In vivo screening characterizes chromatin factor functions during normal and malignant hematopoiesis

David Lara-Astiaso[1,2]*†, Ainhoa Goñi-Salaverri[3]†, Julen Mendieta-Esteban[3]†, Nisha Narayan[1,2], Cynthia Del Valle[3], Torsten Gross[4], George Giotopoulos[1,2], Tumas Beinortas[1,2], Mar Navarro-Alonso[3], Laura Pilar Aguado-Alvaro[3], Jon Zazpe[3], Francesco Marchese[3], Natalia Torrea[3], Isabel A Calvo[3], Cecile K Lopez[1,2], Diego Alignani[3], Aitziber Lopez[3], Borja Saez[3], Jake P Taylor-King[4], Felipe Prosper[3], Nikolaus Fortelny[5]#*, Brian JP Huntly[1,2]#*

[1] Department of Haematology, University of Cambridge; Cambridge, CB2 0AW, UK

[2] Wellcome Trust-MRC Cambridge Stem Cell Institute; Cambridge, CB2 0AW, UK

[3] Centre for Applied Medical Research, University of Navarra; 31008 Pamplona, Spain

[4] Relation Therapeutics; London, UK, W1B 4BD

[5]Department of Biosciences & Medical Biology, University of Salzburg; 5020 Salzburg, Austria

*Corresponding authors.

†These authors contributed equally.

# These authors jointly supervised this work.

Correspondence to: dl627@cam.ac.uk, nikolaus.fortelny@plus.ac.at or bjph2@cam.ac.uk

ABSTRACT

Cellular differentiation requires extensive alterations in chromatin structure and function, elicited by the coordinated action of chromatin and transcription factors (TFs). By contrast with Transcription Factors (TFs), the roles of chromatin-factors in differentiation have not been systematically characterized. Here we have combined bulk *ex vivo* and single-cell *in vivo* CRISPR screens to characterize the role of chromatin factor families in hematopoiesis. We uncover marked lineage specificities for 142 chromatin-factors revealing functional diversity among related chromatin-factors (i.e. BAF-subcomplexes) as well as shared roles for unrelated repressive complexes that restrain excessive myeloid differentiation. Using epigenetic profiling, we identify functional interactions between lineage-determining TFs and several chromatin-factors that explain their lineage dependencies. Studying chromatin-factor functions in leukemia, we show that leukemia cells corrupt homeostatic chromatin-factor roles to block differentiation trajectories by engaging in specific chromatin-factor-TF interactions that suggest therapeutic avenues. Together,

our work elucidates the lineage-determining properties of chromatin-factors across normal and malignant hematopoiesis.

MAIN TEXT

Introduction

Cell-fate decisions are governed by the coordinated activities of Transcription- (TFs) and Chromatin-Factors, which together form gene regulatory complexes (GRC), to orchestrate tissue-specific transcriptional patterns and drive cellular phenotypes[1]. The widescale description of TF binding and its relationship to chromatin accessibility and gene expression, obtained from epigenomic and transcriptomic maps derived across multiple developmental processes, have provided us with a highly developed understanding of the instructional role for TF in governing cell-fates[2,3]. Conversely, although the role of individual chromatin-factors, particularly those mutated across malignancies[4], are being elucidated, we still lack a global understanding of chromatin-factor functions in differentiation. Specifically, whether chromatin-factors have specific or redundant roles during lineage differentiation, how they mechanistically interact with individual TFs, and the identity of these TF-chromatin-factor interactions remain unresolved questions.

We have chosen to address these fundamental questions in the exemplar process of hematopoiesis, where multiple mature cells with diverse specific functions derive from a single self-renewing cell type, the hematopoietic stem cell (HSC)[5]. The study of hematopoiesis benefits from a comprehensive cellular blueprint[3], which describes normal differentiation and malignant transformation, and a well-annotated molecular blueprint[2,3,4], including detailed single-cell transcriptional landscapes and comprehensive maps of TF-activity derived from epigenomic and *in vivo* loss-of-function studies across hematopoietic lineages[6-11]. In addition, the importance of both chromatin-factors and TFs has been further emphasized by recent studies, which have documented mutations that alter the function of TFs and chromatin-factors to be highly recurrent and almost uniform in hematological malignancies such as acute myeloid leukemia (AML)[12].

Understanding chromatin regulation in normal and malignant hematopoiesis is key for harnessing regenerative therapeutics to restore damaged bone marrow function and in the treatment of hematological malignancies. However, many unanswered fundamental questions remain,

including: What are the key interacting components (chromatin-factors and TFs) of the GRCs that orchestrate lineage differentiation? Does their composition change during differentiation? Do chromatin-factors have specificity for lineage determination and, if so, is this dependent on TF recruitment to target loci? What crosstalk occurs within the components of individual GRC or between different GRCs? Which of these mechanisms are corrupted in leukemia and does this drive the induction and/or maintenance of the disease? In this report, we have addressed many of these fundamental questions.

Results

Functional screens of Chromatin Factors in hematopoiesis

To interrogate the roles of chromatin-factors in regulating normal hematopoiesis, we developed 4 screening platforms coupling cytokine-instructed differentiation of primary hematopoietic progenitors with FACS readouts to study key lineage transitions (Fig. 1a,b): 1) self-renewal vs differentiation, 2) branching between myeloid and mega-erythroid lineages, 3) myeloid differentiation of multipotent progenitors and, 4) myeloid differentiation of myeloid-primed progenitors (GMPs). Cross-referencing the expression profiles of each readout population to existing hematopoietic expression, we demonstrated fidelity with their *in vivo* counterparts (Extended Data Fig. 1a,b). Next, we generated a CRISPR library targeting 680 genes, including the vast majority of chromatin-factors expressed by myeloid and mega-erythroid lineages (Supplementary Table 1-2). We then delivered our library to both Cas9 (GFP+) and non-Cas9 (GFP-) progenitors, co-cultured them throughout our differentiation conditions, sorted "readout" populations based on surface markers, and quantified their sgRNA distributions. Next, we calculated a lineage score for each CF by analyzing the differences in sgRNA content between populations, using the non-Cas9 distributions as a background. This analysis identified 142 chromatin-factors with significant lineage scores in any of the four lineage transitions (Fig. 1c,d, Extended Data 1e). Finally, replicate screens for the strongest 200 chromatin-factors demonstrated high correlation between replicates, indicating reproducible methodology (Extended Data Fig. 1d).

Examination of the lineage scores revealed a high degree of phenocopy amongst factors belonging to the same complex, but also antagonistic behavior for specific CF families (Fig. 1c,d, Extended Data Fig. 1e). For instance, several members of the Cohesin and Mediator complexes (*Stag2*,

Med20) were required to elicit differentiation towards Myeloid or Mega-erythroid fates, confirming dynamic chromatin looping as a general requirement for differentiation[13,14]. However, factors associated with the RNA elongation machinery (*Phf5a*, *Ash1l*) operated to preserve progenitor multipotency. As reported in other systems, H3K4 methyltransferases and chromatin remodelers showed high functional diversity[15]. MLL4 factors (*Kmt2d, Kdm6a*) regulated progenitor identities and early Myeloid priming, while SET1 components were required for differentiation and Erythroid priming. BAF members[16] behaved predominantly as pro-Myeloid regulators, but NuRD and ISWI factors[17] facilitated Mega-erythroid fates. Finally, and in contrast, certain repressive complexes demonstrated functional homogeneity, where heterochromatin (*Setdb1*, *Cbx3*), histone deacetylases (*Hdac1* and *Hdac3*) and coREST[20-21] members all functioned as Myeloid repressors. Validating these results, the screen-based phenotypes of 10 individual chromatin-factor-KOs were confirmed by analyzing their *ex vivo* differentiation patterns (Fig. 1e, Extended Data Fig. 2a-d).

Collectively, these findings highlight significant functional diversity within chromatin complexes suggesting that specific chromatin-factor subcomplexes work at different branches and stages along differentiation trajectories.

Chromatin Factor roles during in vivo hematopoiesis

Next, we used Perturb-seq to explore the functional diversity of 80 factors in their proper physiological context, *in vivo* hematopoiesis, at single-cell resolution. This list comprises 60 chromatin-factors with strong dependencies and, as comparators to provide known functional anchors onto hematopoiesis, 20 lineage-specific TFs were also included.

Experimentally, multipotent progenitors (LSKs) were transduced with targeted libraries before transplantation into sublethally irradiated mice. Thereafter, at 2-weeks post-transplant, we isolated their progeny (Lin- and Lin+/ckit+ cells) and used Perturb-seq to jointly measure their transcriptomes and chromatin-factor perturbations (Fig.2a). This approach reconstituted the main hematopoietic lineages: progenitor (HSC) Myeloid (GMP, Gran Prog, Granulocytes and Monocytes), Mega-erythroid (Mkps), Basophil (Baso) and Lymphoid (B-cell), with most cells spanning Myeloid/Erythroid branches (Fig. 2b, Extended Data Fig. 3a-c).

To rule out knockout-independent confounding patterns arising from neutral selection and amplification of individual LSK clones, we compared the distribution of 14 different non-targeting control (NTC) guides across hematopoietic lineages in 7 separate experiments. Different NTCs demonstrated a homogeneous distribution with high correlation, confirming a robust approach (Fig. 2c-d). Conversely, sgRNAs targeting the lineage-determining TFs (*Cebpa*/Myeloid), *Klf1*/Erythroid and *Pax5*/B-cell) were absent from their cognate lineages, as expected (Fig. 2d). In line with the *ex vivo* bulk results, loss-of-function of chromatin-factors demonstrated strong lineage-specific patterns and marked functional diversity (Fig. 2e-f, Extended Data Fig. 3e). To characterize their roles, we used three metrics: (i) assessment of chromatin-factor-KO enrichment/depletion across the different hematopoietic lineages (ii) analysis of the effect of chromatin-factor-KOs at key lineage branching points: Myeloid vs Erythroid and Monocyte vs Granulocyte and (iii) analysis of the progression of CF-KOs along Myeloid and Erythroid differentiation trajectories using pseudotime-analysis (Fig. 2f-h, Extended Data Fig. 4a-d, 5a).

Similar to *ex vivo*, we found that disruption of Cohesin (*Stag2*-KO*)* blocked differentiation, causing accumulation of progenitors and myeloid deficiency (Figure 2e). In line with previous studies[22], perturbation of the COMPASS H3K4 methyltransferases revealed marked functional diversity (Fig 2e-h, Extended Data Fig. 4a-d). As reported[23] and similarly to our *ex vivo* screen, SET1 catalytic subunits (*Setd1a, Setd1b)* were required for erythropoiesis and their perturbation eradicated erythroid differentiation, leading to Myeloid and Megakaryocytic progenitor accumulation. Perturbation of the SET1 structural subunit *Wdr82* partially phenocopied *Setd1a/b*-Kos generating a marked accumulation of Megakaryocyte progenitors, but also defining additional roles for *Wdr82* as a B-cell and granulocyte regulator, and suggesting that Wdr82 co-operates with different chromatin complexes during hematopoiesis. Loss-of-function of MLL1 (*Kmt2a*-, *Men1*-KO) also blocked terminal granulocytic differentiation but further enhanced B-cell priming. Finally, the H3K4 mono-methyltransferase MLL4 (*Kmt2d*-KO) functions as a pleiotropic regulator of HSCs self-renewal, early Myeloid branching, and Monocyte versus Granulocyte specification. Collectively, these results demonstrate a lack of redundancy between H3K4 methyl writers during hematopoiesis.

