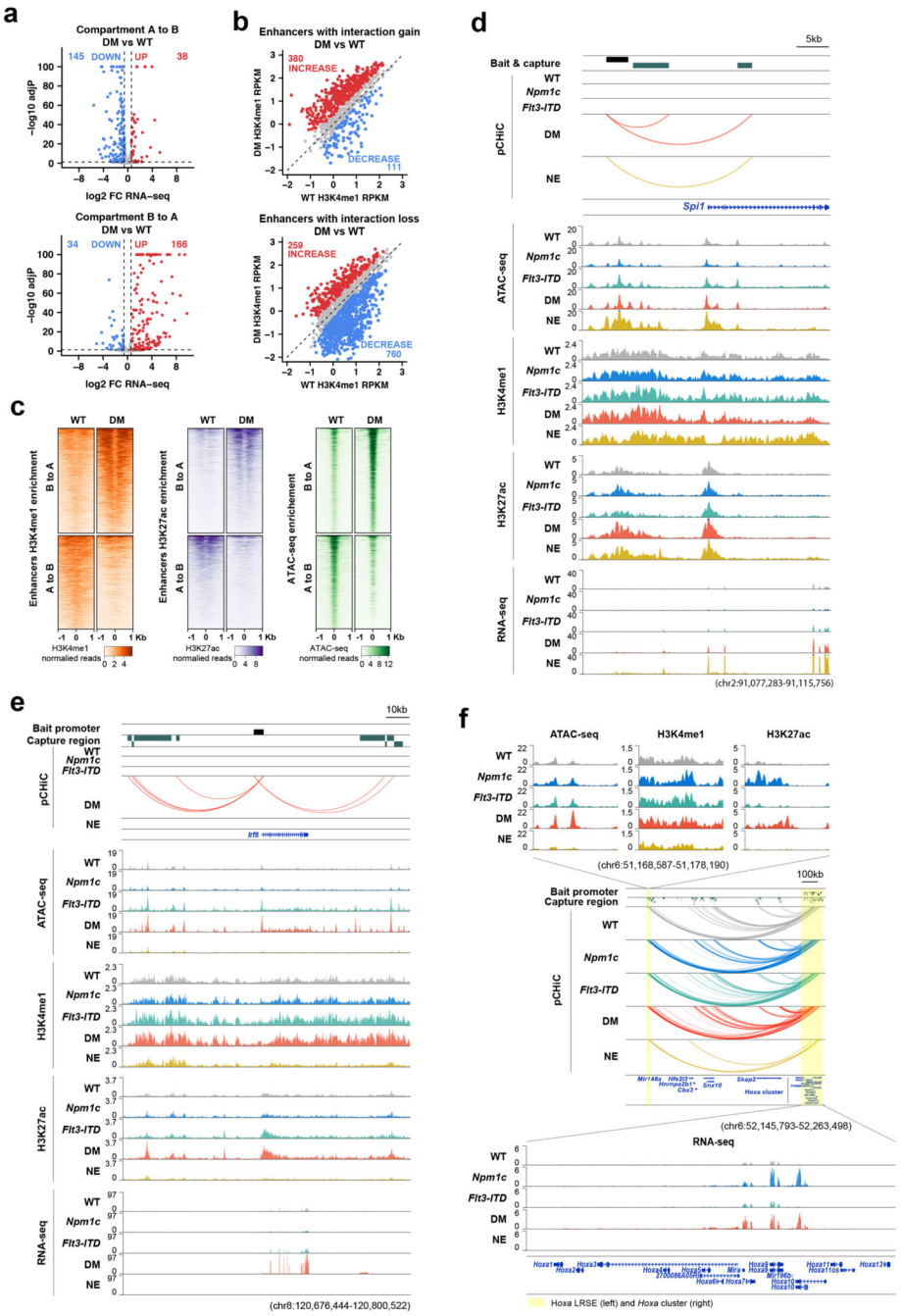


Figure 6. Integrated analyses identifies critical transcriptional networks nodes.

a, Differential expression of genes involved in flipped chromatin compartments in DM vs 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 vs 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 wildtype neutrophils. Significant chromatin interactions were defined by CHiCAGO score ≥ 5 and are represented by arcs.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