Analysis of BAF perturbation patterns showed similar phenotypic diversity (Fig. 2e-g,i). Disruption of the cBAF member *Smarcd2* confirmed the strong Myeloid dependency found *ex*

vivo inducing an early Erythroid skewing and accelerated Erythropoiesis, a similar pattern to the *Smarcd2*-KO mouse[24]. Alternatively, disruption of the ncBAF complex, defined by *Brd9*, caused a major blockade of B-cell development. Moreover, despite a mild Myeloid-priming defect for *Brd9*-KO, these cells did not undergo terminal Myeloid differentiation but accumulated at progenitor stages. In stark contrast, disruption of the pBAF defining subunit *Pbrm1* prevented Erythropoiesis, augmenting Myeloid and B-cell outputs. These results reveal very different roles for BAF sub-complexes during hematopoietic differentiation.

As predicted in the bulk screens, disruption of complexes with repressive functions, including NURD (*Chd4* and *Rbbp4*), ISWI (*Smarca5*) and heterochromatin repressors (*Atf7ip*, *Setdb1*), produced a similar pattern characterized by accelerated Granulocytic vs Erythroid and B-cell trajectories (Fig. 2e-g,h, Extended Data Fig. 5a). This suggests that most epigenetic Repressors act to safeguard diversity of progenitor identities[25,26], and by limiting extensive Myelopoiesis ensure balanced lineage output. In addition, coRest (*Rcor1*, *Hdac1*) and NuRD (*Chd4*) complex repressive activity proved crucial for terminal erythropoiesis and its depletion induced the accumulation of aberrant Erythroid cells, which expressed high levels of both mature and progenitor markers and were rarely found in the unperturbed scenario (Extended Data Fig. 5b-f). Finally, disruption of *Kdm6a, Hmgbx4* and *Ash1l* led to other aberrant populations not present in the unperturbed scenario (Extended Data Fig. 5b-e).

Transcriptional analysis of chromatin-factor perturbations provided an explanation for the lineage dependencies observed for BAF and SET/MLL members, where *Pbrm1*-KO and *Setd1a*-KO caused a downregulation of Erythroid regulators but *Smarcd2*-KO strongly reduced myeloid markers and transcription factors (Extended Data Fig. 6a). In addition, perturbations that blocked differentiation trajectories; *Rcor1*-, *Wdr82*-, *Stag2*- and *Brd9*-KO, upregulated stem-cell transcription factors (*Hoxa7, Meis1*) and surface markers (*Kit, Cd34*), highlighting that these alterations may facilitate leukemic transformation. By contrast, perturbations of chromatin repressors that show a clear myeloid-bias did not significantly alter the balance between Erythroid and Granulocytic programs, suggesting that subtler and cumulative changes drive these phenotypes. Indeed, GSEA analysis of these perturbations revealed a marked upregulation of inflammatory pathways (TNF alpha or JAK/STAT) and Jun/Fos targets, mediators known to enhance Myeloid lineage outputs under inflammatory stimuli[27] (Extended Data Fig. 6b).

Together, these results demonstrate functional diversity of chromatin-factors in hematopoiesis similar to that of TFs. Interestingly and unlike TFs, most chromatin-factor dependencies cannot be simply explained by their expression patterns (Extended Data Fig. 5g), suggesting that other mechanisms like post-transcriptional regulation, protein-assembly, or differential recruitment explain their functional diversity.

Chromatin-Factors regulate lineage-determining TF accessibility

Intrigued by these findings, we sought to elucidate the interactions between chromatin-factors and transcription factors that may explain the observed lineage-specific dependencies. To this end, we CRISPR-engineered loss-of-function of 10 strong lineage-dependent chromatin-factors in multipotent progenitors, induced both lineage priming and Myeloid differentiation. Thereafter, we used ATAC-seq to perform TF footprinting analysis[28] and define synergistic or antagonistic connections between chromatin-factors and lineage-determining TFs that explain chromatin-factor lineage dependencies (Fig. 3a-d, Extended Data Fig. 7a). In line with their *in vivo* effects, disruption of repressive factors (*Rbbp4*, *Hdac3* and *Setdb1*) resulted in increased accessibility of TFs like Cebps, Hlf[29] and AP-1[30] that drive myelopoiesis upon inflammation. This suggests that these repressive complexes attenuate a myeloid TF program triggered by inflammatory cytokines. Conversely, Myeloid-dependent chromatin-factors (MLL4/Kmt2d, Smarcd2/cBAF and Wdr82) regulate chromatin accessibility around the binding sites of key Myeloid-determining TFs (Cebp, PU.1, SpiB). Disruption of ncBAF (*Brd9*-KO) demonstrated milder effects on Myeloid TF binding site accessibility in line with its subtler *in vivo* effects. Interestingly, we detected different degrees of specificity of chromatin-factor-TF connections, where MLL4 (*Kmt2d*-KO) regulated the accessibility of a more restricted set of myeloid TFs (Cebp, AP-1 and Hlf) but Wdr82 mediates the accessibility of a larger TF repertoire. Finally, we interrogated the dynamics of these effects using a time series analysis in *Wdr82*-and *Kmt2d*-KOs (Extended Data Fig. 7b-d). *Wdr82*-KO induced reduction in Myeloid TF accessibility at early time points (day 3), followed by increased Mega-erythroid accessibility at later time points (days 5-7), suggesting that Wdr82 directly interacts with myeloid TFs and that its depletion indirectly enhances Mega-erythroid TFs activity as a secondary effect. Taken together, these results highlight how specific interactions between chromatin-factors and lineage-determining TFs coordinately drive lineage differentiation.

ncBAF is required for terminal myeloid differentiation

Ex vivo chromatin footprinting of BAF perturbation could not explain the sequential requirement of cBAF vs ncBAF-complexes in, respectively, early Myeloid priming and Myeloid maturation. We speculated that this may reflect incomplete recapitulation of Myeloid progression in our *ex vivo* system, therefore we used an *in vivo* system to dissect the roles of BAF subcomplexes in Myelopoiesis (Fig. 4a). Firstly, using Perturb-seq at a later time point (4-weeks), we detected a massive expansion of *Brd9*-KO clones that mapped predominantly to the multipotent (HSC) and myeloid progenitor (GMP) compartments (Fig. 4b-c), confirming that Brd9 is required for full Myeloid maturation and revealing that *Brd9*-disruption confers a competitive advantage to progenitors.

Chromatin accessibility analysis of *Brd9*-KO myeloid progenitors (Fig. 4d) demonstrated an aberrant pattern, with a marked loss of accessibility at Myeloid maturation loci (*Mpo*, *S100a7-9* and *Ctsg*) and at the motifs of Cebp and AP1, TFs mediating terminal Myeloid maturation (Fig. 4e-f). By contrast, progenitor loci (*Hoxa7-9-10*, *Gata2*, *Meis1)* and progenitor associated TF motifs, including Gata2, demonstrated increased accessibility (Fig. 4e-f). ChIP-seq of Cebpa and Cebpe in freshly sorted GMPs validated the motif analysis derived from ATAC-seq (Fig. 4g-i), confirming that *Brd9*-KO leads to reduced accessibility at the binding sites of these two pro-granulocytic TFs.

Finally, corroborating these results, ChIP-seq analysis of Brd9 in Myeloid Progenitors (GMPs) and mature Myeloid cells (Monocytes) showed a strong enrichment for Cebp and AP1 (Atf4) motifs in mature Myeloid cells and highlighted a switch in the TF partnership of Brd9; from a broader spectrum in GMPs to a Cebp-AP1-centric association in mature Myeloid cells (Fig. 4j-l). We expanded this approach to other CFs/complexes and hematopoietic lineages to highlight lineage-specific TF- chromatin-factor interactions with regulatory potential; for example, a strong connection between Brd9 binding and the Ebf1 motif may explain the strong B-cell dependency of ncBAF-complex members (Extended Data Fig. 8 a-e).

Together, these results demonstrate that specific TF-chromatin-factor interactions mediate lineage specification *in vivo*. In particular, Myeloid maturation is governed by a TF-chromatin-factor switch, where ncBAF/Brd9 transits from a broad range of TF partners towards Cebp. Furthermore, this suggests that Brd9 depletion induces a preleukaemic-like state, related, at least in part, to the failure to transition from "progenitor TF-programs" to later differentiation programs.

Chromatin-Factors enforce differentiation blockade in leukemia

Having extensively dissected chromatin-factor function in normal hematopoiesis, we decided to explore their roles in the aberrantly blocked differentiation states typical of leukemia. To this end, we chose an aggressive *Npm1c/Flt3-ITD* model, driven by the two most common co-occurring mutations in AML that synergize to generate a highly corrupted chromatin landscape, which recapitulates many aspects of human AML with the same genotype[31]. We isolated primary leukemia cells from *Npm1c/Flt3-ITD*/Cas9 mice, cultured them and used CITE-seq to interrogate their differentiation status (Fig. 5a-b). This identified a core of progenitor-like cells with high levels of stem-cell transcripts (*Bcat1*, *Sox4*) and markers (Cd34) and also subpopulations with more differentiated transcriptomes resembling Granulocytic, Erythroid, Basophilic and Megakaryocytic states, which demonstrated decreased fitness when isolated and grown in liquid cultures and clonogenic assays (Fig. 5c-d, Extended Data Fig. 9a-e).

Importantly, Perturb-seq analysis of 50 chromatin-factor-KOs uncovered latent trajectories towards such differentiated endpoints (Fig. 5e, Extended Fig. 9f). Specifically, the depletion of MLL1-COMPASS (*Kmt2a*-KO and *Men1*-KO), NCoR (*Hdac3*-KO), ncBAF (*Brd9*-KO and *Smarcd1*-KO), *Prmt5* and Heterochromatin Regulators *Setdb1* and *Atf7ip* induced transition towards various Granulocytic states. In contrast, loss-of-function of cBAF (*Smarcd2*-KO and *Smarcb1*-KO), MLL4-COMPASS (*Kmt2d*-KO) and *Prmt1* pushed leukemia towards Basophil and Erythroid fates. Interestingly, contrary to their roles in normal hematopoiesis, disruption of repressive factors *Smarca5*, *Chd4* and *Rbbp4* induced the same Erythroid trajectory. Finally, similarly to the normal setting, loss of function of *Wdr82* generated a Megakaryocytic state, which was absent from the unperturbed scenario (Extended Fig. 9g).

Analysis of the growth dynamics of several individual chromatin-factor-KOs revealed a marked defect in cell growth indicating that the differentiation caused by their depletion leads to leukemic exhaustion (Fig. 5e-f). This was especially pronounced for cBAF and COMPASS members, revealing *Npm1c/Flt3-ITD* leukemia to be highly dependent on the epigenetic activities regulated by these complexes. However, of interest and unlike previous reports for MLL-driven leukemias[32,33], our *Npm1c/Flt3-ITD* model did not show vulnerability to *Brd9* (ncBAF) disruption, highlighting that different leukemia-mutations produce specific chromatin states that are variably dependent on individual chromatin-factors.

Of note, our analysis finds potential therapeutic targets in *Npm1c/Flt3ITD* leukemia, including *Prmt1* (Fig. 5e, Extended Fig. 9h), that may be amenable to therapeutic exploitation. As a proof-of-principle for the therapeutic implications of our approach, treatment of the cells with the clinical grade Menin Inhibitor (SNDX-5613/Revumenib) that is currently producing promising results in a clinical trial in KMT2A- and NPM1-mutated AML (AUGMENT-01, NCT04065399)[34] recapitulated the *Men1*- and *Kmt2a*-KO single cell phenotype to induce a dose-dependent granulocytic differentiation and decrease in proliferation (Figure 5g-h).

These collective findings demonstrate how leukemias corrupt chromatin-factors involved in homeostatic differentiation to aberrantly block latent differentiation pathways and how this can be therapeutically exploited to facilitate leukemia exhaustion.

Chromatin-Factors engage in corrupted TF interactions in leukemia

Finally, to interrogate the molecular mechanisms that underpin the requirement for the cBAF and MLL-COMPASS complexes in *Npm1c/Flt3-ITD* AML, and how these differ from normal hematopoiesis, we compared the genome-wide binding patterns of Smarcb1 (an exemplar of cBAF), Kmt2a (COMPASS-MLL1) and Kmt2d (COMPASS-MLL4) using ChIP-Seq across leukemia, normal Myeloid progenitors (GMP) and mature Myeloid subsets (Fig. 6a). Of note, these analyses demonstrated marked redistribution of the cBAF and MLL1/4-COMPASS complexes upon leukemia induction (Fig. 6a, Extended Data Fig. 10b-d), identifying 3 major binding patterns: (1) Leukemic-specific, enriched in molecular functions like Tyrosine Kinase Signaling related to the *FLT3-ITD* mutation (2) Common to leukemia and Myeloid progenitors and (3) Normal-specific.

Motif analysis across the three binding patterns revealed corrupted leukemic TF-chromatin-factor interactions (Fig. 6b), specifically between Stat and Runx TFs and MLL1/4 and cBAF complexes. In contrast, Pu.1 and IRF factors were associated with cBAF specifically in normal cells. ChIP-seq profiling of Stat5a, Runx1, and Runx2 confirmed the motif analysis, demonstrating co-binding of these TFs and BAF or MLL1/4 complexes at key pro-leukemia genes including *Hoxa7* and *Hoxa9* (Fig. 6c-d, Extended Data Fig. 10e). Moreover, chromatin accessibility profiling of leukemia cells disrupted for cBAF (*Smarcd2*-KO), MLL1 (*Kmt2a*-KO) or MLL4 (*Kmt2d*-KO) showed that these chromatin-factors are required to maintain optimal accessibility at Stat5a and

Runx2 binding loci (Fig. 6d-e, Extended Data Fig. 10f). Finally, loss-of-function of *Stat5a* in *Npm1c/Flt3-ITD* cells significantly reduced their proliferative fitness, confirming the importance of these TF-chromatin-factor switches for leukemic maintenance (Fig. 6f). Taken together, these findings demonstrate how individual chromatin-factors required for normal myeloid lineage determination engage in corrupted interactions with alternative TF partners to promote differentiation blockade, thereby maintaining cellular fitness in AML.

Discussion.

In this study, we have generated a detailed lineage-dependency map for CFs during hematopoiesis (Fig 7a-b). We uncover a remarkable phenotypic diversity that, for some chromatin-factors, phenocopies key lineage-determining TF functions (i.e Smarcd2/Cebpa or Pbrm1/Klf1). Demonstrating the complex nature of chromatin-factor regulation, we show highly divergent roles for complexes that mediate the same, or very similar epigenetic activities, including the different COMPASS H3K4methyltransferases or the various BAF subcomplexes. The lack of redundancy amongst COMPASS complexes has been described in other systems[35], suggesting lineage-specific requirements for H3K4 methylation deposition by particular COMPASS members or regulation via catalytic-independent roles[36,37]. In addition, reshaping of BAF complexes has been shown to regulate cellular fates in pancreatic Beta-cells[38], and here we show evidence for another switch, from cBAF to ncBAF, that regulates myeloid differentiation ensuring full lineage progression. Of note, *Brd9*/ncBAF perturbation led to the accumulation of Myeloid progenitors with a preleukaemic gene-expression program, mimicking the aberrant splicing of *BRD9* that results in its degradation, a process mechanistically implicated in the AML-precursor lesion, MDS[39]. Lastly, in stark contrast to functional diversity for some chromatin-factors, we observe a common function for different Chromatin Repressors as attenuators of excessive granulopoiesis, suggesting repressive CF as a key buffering mechanism in the interplay between inflammatory signaling and chromatin state.

What then underlies such chromatin-factor specificity? Inspired by previous studies[40], we demonstrated that specific TF-chromatin-factor interactions mediate lineage diversification via regulating the accessibility of key lineage-determining TFs. However, as the chromatin-factors investigated include a large number of proteins with diverse functions (Remodelers, epigenetic-readers, -writers, -erasers, etc), we think it unlikely that a simple "one size fits all" mechanism

governs chromatin-factor-TF interactions. We believe it more probable that the interactions are usually directed by TFs, that physically recruit specific chromatin-factors to induce lineage-specific chromatin configurations [41-45]. However, an alternative and non-mutually exclusive explanation could be a sequential model, where specific chromatin-factors, already deployed through multivalent interactions with epigenetic modifications, regulate subsequent TF activity by modulating the chromatin state at the TF binding sites. Specifically, we propose this as a possible mechanism whereby the Chromatin Repressors attenuate Myeloid pro-inflammatory TF responses.

Regardless of the specific detail, central to the chromatin-factor-TF collaboration appears to be the chromatin-factor-mediated regulation and maintenance of TF accessibility and binding[46]. As evidenced in our dynamic ATAC-seq studies, early alterations in accessibility are observed even following KO of methyltranseferase complex components (*Kmt2d* and *Wdr82*) that lack chromatin remodeling activity. Therefore, these data also inform interactions among chromatin-factors with regulatory potential; the changes in accessibility suggest that the MLL4 and SET1 complexes recruit chromatin remodelers[47], likely BAF members, an observation reinforced by the strong KO phenocopy observed between *Kmt2d* and several cBAF subunits. However, whether this recruitment involves direct protein-protein interactions, or indirectly and dependent upon catalysis, through configuring a local pattern of histone methylation that recruit remodelers via their own reader modules[48], will require further investigation.

Finally, studying chromatin-factor requirement in AML, we highlight corrupted roles for certain chromatin-factors in malignant hematopoiesis. Here, they alter their normal lineage regulatory role to conversely block differentiation in leukemia, reiterating that this blockade is an active process required for leukemic fitness (Fig. 7c). In characterizing these patterns, we have also identified leukemic dependencies that may be amenable to inhibition/targeting, including *Prmt1*, *Hdac3*, *Setdb1* and *Kmt2d*. Furthermore, we identify that these altered roles relate to leukemia-specific TF-chromatin-factor interactions, including COMPASS and BAF factors that rewire their TF networks towards Runx and Stat factors (Fig. 7d). These observations also have clinical implications; TFs and chromatin-factors have pleiotropic effects across multiple tissues, however targeting leukemia-specific TF-chromatin-factor interactions, important only for leukemia cells, will likely have much lower toxicity and higher specificity. This can be achieved chemically or

through synthetic approaches[49,50] and requires not only a detailed structural understanding of specific interactions, but also knowledge of the mechanisms governing individual TF-chromatin-factor associations[51]. Our study provides a blueprint to expand such approaches to leukemia.

Our approach combining large-scale CRISPR-screening with downstream single-cell analysis *in vivo* could be readily deployed to assess the role of other classes of proteins in hematopoiesis or adapted to other organ and tumor systems. However, several limitations must be considered. Firstly, our currently limited *ex vivo* differentiation readout could be supplemented by the use of other cytokine cocktails that permit interrogation of Lymphoid lineages[52]. In addition, our *in vivo* approach does not completely reflect steady-state hematopoiesis, but more regenerative hematopoiesis in the post-transplant setting. Thus, some of the roles described for individual CF, including the blockade of granulopoiesis by several repressors, may differ under steady state conditions. Screening in homeostasis could be achieved by combining inducible Cas9 systems[53,54,55] that permit inducible loss-of-function of specific factors in steady-state conditions, following transplantation, full reconstitution and a return to homeostatic hematopoiesis.

Taken all together, this study shows that chromatin factors constitute a specific regulatory layer that should be accorded equal weighting with TFs when studying cell-fate decisions. It lays the basis for additional, in-depth interrogation of specific chromatin-factor-TF interactions and functions, using multidisciplinary approaches ranging from *in vivo* functional approaches to protein-protein interactions, that we feel are warranted to further elucidate chromatin-factor-TF functions.

ACKNOWLEDGMENTS

The authors thank Dr. Markus Wiederstein from the University of Salzburg for HPC support, Dr. Katarzyna Kania from CRUK Genomics Core for assistance with Single-cell methods and Nerea Cano and Martin Dawes for superb administrative support. We thank David Kent and Arnau Sebe-Pedrós for their constructive advice and discussions regarding the manuscript. CRISPseq-BFP-backbone (Addgene plasmid #85707) was produced by Professor Ido Amit, Weizmann Institute

of Science. pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260) was produced by Dr Didier Trono, EPFL Lausanne.

This project was funded by ”La Caixa” Foundation (agreement LCF/PR/HR20/52400016. BJPH, FP), Marie-Skłodowska-Curie International Fellow (886474. DLA), European Hematology Association Junior Research Grant, Cancer Research UK (C18680/A25508. DLA), European Research Council (647685. BJPH), Cancer Research UK Cambridge Major Centre (C49940/A25117. BJPH), Kay Kendall Leukaemia Fund Junior Fellowship (KKL1440. NN), NIHR Cambridge Biomedical Research Centre (BRC-1215-20014. BJPH), UKRI Medical Research Council (MC_PC_17230. BJPH) supporting the Wellcome – MRC Cambridge Stem Cell Institute, Wellcome Trust (203151/Z/16/Z. BJPH) supporting the Wellcome – MRC Cambridge Stem Cell Institute. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. 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.

AUTHOR CONTRIBUTIONS

Conceptualization: DLA, BJPH

Methodology and Data Collection: DLA, AGS, NN, CDV, GG, MN, TB, JZ, FM, NT, IAC, CKL, DA, AL, BS.

Data Analysis: DLA, JME, TG, JTK, NF, BJPH

Visualization: DLA, JME, NF, BJPH

Funding acquisition: DLA, FP, BJPH

Project administration: BJPH

Supervision: DLA, NF, BJPH

Writing – original draft: DLA, BJPH

Writing – review & editing: DLA, JME, NF, BJPH

COMPETING INTERESTS

TG and JPTK are employees receiving compensation from Relation Therapeutics. JTK is a founder of Relation Therapeutics. DLA is a consultant of Relation Therapeutics.