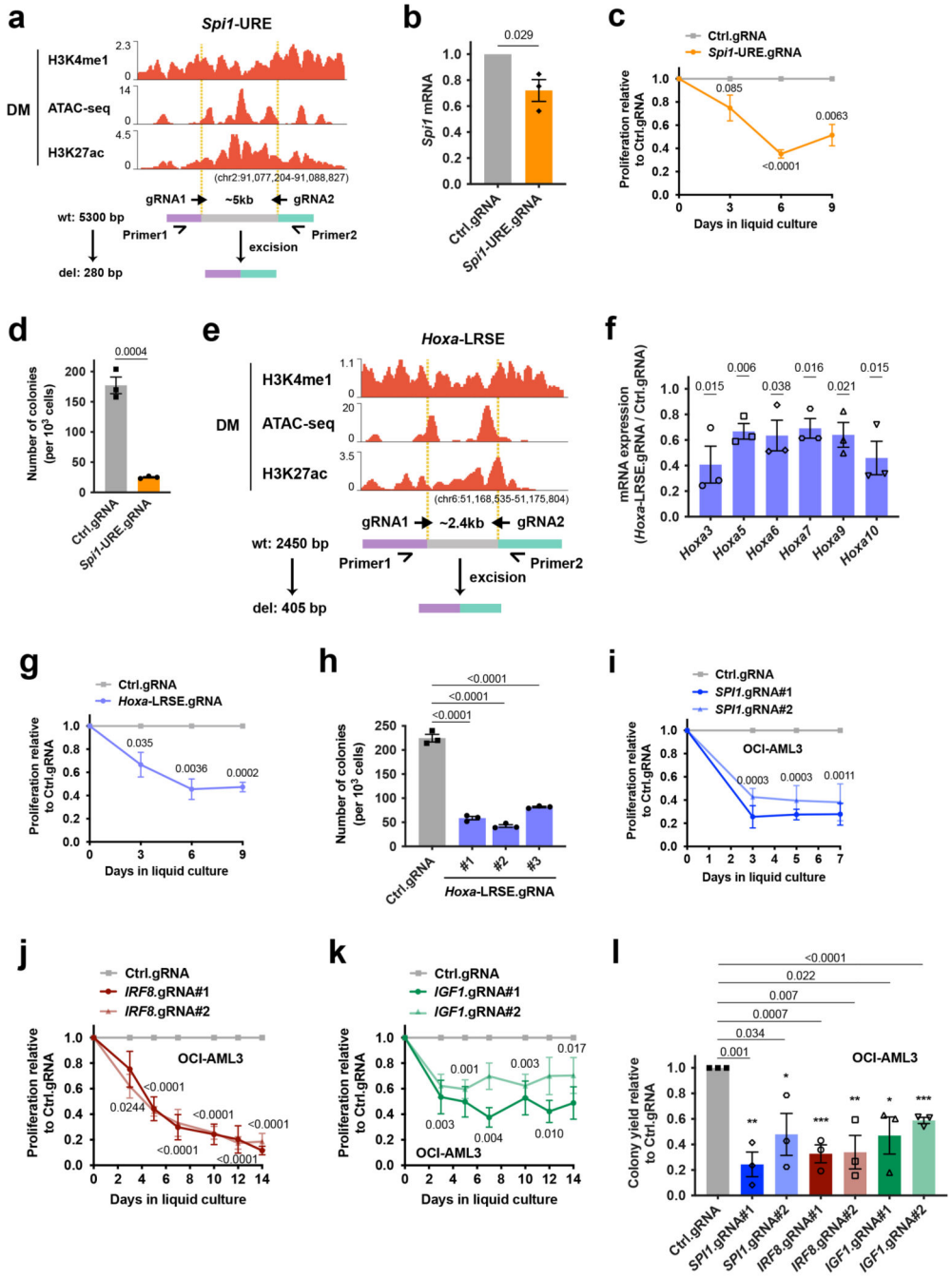


Figure 7. Network perturbation abrogates leukemia maintenance.

a, Experimental strategy targeting *Spi1*-URE for deletion using CRISPR-Cas9 plus dual guide RNAs (gRNA) flanking its ~5kb central region. Primer 1 and 2 are oligos for PCR confirmation of deletion. **b**, *Spi1* mRNA expression detected by RT-qPCR in DM-Cas9 cells expressing *Spi1*-URE gRNAs relative to control gRNAs (n=3 independent experiments). **c** and **d,** *Ex vivo* proliferation (**c**) and colony-forming unit (CFU) assays (**d**) of DM-Cas9 cells expressing *Spi1*-URE gRNAs relative to control gRNAs (n=3 independent experiments). **e**, Experimental strategy targeting *Hoxa*-LRSE for deletion using

CRISPR-Cas9 plus dual gRNAs flanking its ~2.4kb central region. **f**, mRNA expression of multiple *Hoxa* genes detected in DM-Cas9 cells expressing *Hoxa*-LRSE gRNAs relative to control gRNAs (n=3 independent experiments). **g**, *Ex vivo* proliferation of DM-Cas9 cells expressing *Hoxa*-LRSE gRNAs relative to control gRNAs (n=3 independent experiments). **h**, CFU assays of DM-Cas9 cells expressing *Hoxa*-LRSE gRNAs or control gRNAs (n=3 independent experiments). # indicates three independent single cell derived clones. **i-k**, Cell proliferation of OCI-AML3 cells expressing human gRNAs targeting *SPI1* (**i**), *IRF8* (**j**) and *IGF1* (**k**) relative to control gRNAs in liquid culture (n=3, 4, 4 independent experiments, respectively). # indicates two independent gRNAs. **l**, CFU assays of OCI-AML3 cells expressing human gRNAs targeting *SPI1*, *IRF8* and *IGF1* relative to control gRNAs in methylcellulose culture (n=3 independent experiments). 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) were shown; error bars represent mean and standard error of mean.

Europe PMC Funders Author Manuscripts