FIGURE LEGENDS

Fig. 1. Functional screens identify lineage specificities for Chromatin Factors (CFs) in hematopoiesis. (**a**) Experimental approach. **(b)** Differentiation systems and FACS-based Readouts. 1- Self-Renewal vs Differentiation. 2- Lineage priming: Mega-erythroid vs Myeloid. 3- Myeloid Differentiation: Mature Myeloid vs Non-Myeloid. 4- Terminal Myeloid Maturation vs Immature. **(c)** Averaged lineage scores of "Differentiation vs Self-renewal" (y-axis) versus "Lineage priming" (x-axis). Significant hits are shown with a red cross. NTC-sgRNAs are shown with a blue cross. Data for Non-Cas9 cells are shown in the background using a yellow-blue density. (**d**) Lineage scores for CFs grouped on the basis of complex membership. Dot color represents the lineage score, dot size the percent of significant guides (Supplementary Table 3). (**e**) Exemplar immunophenotypic validations for CFs in the lineage priming (top) and myeloid differentiation (bottom) systems.

Fig. 2. Perturb-seq highlights disparate lineage dependencies for chromatin complexes during hematopoiesis. (**a**) Schematic drawing of the *in vivo* Perturb-seq. **(b)** UMAP of the single-cell transcriptomes. Clusters are annotated using external reference maps[56]. The analysis integrates 7 different experiments. (**c**) UMAP showing the distribution of unperturbed cells (NTC-sgRNAs) and specific perturbations. **(d)** Scatterplot showing correlation between experimental batches. Each dot represents the abundance of two NTC-sgRNAs in a given population. **(e)** Enrichment/depletion of CF-KOs across hematopoietic populations (Supplementary Table 4). Dot color and size relate to the log2 odds ratio and the percent of significant enrichments. **(f)** Effect of specific CF-KOs on Myeloid versus Erythroid priming. Positive values (red) show enhanced Myeloid priming. Negative values (blue) indicate reduced Myeloid priming. **(g)** Trajectory analysis of specific CF-KOs along Myeloid differentiation. Cells are ordered from HSCs to mature Granulocytes using pseudotime. **(h-j)** Graphic representation of the roles of key Chromatin Regulatory Complexes.

Fig. 3. Chromatin Factors regulate the accessibility of Lineage Determining Transcription Factors. (a) Schema of the experiment. **(b)** Chromatin accessibility for representative myeloid and mega-erythroid loci after disruption of specific CFs. The accessibility profiles correspond to two merged replicate experiments. *S100a9*, *Ly6g*, *Hbba*, and *Pf4* loci coordinates are chr3:90685583-90703276, chr15:75134355-75144437, chr7:103869996-103874051, and chr5:90771031-90773155, respectively. Y-axes ranges for all KOs in the same regions are 0.02-1.5, 0.02-1, 0.02-2, 0.02-4, respectively. **(c)** Volcano plot showing differentially bound TF motifs (estimated by TOBIAS) in *Wdr82*- and *Hdac3*-KOs under lineage priming conditions. TF motifs demonstrating gained and lost accessibility in each CF-KOs (compared to NTCs) are shown in green and purple, respectively. n = 2 biologically independent experiments (**d**) Heatmap summarizing the effect of 10 CF-KOs on TF motif footprints estimated by TOBIAS. n = 2 biologically independent experiments. Dot color and size relate to log2 fold change and the -log10 adjusted p-value, respectively[28].

Figure 4 | Disruption of ncBAF leads to a pre-leukemic accumulation of Myeloid progenitors with diminished Cebp-AP1 activity. (a) Schema of the *in vivo* perturbation of cBAF (*Smarcd2*-sgRNA) and ncBAF (*Brd9*-sgRNA) **(b)** Proportions of cells with a specific sgRNA at 14 and 28 days post-transplant. (**c)** UMAP showing the distribution of *Brd9*-KO and Control (NTC-sgRNA) cells at 28 days post-transplant. **(d)** Schema of the *in vivo* ATAC-seq experiment. **(e)** ATAC-seq signal at Myeloid progenitor (top) and differentiated (bottom) loci. Coordinates, *Hoxa7/9/10*: chr6:52214971-52236669, chr11:18912448-19036437*; Meis1*: chr6:88186822-8820729; Gata2: chr11:87788022-87796472, *Mpo*: chr3:90651905-90703284; *S100a7/a8/a9*: chr14:56098019-56107286; *Ctsg*: chr14:56098019-56107286. **(f)** Volcano plot showing differentially bound TF motifs (estimated by TOBIAS) between Control (NTC) and *Brd9*-KO GMPs. N = 2 biologically independent experiments (**g**) Genome browser tracks showing ATAC-seq, Cebpa-ChIP-seq and Cebpe-ChIP-seq in *wild-type* myeloid progenitors (GMPs). Loci coordinates are the same as in Fig. 4e. (**h-i**) Quantification of accessibility changes between *Brd9*-KO and Control (NTC) GMPs. **(h)** MA-plot showing loci overlapping with Cebpa binding (red) **(i)**. Boxplots showing accessibility loss (statistically tested using a two-sided KS test, n=2) at Cebpa (m=11316, statistic=0.86, pvalue=0.0) and Cebpe sites (n=10409, statistic=0.85, pvalue=0.0). The boxplot displays the median as the center line of the box, with the box representing the distribution's 25th

(minima) and 75th (maxima) percentiles. The whiskers extend up to 1.5 times the interquartile range (Q3-Q1) from the minima and maxima. **(j)** Schema of the ChIP-seq analysis. **(k)** Heatmaps showing specific binding of Smarcb1 (cBAF) and Brd9 (ncBAF) in Myeloid progenitors (GMPs) and bone-marrow Monocytes (Mono). Two merged independent ChIP-seq experiments were used. **(l)** TF motif enrichment at cBAF and ncBAF sites between progenitor (GMP) and mature (Monocytes) myeloid cells. The axes represent TF motif odds ratio. All colored TF motifs have adjusted p-values < 0.001. The analysis was performed with 2 independent experiments per population.

Figure 5 | Npm1c/ Flt3ITD leukemia abrogates normal Chromatin Factor CF function to maintain leukemic fitness through enforcing differentiation blockade. (**a)** UMAP projection of single-cell transcriptomes from *Npm1c/Flt3-ITD* primary leukemia. Color-coded clusters correspond to cells with specific signatures: Leukemic Stem Cells (green), Granulocytic-like (blue), Erythroid-like (red), Basophil-like (purple) and Megakaryocytic-like (violet). The analysis integrates datasets from 6 different Perturb-seq experiments. **(b)** mRNA and CITE-seq derived expression of lineage makers in the differentiated leukemia subpopulations. **(c-d)** Clonogenic and proliferation assays for differentiated leukemia subpopulations, isolated following the strategy in Extended Fig. 9e. Colonies were counted after 7 days of culture in methylcellulose. Proliferation and clonogenic values are obtained from n = 4 biologically independent experiments. ****P<0.0001 (Two Way ANOVA). Error bars are SEM. **(e)** Enrichment analyses of specific CF-KOs across differentiated leukemia subpopulations. Dot color and size relate to the log2 odds ratio and the percent of significant enrichments versus NTCs, respectively. The analysis is based on measurements for two sgRNAs per CF target. All values are shown in Supplementary Table 6. **(f)** Growth curves for leukemic CF-KOs. The assay measures the change in the proportion of BFP (sgRNA) expressing cells over time, n=4 biologically independent experiments. All ***P<0.001 but Smarcd1 vs NTC (day 9) where ***P=0.0009 (Two Way ANOVA). Error bars are SEM. **(g)** FACS analysis of mega-erythroid (CD55) and myeloid (CD11b, Gr1) surface differentiation markers in leukemia cells depleted for cBAF (*Smarcb1*-KO and *Smarcd2*-KO) and MLL (*Kmt2a*-KO) components. **(h)** FACS analysis of myeloid surface differentiation markers (CD11b, Gr1) in leukemia cells treated with increasing doses of Men1 inhibitor (SNDX-5613).

Figure 6 | Chromatin Factors enforce differentiation blockade in AML through corrupted TF interactions. (a) Heatmaps showing ChIP-seq signal for cBAF (Smarcb1), MLL (Kmt2a) and MLL4 (Kmt2d) in leukemia (*Npm1c/Flt3-ITD*), *in vivo* myeloid progenitors (GMPs), *in vivo* monocytes and *ex vivo*-derived primary monocytes. n=2 independent ChIP-seq experiments per factor. **(b)** Motif Enrichment analysis of cBAF, MLL and MLL4 binding patterns specific for leukemia (Leukemic), common between leukemia and myeloid progenitors (Common), and specific for normal myeloid cells (Normal). **(c)** Boxplot showing Stat5a, Runx1 and Runx2 binding signal (ChIP-seq) at the Leukemic, Common and Normal loci defined in figure 6a (n=2). The number of loci comprising the Leukemic, Common and Normal categories are 2019, 581, and 2361 for Smarcb1; 2427, 2346, and 5789 for Kmt2a and; 3144, 810, and 4129 for Kmt2d. (**d)** Genome-browser tracks showing ChIP-seq signal for Smarcb1, Kmt2a, Kmt2d, Stat5a and Runx2 in leukemia and, ATAC-seq for control and chromatin-factor depleted leukemia cells. The chosen loci are leukemic-specific. n=2 independent experiments. **(e)** Boxplots showing changes in chromatin accessibility at leukemic loci bound by Stat5a, Runx1 and Runx2 upon depleting specific chromatin factors. cBAF, n = 2 independent experiments. *Smarcd2*-KO: n=1385, 1050, 1711; statistic=0.73, 0.75, 0.76; pvalue=$5e^{-15}$, 0, 0. *Kmt2a*-KO: n=1288, 695, 1427; statistic=0.78, 0.86, 0.78; p-value=0, $9e^{-16}$, 0. *Kmt2d*-KO: n=1482, 990, 1913; statistic=0.70, 0.70, 0.69; p-value=$2e^{-15}$, 0, 0. The decay in accessibility was statistically tested using a two-sided KS test. **(f)** Growth curves for Cas9-leukemic cells expressing NTC and anti-Stat5a sgRNAs, n = 3 independent experiments. ***P<0.001 (Two Way ANOVA). Error bars are SEM. Boxplots in **c** and **e** display the median and the distribution's 25th (minima) and 75th (maxima) percentiles. The whiskers extend up to 1.5 times the interquartile range (Q3-Q1) from the minima and maxima.

Figure 7 | Summary model of Chromatin Factor function and CF-TF interactions in normal and malignant hematopoiesis (see discussion in text).

REFERENCES

1. Stadhouders, R., Filion, G.J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345-354 (2019)
2. Ding, Y., Liu, Z., Liu, F.Transcriptional and epigenetic control of hematopoietic stem cell fate decisions in vertebrates. *Dev Biol* **475**, 156-164 (2021)
3. Al-Mousawi, J. & Boskovic, A. Transcriptional and epigenetic control of early life cell fate decisions. *Curr Opin Oncol* **34**, 148-154 (2022)
4. Valencia, A.M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat Cell Biol* **21**, 152-161 (2019)
5. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553,** 418–426 (2018)
6. Orkin, S.H. & Zon, L.I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644 (2008)
7. Cedar, H. & Bergman, Y. Epigenetics of haematopoietic cell development. *Nat Rev Immunol* **11**, 478-488 (2011)
8. Wilson, N.K. et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532-544 (2010)
9. Downing, J.R. The core-binding factor leukemias: lessons learned from murine models. *Curr Opin Genet Dev* **13**, 48-54 (2003)
10. Robb, L. et al. The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J* **15**, 4123-4129 (1996)
11. Rosenbauer, F. & Tenen, DG. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat Rev Immunol* **7**, 105-117 (2007)
12. Papaemmanuil, E. et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209-2221 (2016)
13. Cuartero, S. et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat Immunol* **19**, 932–941 (2018)
14. Sasca, D. et al. Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood* **134**, 2195-2208 (2019)
15. Meeks, J & Shilatifard, A. Multiple Roles for the MLL/COMPASS Family in the Epigenetic Regulation of Gene Expression and in Cancer. *Ann Rev Cancer Biol.* **1**, 425-446 (2019)
16. Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C. & Chan, H.M. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet* **36**, 936-950 (2020)
17. Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735-51 (2014)
18. Scacchetti, A. & Becker, P.B. Variation on a theme: Evolutionary strategies for H2A.Z exchange by SWR1-type remodelers. *Curr Opin Cell Biol* **70**, 1-9 (2020)
19. Brahma, S. et al. INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat Commun* **8**, 15616 (2017)
20. Yu, L. et al. An erythroid-to-myeloid cell fate conversion is elicited by LSD1 inactivation. *Blood* **138** 1691-1704 (2021)
21. Gillespie, M.A. et al. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Mol Cell* **78**, 960-974 (2020)

22. Kerimoglu, C. et al. KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Rep* **20**, 538-548 (2017)
23. Li, Y. Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. *Nucleic Acids Res* **44**, 7173-7188 (2016)
24. Priam, P. et al. SMARCD2 subunit of SWI/SNF chromatin-remodeling complexes mediates granulopoiesis through a CEBPε dependent mechanism. *Nat Genet* **49**, 753-764 (2017)
25. Koide, S et al. Setdb1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. *Blood* **128**, 638-649 (2016)
26. Summers, A.R. et al. HDAC3 is essential for DNA replication in hematopoietic progenitor cells. *J Clin Invest* **123**, 3112-3123 (2013)
27. Paudel, S. et al. Regulation of emergency granulopoiesis during infection. *Front. Immunol.* **13**, 961601 (2022)
28. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**, 4267 (2020)
29. Gombart, A.F. et al. ATF4 differentially regulates transcriptional activation of myeloid-specific genes by C/EBPepsilon and C/EBPalpha. *J Leukoc Biol* **81**, 1535-1547 (2007)
30. Wahlestedt, M. et al. Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. *Cell Rep* **21**, 2251-2263 (2017)
31. Yun, H. et al. Mutational synergy during leukemia induction remodels chromatin accessibility, histone modifications and three-dimensional DNA topology to alter gene expression. *Nat Genet* **53**, 1443-1455 (2021)
32. Zhou, L., Yao, Q., Li, H. & Chen, J. Targeting BRD9 by I-BRD9 efficiently inhibits growth of acute myeloid leukemia cells. *Transl Cancer Res* **10**, 3364-3372 (2021)
33. Hohmann, A., et al. Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat Chem Biol* **12,** 672–679 (2016)
34. Wang, E.S. et al. Preliminary Data on a Phase 1/2A First in Human Study of the Menin-KMT2A (MLL) Inhibitor KO-539 in Patients with Relapsed or Refractory Acute Myeloid Leukemia. *Blood* **136**, 7-8 (2020)
35. Crump, N.T. & Milne, T.A. Why are so many MLL lysine methyltransferases required for normal mammalian development? *Cell Mol Life Sci* **76**, 2885-2898 (2019).
36. Douillet, D. Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat Genet* **52**, 615-625 (2020).
37. Dorighi, K.M. et al. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell* **66**, 568-576 (2017) .
38. Wei, Z. Vitamin D Switches BAF Complexes to Protect β Cells. *Cell* **173**, 1135-1149 (2018)
39. Inoue, D. et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432-436 (2019)
40. Barisic, D., Stadler, M.B., Iurlaro, M. & Schübeler, D. Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* **569**, 136-140 (2019)
41. Liu, L. et al. Histone methyltransferase MLL4 controls myofiber identity and muscle performance through MEF2 interaction. *J Clin Invest* **130**, 4710-4725 (2020)
42. Sandoval, G.J. et al. Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. *Mol Cell* **71**, 554-566 (2018)
43. Vierbuchen, T. et al. AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol Cell* **68**, 1067-1082 (2017)
44. Huang, X. et al. OCT4 cooperates with distinct ATP-dependent chromatin remodelers in naïve and primed pluripotent states in human. *Nat Commun* **12**, 5123 (2021)

45. Ninkovic, J. et al. The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell Stem Cell* **13**, 403-418 (2013)
46. Schick, S. et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nat Genet* **53**, 269-278 (2021)
47. Park, Y.K. et al. Interplay of BAF and MLL4 promotes cell type-specific enhancer activation. *Nat Commun*, **12**, 1630 (2021)
48. Lange, M. et al. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev* **22**, 2370-2384 (2008)
49. Holmes, A.G. et al. A MYC inhibitor selectively alters the MYC and MAX cistromes and modulates the epigenomic landscape to regulate target gene expression. *Sci Adv* **8** (2022)
50. Speltz, T.E. et al. Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains. Nat Biotechnol (2022)
51. Kadoch, C. Diverse compositions and functions of chromatin remodeling machines in cancer. *Sci Transl Med* **11**, eaay1018 (2019)
52. Trotman-Grant, A.C et al. DL4-µbeads induce T cell lineage differentiation from stem cells in a stromal cell-free system. *Nat Commun* **12**, 5023 (2021)
53. Lundin, A. et al. Development of an ObLiGaRe Doxycycline Inducible Cas9 system for pre-clinical cancer drug discovery. *Nat Commun* 11, 4903 (2020).
54. Bowling, S. et al. An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **181**, 1410-1422 (2020)
55. Dubrot, J. et al. In vivo screens using a selective CRISPR antigen removal lentiviral vector system reveal immune dependencies in renal cell carcinoma. *Immunity* **54**, 571-585 (2021).

METHODS

Mouse models.

C57BL/6J (Jackson Laboratory #JAX_000664) and Gt(ROSA)26Sortm1.1(CAG-cas9*/EGFP)Rsky (Jackson Laboratory #JAX_026179) were used for all experimental procedures. The *Npm1c/Flt3-ITD/Cas9* model has been extensively described previously[31,57,58]. The maximal tumour size allowed by the Home Office Licence for this project and authorised by the Animal Welfare Ethical Review Body at the University of Cambridge is 1.2cm in diameter; however as the animals develop a liquid not a solid tumor in none of the experiments was this tumour size exceeded. Housing conditions were: 12h-12h dark/light cycle, temperature 21º +/-1, 40% humidity.

Bulk CRISPR Screens

CRISPR Libraries Construction.

sgRNA-CRISPR libraries (Supplementary Table 2) were ordered from IDT Technologies and cloned using Gibson assembly in the CRISP-seq backbone (Addgene #85707).

Cloning of individual sgRNAs

Individual oligos were cloned in the CRISP-seq backbone (Supplementary Table 2) using a Golden-Gate protocol and sequenced using a U6 Forward primer (Supplementary Table 8).

Lentiviral production.

HEK293T (Sigma, 12022001-DNA-5UG) cells were transfected with the CRISP-seq vectors, pMD2-G (Addgene #12259) and psPAX2 (Addgene #12260) using lipofectamine 3000 following the manufacturer's protocol. After 10 hours, media was replaced with Opti-MEM (ThermoFisherScientific) plus 1% Pen/Strep. The viral supernatant was collected at 48 hours, filtered using 0.45 uM filters and concentrated with a 100 Kda Centricon at 3000 xg at 4 ºC.

Isolation of murine hematopoietic progenitors (LSK) from bone marrow.

Femora, tibiae, ilia, humerus, sternum and scapula were harvested from 12-14 week old C57BL/6J and ROSAxCas9 mice (equal ratio of males and females), crushed in cold autoMACS Running Buffer and filtered through a 70 μM strainer. Erythrocytes were lysed and c-Kit+ cells were enriched using mouse CD117 magnetic beads (Miltenyi). The c-Kit enriched fraction was stained with anti-Lineage (B220, CD3, CD11b, Gr1, Ter-119), anti-CD117 (cKit) and anti-Sca1 (Supplementary Table 9). Lin-/cKit+/Sca1+ cells (LSKs) were FACS-sorted in 1 mL of DMEM/F12 + 1X Pen/Strep, centrifuged at 350xG for 10 minutes an processed for *ex vivo* screens (see below).

Dilution 1:100 for all antibodies, except 1:50 for CD34.

An exemplar gating strategy can be found in Supplementary Figure 1

Ex vivo CRISPR screens Cultures.

After FACS-sorting, multipotent **(**LSK) or myeloid (GMP) progenitors were resuspended at 250 cells/ul in DMEM/F12 plus 1% PSG, PVA 87%, ITS-X, 1X Hepes, 100 ng/mL mTpo and 10 ng/mL mSCF and plated in 96w plates with 25,000 cells (100 ul) per well. Immediately after plating, cells were transduced with the CRISPR libraries to reach a MOI of ~20% infection. After

12 hours, 2.5 volumes of fresh medium were added. Then, at 48 hours post-infection, cells were transferred (1000 cells/ ml) to the "Screen Media" and cultured:

Stem cell vs Differentiation:

Time: 7 days.

Media. Complete DMEM/F12 (1% PSG, PVA 87%, ITS-X, 1X Hepes) + 100 ng/mL mTpo +10 ng/mL mSCF[59].

Lineage Priming:

Time: 5 days.

Media: Complete DMEM/F12 + 100 ng/mL murine Tpo + 1 ng/mL mFlt3L + 10 ng/mL mSCF and 1 U/mL Epo.

Myeloid differentiation:

Time: 4 days.

Media: IMDM, 20% FBS, 1% PSG, 10 ng/mL mGM-CSF, 10 ng/mL mSCF, 5 ng/mL mG-CSF, 5 ng/mL mIL-3, 5 ng/mL mIL-6, 5 ng/mL mIL-5, 5 ng/mL mFlt3L, 2 ng/mL mTpo and 2 U/mL Epo .

Terminal Myeloid Differentiation:

Time: 2 days.

Media: IMDM, 20% FBS, 1% PSG, 10 ng/mL mGM-CSF, 10 ng/mL mSCF, 5 ng/mL mG-CSF, 5 ng/mL mIL-3, 5 ng/mL mIL-6, 5 ng/mL mIL-5, 5 ng/mL mFlt3L.

CRISPR FACS Readouts.

Cultures were harvested and stained with the Readout-specific antibody cocktails (below, Supplementary Table 9) plus a viability marker (TOPRO or Propidium Iodide). Cas9 (GFP+) and Non-Cas9 (GFP-) populations were sorted in 1.5 mL of PBS + 0.1% BSA.

Stem cell vs Differentiation:

Multipotent Progenitor (Lin-, cKit+, Sca1+)

Differentiated (Lin-, cKit+, Sca1-).

Lineage Priming:

Myeloid Progenitor (Lin-, cKit+, Sca1-, FcgRIII+)

Mega-erythroid progenitor (Lin-, cKit+, Sca1-, FcgRIII-).

Myeloid Differentiation:

Mature Myeloid (FcgRIII+, CD11b+)

Non-myeloid (FcgRIII-, CD11b-).

Terminal Myeloid Differentiation:

Mature Myeloid (CD11b+, Gr1+).

Immature Myeloid (CD11b-, Gr1-).

Dilution 1:100 for all antibodies.

Exemplar gating strategies can be found in Supplementary Figures 2-3.

Bulk CRISPR Library Preparation.

Sorted cells were lysed in 40 uL of 0.2% SDS and 2 uL of Proteinase K at 42 ºC for 30 min. Then, gDNA was isolated with a 2X SPRI Cleanup and NGS libraries were prepared from purified gDNA using 2-step PCR protocol (2X Kappa Hifi):

1st PCR: 10 uM Read1-U6 and Read2-scaffold primer mix (Table S2). 3 min at 98 ºC; 20x [10 s at 98 ºC, 10 s at 62 ºC, 25 s at 72 ºC]; 2 min at 72 ºC.

2nd PCR: 10 uM P5-index and P7-index mix. 3 min at 98 ºC; 10x [10 s at 98 ºC, 10 s at 62 ºC, 25 s at 72 ºC]; 2 min at 72 ºC.

Libraries were purified with 1X SPRI cleanup and sequenced at 10M reads per sample.

Computational analysis of FACS-based CRISPR screens.

Analyses were performed in R (version 4.0.2). Biological replicates were merged by summing the counts. Aggregated counts were normalized by calculating normalizing factors using the function *calcNormFactors* from edgeR (version 3.32.1)[60] on non-targeting guide counts. Counts were transformed to counts per million (CPM) and $\log_2$ normalized using limma (version 3.46.0)[61]. A raw lineage score comparing pairs of populations (A and B) was then calculated by subtracting the $\log_2$ CPMs of population A from population B, for each library of guides. To assess significance , we next calculated the probability of observing a given score in the Cas9 data given the Non-Cas9 data, where no effective KO occurs. For each comparison and each library, we centered and scaled Cas9 data based on the mean and standard deviation calculated from the Non-Cas9 data. The resulting normalized scores were used to calculate the probabilities of observing values as extreme (two-sided) using the function *pnorm*. The resulting probabilities represent the

probability of an observed value given a background distribution but with the important difference to p-values that in our analyses the background distribution was not based on replicates but on the Non-Cas9 data. We next corrected these probabilities for multiple testing using the function *p.adjust* with the method "BH" and selected values smaller than 0.05 as significant.

Validation of single candidates with Flow cytometry.

Cas9 Progenitor cells (LSK) were transduced with CRISPR-seq lentiviral vectors, cultured and stained using the conditions described above and analyzed with a FACS Aria. FACS data was analyzed using FlowJo v10.8.0 Software (BD Biosciences).

Perturb-seq

Perturb-seq libraries.

For each target, we cloned the top 2 performing sgRNAs in the Lenti-PerturbSeq-BFP vector, which we build by modifying the original Lenti-CRISPR-BFP vector by replacing the original sgRNA scaffold for a sgRNA scaffold containing the 10X capture sequenced CR1Cs1[62]. Lentiviral particles were prepared as specified for Bulk screens.

In vivo Perturb-seq.

We performed 7 experiments with 10-15 factors and 2 Non-Targeting Control sgRNAs per batch. In each batch, 300,000 LSKs were isolated from 12-14 week-old ROSA26xCas9 mice (equal ratio of males and females), and transduced with the Perturb-seq library to reach 10% infection (>1000X coverage). After transduction, cells were left to recover for 36h in Stem Cell media (DMEM/F12, ITSX, PVA, 100 ng/uL mTpo, 10 ng/uL mSCF)[59]. Then, cell number and viability were assessed with Cellometer K2 Image Cytometer (Nexcelom Bioscience) and 50,000 viable cells were transplanted to each irradiated (902 cGy, 1 minute) 12 week-old adult B6.SJL-Ptprca Pepcb/BoyJ (CD45.1) mice (Jackson #002014) via tail injection.

After two weeks, mice were euthanized, cKit+ cells were isolated and stained with TOPRO (viability), anti-Lineage (CD3, CD19, Ter119, CD11b, Gr1) and anti-CD117 (cKit) antibodies (1:100 dilution). Then, we gated GFP+ (Cas9), BFP+ (sgRNA) cells and FACS-sorted Lineage- and Lineage+/cKit+ fractions. Cells from each of these gates were processed in the Chromium Controller to reach 500 cells per sgRNA. An exemplar gating strategy can be found in Supplementary Figure 4.

Perturb-seq in leukaemia

0.25-0.5x10^6 DM Cas9 cells were transduced with Perturb-seq libraries using Retronectin-mediated (Takara Bio) and maintained in culture for 6 days. Transduced, BFP-positive and 7AAD-negative live cells were FACS-sorted (BD Influx; BD Bioscience). Finally, 16,000 live cells (cell number and viability assessed with Cellometer K2 Image Cytometer) were processed in a 10X single-cell RNA-seq partition aiming at a final coverage of 500 single-cells per sgRNA.

Perturb-seq library preparation.

Single-cell libraries were generated using the Chromium Next GEM Single-Cell 3' Reagent Kits v3.1 (Dual Index) using the manufacturer recommended protocol. The resulting libraries were sequenced to a final coverage of 50,000 reads/cell for 3' Gene Expression libraries and 5,000 reads/cell for CRISPR Feature Barcode libraries.

CITE-seq

CITE-seq was performed on 2x10^6 DM Cas9 murine leukemic cells, stained with Total-Seq B antibodies (Biolegend) for CD11b, Ly6C, CD115, CD14, CD150, CD48, CD34, CD117, CD55, CD41, CD326 and FcgRI (Supplementary Table 9) as per manufacturers protocol. Stained cells were FACS-sorted for 7AAD (BD Bioscience) negative, live cells (BD Influx; BD Biosciences). Single-cell RNA-seq libraries were prepared in the Cancer Research UK Cambridge Institute Genomics Core Facility using the Chromium Single Cell 3′ Library & Gel Bead Kit v3.1, Chromium Chip G Kit and Chromium Single Cell 3' Reagent Kits v3.1 User Guide (Manual Part CG000317 for CITEseq). Libraries were sequenced a final coverage of 50,000 reads/cell for 3' Gene Expression libraries and 5,000 reads/cell for Antibody Feature Barcode libraries.

Perturb-seq and CITE-Seq analysis.

Analyses were performed in R (version 4.0.2) unless otherwise stated.

Basic processing and alignment. Raw reads were processed and aligned to the GRCm38/mm10 reference genome assembly (GENCODE vM23/Ensembl 98) using the software cellranger count (version 6.1.1).

Quality control and integration. Starting with the "filtered" data matrix from cellranger, additional quality control and processing was performed. First, cells with low quality were filtered based on the number of detected genes, UMIs, and the percent of mitochondrial reads using Seurat (version 4.0.0)[63]. For each sample, the 90-th percentile of cells was calculated based on the number of detected UMIs and the number of detected genes. Cells with less than 20% of either 90-th percentile (and less than 500 genes and 1000 UMIs as minimum cutoffs) were removed. Cells with more than 10% of mitochondrial reads were also removed. Second, cell cycle phases were inferred using the function *CellCycleScoring* from Seurat. Third, guide RNAs were assigned to cells using the guide RNA matrix provided by cellranger. In the case of multiple detected guide RNAs per cell, guides matching ≥75% of reads per cell were used. If no guide matched ≥75% of reads in a cell, this cell was left unassigned. Fourth, data were aligned across samples and cell cycle effects were removed using the function *align_cds* from Monocle3 (version 0.2.3.0)[64]. Finally, UMAP projection and clustering was performed using the functions *reduce_dimension* and *cluster_cells* from Monocle3.

Cell type assignment (in vivo and ex vivo). Cell types were predicted using the package singleR (version 1.4.1), based on a dataset from Izzo and colleagues[56] and a dataset from the packages CytoTRACE (version 0.3.3)[65]. SingleR[66] was run using the Wilcoxon method for differential analysis. Cells in clusters with more than 80% of cells predicted as granulocytes, granulocytes progenitors, or immature B cells in the bone marrow dataset from CytoTRACE were assigned based on this dataset. All other clusters were assigned based on the predictions by Izzo et al[45]. Eosinophils and basophils were combined in one label. Cells predicted as erythrocytes were further split into MEPs, Erythroid Progenitors or Erythrocytes based on a comparison of gene signatures with external datasets[67] and on the expression of key marker genes: 1) Low *Gata2*/High *Gata1, Epor* & *Klf1* marked the transition from MEPs to Erythroid progenitors 2) Induction of *Hba-a1* and *Hbb-b1*, increased *Tfrc* expression, enrichment of S and G2 cell-cycle signatures mark the transition from Erythroid Progenitors to Erythrocytes. Finally, cells co-expressing high levels of marker genes from the two distinct mature lineages (Erythroid and Myeloid) were removed as likely doublets or cells with contamination of ambient RNA .

Next, cells constituting <10% of a cluster were re-assigned to the majority in each cluster. MEPs with *Gata2* expression greater than *Gata1* expression were labelled as early MEPs. MEPs clusters with strong cell cycle phase signatures were labelled accordingly. A cluster of MEPs harboring predominantly *Rcor1* KOs was labelled as "MEP (perturbed)".

Cell type enrichment analyses. To test differences in distributions of KOs and NTCs, we tested the enrichment of KOs compared to NTCs within each cell type. Clusters with less than 5 NTCs or <25% NTCs were removed from this analysis. Fisher's exact test was used with the function *fisher.test*. Enrichment was tested against each NTC separately. P-values were adjusted using the function *p.adjust* with the method "BH".

Viability analysis. Counts of cells harboring KOs at day 14 were transformed to counts per million (CPMs) and normalized to the number of cells harboring NTCs. The thus normalized cell counts were compared to the equally normalized read counts in the guide RNA pool before cell infection, resulting in a log fold change that represents viability. Thus, negative or positive values represent an enrichment or loss, respectively, of KO-harboring cells at day 14 (after cell infection) relative to NTCs.

Cross-projection of ex vivo and leukemia samples to in vivo data. To project *ex vivo* and leukemia samples onto the *in vivo* data, we adapted the ProjecTILs algorithm (version 2.0.2)[68] predicting UMAP coordinates for each cell from the ex vivo and leukemia samples based on the UMAP coordinates of *in vivo* cells using a k-nearest-neighbor approach with (k=20 neighbors).

Differential expression analysis. We performed differential expression comparing cells with CF-KOs to NTCs using nebula (version 1.1.8). We removed cluster with less than 31 cells and genes with less than 21 reads. We ran nebula[69] with default parameters, testing differences of KOs to NTCs with fixed effects (parameter "pred") and adding sample information as random effects (parameter "id"). For genes, where the algorithm did not converge, we reran nebula with the "NBLMM" model. P-values were corrected for multiple testing using the function *p.adjust* with the method "BH". Gene set enrichment was performed using the function *fgsea* from the *fgsea* package[70].

Pseudotime trajectory analysis. Pseudotime for the different lineage trajectories were identified using diffusion maps[71] applied to the non-targeting control population, using Scanpy (version 1.9.1)[72] functions *tl.diffmap* and *tl.dpt*. Perturbed cells were then mapped to their nearest k=15 non-targeting cells in PCA space, considering the first N=8 principal components, and then assigned the mean pseudotime value across these cells. Principal component analysis cut offs were found via elbow plots through assessing the variance accounted for in the first N principal components. Each branch was extracted for separate analysis using the aforementioned cell labels (see Figure 2a) and pseudotime was scaled to the unit interval. We plot perturbations with crucial

biological significance, which incidentally show visually striking distribution differences compared to the non-targeting cell population.

CITE-seq read counts obtained from cellranger were normalized to log counts per million (CPMs) and then scaled. Perturb-seq data in leukemia were processed in the same way as the in vivo data, as described above. Enrichment analyses were performed on clusters instead of cell types.

Chromatin Accessibility Analysis of Chromatin Factor Knockouts

Isolation of progenitors, CRISPR loss-of-function and *ex vivo* differentiation.

20,000 Cas9-LSK cells were transduced with Lenti-CRISPR-BFP virus, expressing the top performing sgRNA against each CF and, cultured for 48h under multipotent conditions (detailed above). Then, cells were stimulated with cytokine cocktails for lineage-priming or myeloid differentiation for 5 days. For the time-course experiment, cells were perturbed and immediately grown for 3, 5 and 7 days under lineage priming or myeloid conditions. Finally, the CRISPR edited progeny was FACS-Sorted (BFP+, GFP+) and harvested by centrifugation for ATAC-seq.

Fast ATAC-seq was perfomed as described in Corces *et al*[73]. ATAC-seq libraries were sequenced at 50 million reads on a NextSeq 1000.

Data processing and analysis.

Based on the ATAC-seq nf-core pipeline (Patel et al, 2021) we ran Trim Galore (version 0.6.6)[74] with Cutadapt (version 3.4)[75] using the default parameters to trim low-quality and adaptor sequences. We then aligned these reads to the GRCm38/mm10 reference genome assembly with Decoy sequences using Bowtie2 (version 2.3.4.2)[76] with parameters -X 1000 --no-discordant --no-mixed --very-sensitive. Subsequently, we removed duplicated regions with Picard (version 2.25.4) (Broad Institute, G. R. Picard Toolkit. https://broadinstitute.github.io/picard/ (2019)), non-interesting chromosomes (chrM, chrUn, …), and blacklist regions included in the ENCODE blacklist (version 2.0) (*65*). Finally, we removed Tn5 adaptors with alignmentSieve's (version 3.5.1)[77] –ATACshift parameter and indexed the final BAM files with SAMtools (version 1.3.1)[67]. These BAM files were then processed to CPM scaled Bigwig files with bamCoverage (version 3.5.1)[78].

To identify the ATAC peaks, we pooled replicates and converted the paired BAM files to single-read bed format using function bamToBed from BEDTools (version 2.27.1)[79-80]. Then, we used

MACS (version 2.2.7.1)[81] with parameters –broad -f BED --keep-dup all --nomodel --shift -75 --extsize 150 to call peaks. To compare peak strength between conditions, we generated a unified peak set for all experiments, ending with, respectively, 376,658 and 207,724 peaks in LSKs and Npm1c/Flt3-ITD murine leukemic cells (DM- Double Mutant), respectively. We then annotated these consensus peaks with the function *annotatePeaks* from HOMER (version 4.10)[82] counted the reads on them with featureCounts (version 2.0.1)[83] and calculated adjusted CPM values with edgeR's (version 3.34.1) TMM method[84]. Additionally, we used DESeq2 (version 1.32.0)[72] to measure the Fold Change (FC) between conditions, defining the peaks with absolute log2(FC) values greater than 0.75 and adjusted p-values lower than 0.01 as differentially enriched. Finally, we filtered out peaks with less than 2 CPM and/or 10 reads in the compared two conditions (KO vs. WT).

Motif analysis.

To look for differential TF motifs enrichment between KO and WT, we used TOBIAS (version 0.13.2)[28]. Following the program guidelines, we generated a consensus set of peaks with all the peaks called previously in both the control and the compared KO sample. We then renamed and formatted HOMER's list of vertebrate known.motifs to make it suitable for TOBIAS. After generating the TF footprint bigwigs (using the function *ATACorrect* with parameters --read_shift 0 0 and the function *ScoreBigwig* with default parameters), we computed the differentially bound motifs with the function *BINDetect*.

Additionally, we used the output TF binding coordinates of each of the motifs to measure the gain/loss of TFs union to chromatin for each chromatin factor KO.

ChIP-seq analysis of Normal and Leukaemic Populations

Isolation of *in vivo* hematopoietic cells from bone marrow.

Bone marrow cells were harvested from 12-14 week old C57BL6 mice as described above and stained for the isolation of:

GMP: Lineage (CD3, CD19, CD11b, Gr1, Ter119, B220)-, cKit+, Sca-1-, FcgRIII+, CD34+.

MEP: Lineage (CD3, CD19, CD11b, Gr1, Ter119, B220)-, cKit+, Sca-1-, FcgRIII-, CD34-.

Monocytes: CD3-, CD19-, Ter119-, CD11b+

B-cell: CD3-, CD19+, Ter119-, CD11b-

Erythroid cells were FACS-sorted from spleens of 12 week-old C57BL6 mice as: CD3-, CD19-, Ter119+, CD11b-, Gr1-.
Dilution 1:100 for all antibodyes, except 1:50 for CD34
Cells were sorted in PBS + 0.1% BSA and crosslinked immediately after sorting.
FACS antibodies in Supplementary Table 9.

Isolation of leukemic cells.

Npm1c/Flt3-ITD/Cas9 double mutant (DM) cells were generated from lineage-depleted, bone marrow cells of primary transgenic mice post-leukemic onset (female, 12 weeks old) as previously described[57-58]. Cells were maintained in XVIVO-20 medium (Lonza) supplemented with 5% Fetal Bovine Serum (FBS) (ThermoFisherScientific), 1% PSG (Gibco), mSCF 50 ng/mL (PeproTech), mIL-3 10 ng/mL (PeproTech) and mIL-6 10 ng/mL (R&D Systems), in a 37ºC and 5% CO_2 atmospheric environment. *Npm1c/Flt3-ITD/Cas9* double mutant (DM) cells were passaged every 2 days and cultured for short time (passage 3-5) to maintain the original leukemic properties.

Crosslinking.

Freshly sorted normal cell or early passage (3-5) leukemic cells were crosslinked at room temperature with 3 mM EGS, DSG, and DMA for 20 min followed by 1% Formaldehyde for another 5 minutes. Then, Glycine was added to 125 mM and incubated for 5 minutes to quench the crosslinkers. Finally, cells were pelleted at 750xG for 7 minutes, washed twice with cold 0.5% BSA/PBS containing 1X Protease Inhibitors (Roche) and flash frozen at -80C.

Chromatin Immuno-precipitation.

Crosslinked cells were thawed and resuspended in 1.5 mL of ice-cold Cell Lysis Buffer (10 mM HEPES pH 7.5, 10 mM NaCl, 0.2% IGEPAL) + Protease Inhibitors for 10 minutes on ice. Then, nuclei were pelleted at 5,000 g for 7 minutes, resuspended in Sonication buffer (0.5% SDS, 5mM EDTA) and pelleted again at 8,000 g, resuspended in 50-100 uL of Sonication buffer and sonicated for 5 cycles (30s ON, 30s OFF) in a Bioruptor Nano. Then, chromatin extracts were diluted in 4 volumes of ChIP dilution buffer (25 mM HEPES, 185 mM NaCl, 1.25% Tx-100 + Protease Inhibitors) and incubated with the relevant antibodies (Supplementary Table 10) at 4C for 10-12 hours. The following day, 25 uL Magna ChIP Prot A+G were added and incubated for 3h at 4C. Bead-bound chromatin was washed 2X with RIPA, 2X with RIPA-500, 2X with LiCl buffer and

1X with TE buffer. ChIPed DNA was reversed crosslinking by 30 min incubation with 2 uL of Proteinase K in 50 ul of ChIP Elution buffer (10 mM Tris-Cl pH 8, 300 mM NaCl, 0.2 mM EDTA, 0.4% SDS) at 55ºC followed by 1-hour incubation at 68 ºC. Finally, the ChIPed DNA was purified with a 2.2X SPRI Cleanup and quantified using the Qubit dsDNA High Sensitivity Assay Kit.
Every ChIP-seq experiment was performed in replicates except for Kmt2d and Kmt2a in early myeloid (GMP) and erythroid (MEP) progenitors. Alla attempts of replication were successful except for a failed Smarcb1 ChIP-seq experiment in MEPs, which was removed from the analysis and substituted by a third ChIP-seq experiment to reach n=2.

Preparation of ChIP-seq libraries.

ChIP-seq libraries were prepared from 0.5-10 ng of ChIPed DNA using the NEB Next Ultra II kit (New England Biolabs). ChIP-seq libraries were sequenced to 100 million reads per sample.

ChIP-seq data processing and analysis.

Based on the ChIP-seq nf-core pipeline[85] we first processed the FASTQ files to BAM files as described for ATAC-Seq skipping the Tn5 adapter removal (Statistics for each ChIP-seq experiment detailed in Supplementary Table 11). Next, to identify peaks for each sample, we pooled replicates and used MACS with parameters -f BAMPE --keep-dup all. To compare peak strength between cell types, we generated a unified peak set per CF (Brd9, Kmt2a, Kmt2d, and Smarcb1). We then followed the steps explained in the ATAC-Seq data processing and analysis section to annotate the peaks and calculate the CPM reads on them. Finally, we measured the FC between cell types using DESeq2[84] to get peaks with an absolute log2(FC) greater than 0.75 and an adjusted p-value lower than 0.01.

Motif analysis in ChIP-seq peaks.

We first generated a list of cell-type-specific peak coordinates for each of the analysed CFs. To do so, we selected all peaks that were significantly enriched or depleted in pairwise comparisons of GMP, Myeloid, MEP, Ery, and B-cell on each of the CFs. Then we clustered and manually curated these coordinates to get a list of cell-type-specific peaks per CF. Enrichment of transcription factors in cell-type-specific peak coordinates was then analyzed using the function *findMotifsGenome* from HOMER (*69*). For each CF, cell-type specific peaks were compared to all

peaks found across all five cell types and all four CFs as background. Motif enrichment analyses were centered on the 100 bp surrounding the peak summit.

Comparison of normal and leukemic patterns.

To identify transcription factor switches in leukemia, we defined subsets of peaks that were gained in DM AML (Leukemic), shared in DM and GMP (Common), and not present in DM but present in GMP and Monocytes (Normal). Gained peaks were defined by log2(FC) values greater than 1, lost peaks by log2(FC) values smaller than -1, and shared peaks by absolute log2(FC) values lower than 0.5. The subsetting was done per CF in a consensus peak dataset of Smarcb1, Kmt2a, and Kmt2d. Similar to Motif analysis in ChIP-seq peaks section, we looked for enrichment of transcription factors in each of the subsets using the function *findMotifsGenome* from HOMER, and all peaks from the consensus (across all subsets) were used as the background set.

To measure Stat5a binding signal over the CF-bound sites we collected Stat5a ChIP-seq sequencing data at the same consensus coordinates stated right above and computed the CPM values as described in ATAC-Seq data processing and analysis section.

Functional assays in Leukemia

Cell transduction and sorting of differentiated populations.

Npm1/Flt3-ITD/Cas9 DM murine cells were transduced using a Retronectin-transduction protocol (Takara Bio) with Lenti-PerturbSeq-BFP vectors targeting *Smarcb1, Smarcd2, Brd9* and *Smarcd1*, *Kmt2a, Kmt2d, Wdr82*, *Hdac3, Setdb1*, *Stat5a* or non-targeting control (NTC). Then, growth was monitored by flow cytometric assessment (BD LSR Fortessa II; BD Bioscience) of BFP+ cells in culture at 2, 6, 9 and 13 days post-transduction. Fold change in BFP+ cells at each time point relative to the proportion of BFP+ cells at day 2 was calculated. Immunophenotypic analysis was performed by flow cytometry (BD LSR Fortessa II; BD Bioscience) at 6 days by staining with anti-CD11b, anti-Ly6G/Ly6C (Gr1) and anti-CD55 (Supplementary Table 9). Proportions of Gran-like (CD11b-high/Gr-1+) and Ery/Baso-like (CD55-high) were quantified with FlowJo (Tree Star; v10.8.1). P-values were calculated using a two-way ANOVA statistical test or ratio-paired t-test (Prism; GraphPad software; v9.1).

Clonogenic and Cell proliferation assays of DM AML cells.

$2x10^6$ DM murine leukemic cells were stained for CD11b, Gr-1, CD55, CD41 and CD34 (Supplementary Table 9). Gran-like (CD11b-high/Gr-1+); Ery/Baso-like (CD55-high/CD41-) and CD34+ fractions were subsequently FACS-sorted (BD Influx; BD Bioscience). 1000 cells of each sorted population were seeded in 1 mL of methylcellulose medium (M3434, STEMCELL Technologies) supplemented with recombinant mSCF and mIL-3 with recombinant IL-6 and Epo in duplicate. Methylcellulose cultures were maintained at 37 ^oC and 5 % CO_2, and total colony forming units (CFU) enumerated 7 days later. Photographs of colonies were obtained on the STEMvision instrument and software (STEMCELL Technologies). Data is presented as a fold change of average CFU per 1000 cells seeded, relative to the CD34+ fraction. 10,000 Gran-like, Ery/Baso-like or CD34+ DM cells were also maintained in standard culture conditions for 21 days, and number of cells counted every 7 days. Total cell number is presented as a fold to corresponding CD34+ counterpart for each timepoint. P-values were calculated using a two-way ANOVA statistical test (Prism; GraphPad software; v9.1).

Statistics & Reproducibility

No statistical method was used to predetermine sample size. Lineage scores (bulk CRISPR screens) of the top 200 hits were validated in replicate screens. Likewise, Perturb-seq was performed in replicate for the top 40 chromatin-factors. Epigenetic profiling (ATAC-seq and ChIP-seq) was performed in two biologically independent replicates (except for Kmt2d and Kmt2a ChIP-seq in MEPs) following a common practice in the field. Growth curves of chromatin factor KOs were calculated from 3-4 independent experimental batches. Immuno-phenotypic patterns derived from chromatin-factor perturbation were replicated at least in 2 biologically independent experiments.

Data exclusion: No data points were excluded from any of the analysis except for one ChIP-seq experiment (Smarcb1 in MEPs) was excluded for analysis due to low signal-to-noise. This data point was substituted by another ChIP-seq experiment to reach n=2.

Randomization: Cell-based assays (Screens and validations) and mouse allocation was randomized a proper batch designs were ensured to avoid confounding effects. In the Perturb-seq analysis we used several Non-targeting controls across all experimental batches.

The Investigators were not blinded to allocation during experiments and outcome assessment.

Ethical compliance.

Murine ethical compliance was fulfilled under the Guidelines of the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at University of Navarra, Spain, and the Animal Welfare Ethical Review Body at the University of Cambridge, UK. Research in the UK was conducted under Home Office license PP3042348.

DATA AVAILABILITY

Bulk Expression patterns of hematopoietic populations: Lara-Astiaso *et al,* Science 2012 (*52*), GEO accession (GSE60103)

Single-cell expression patterns of hematopoiesis: Izzo *et al,* Nat Genetics 2020 (*45*), GEO accession (GSE124822)

Perturb-seq datasets: *in vivo, ex vivo* and leukemic: GEO accession (GSE213511)

Chromatin accessibility of CF-KOs: GEO accession (GSE213506)

ChIP-seq datasets of CFs *in vivo*, *ex vivo* and leukemic: GEO accession (GSE213507)

CODE AVAILABILITY

Analysis code is available in a dedicated GitHub repository https://github.com/csbg/tfcf[68].

METHODS-ONLY REFERENCES

56. Izzo, F et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat Genet* **52**, 378-387 (2020)
57. Mupo, A.et al. A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice. *Leukemia* **27**, 1917-1920 (2013)
58. Vassiliou, G.S. et al. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet* 43, 470-475 (2011)
59. Wilkinson, AC. et al. Long-term ex vivo haematopoietic stem cell expansion allows nonconditioned transplantation. *Nature* **571**, 117-121 (2019)
60. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010)
61. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi ,W. & Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res* **43**, e47 (2015)
62. Replogle, J.M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* **38**, 954–961 (2020)

63. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021)
64. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014)
65. Gulati, G.S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411 (2020)
66. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019)
67. Lara-Astiaso, D. et al. Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943-949 (2014)
68. Andreatta, M. et al. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun* **12**, 2965 (2021)
69. He, L. et al. NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. *Commun Biol*. **4,** 629 (2021)
70. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at https://www.biorxiv.org/content/10.1101/060012v3 (2021)
71. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845–848 (2016)
72. Wolf, F., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19,** 15 (2018)
73. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203 (2016)
74. Krueger, F., James, F.. Ewels, P., Afyounian, E. & Schuster-Boeckler, B. FelixKrueger/TrimGalore: v0.6.7 (2021) https://doi.org/10.5281/zenodo.5127899
75. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011)
76. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **4**, 357–359 (2012)
77. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019)
78. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016)
79. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021)
80. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010)
81. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, 1–9 (2008)
82. Heinz, S. et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**, 576–589 (2010)
83. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014)
84. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014)
85. Wolf, E. et al. Miz1 is required to maintain autophagic flux. *Nat Commun* **4**, 2535 (2013)
86. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019)
87. Analysis Code: https://doi.org/10.5281/zenodo.7891703

Fig. 1

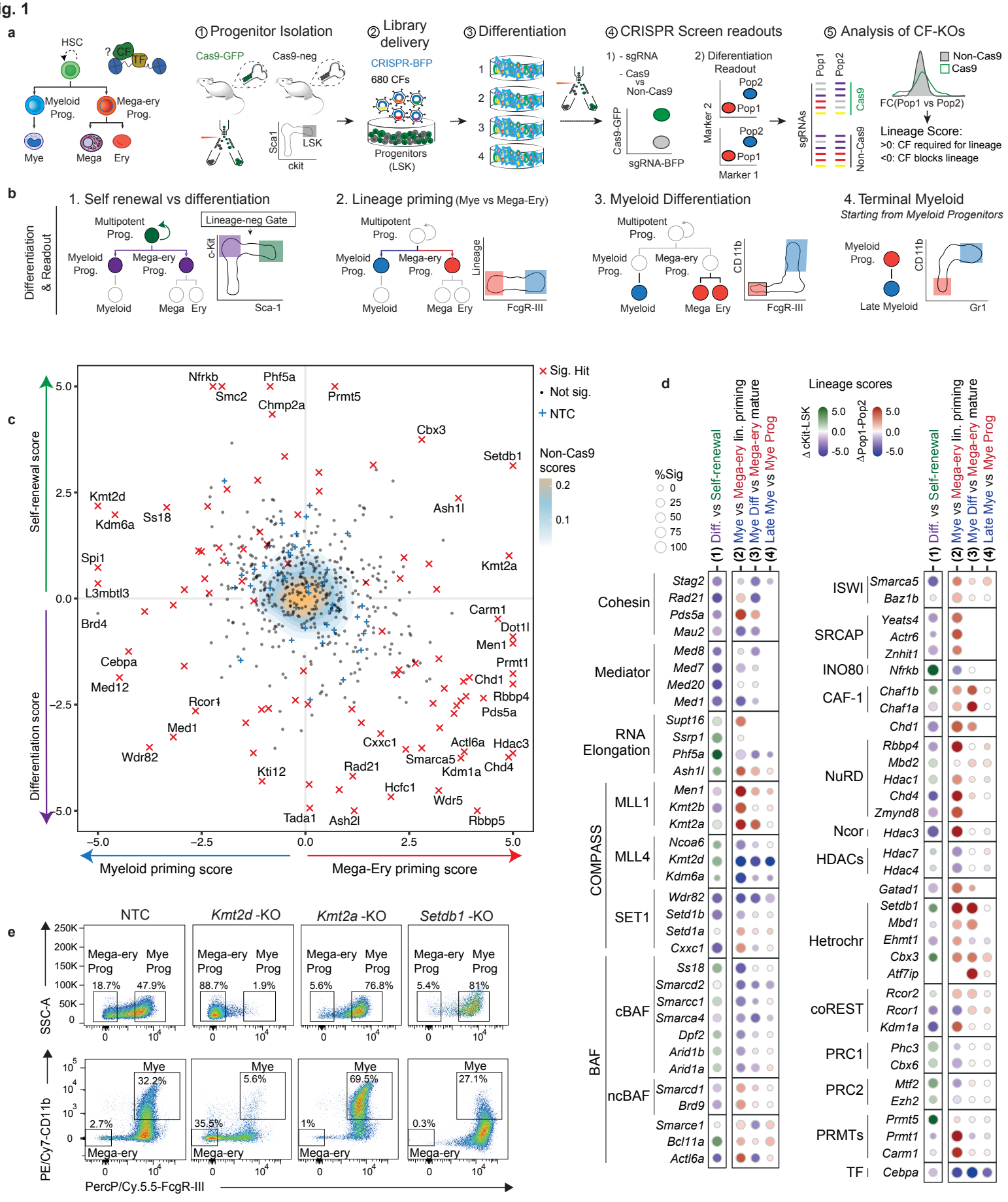

a
HSC
Myeloid Prog.
Mega-ery Prog.
Mye
Mega
Ery
① Progenitor Isolation
Cas9-GFP
Cas9-neg
Sca1
LSK
ckit
② Library delivery
CRISPR-BFP
680 CFs
Progenitors (LSK)
③ Differentiation
④ CRISPR Screen readouts
1) - sgRNA
- Cas9 vs Non-Cas9
Cas9-GFP
sgRNA-BFP
2) Diferentiation Readout
Marker 2
Marker 1
Pop1
Pop2
⑤ Analysis of CF-KOs
sgRNAs
Cas9
Non-Cas9
FC(Pop1 vs Pop2)
Non-Cas9
Cas9
Lineage Score:
>0: CF required for lineage
<0: CF blocks lineage
b
Differentiation & Readout
1. Self renewal vs differentiation
Multipotent Prog.
Lineage-neg Gate
c-Kit
Sca-1
2. Lineage priming (Mye vs Mega-Ery)
Lineage
FcgR-III
3. Myeloid Differentiation
CD11b
4. Terminal Myeloid
Starting from Myeloid Progenitors
Late Myeloid
Gr1
c
Self-renewal score
Differentiation score
Myeloid priming score
Mega-Ery priming score
Sig. Hit
Not sig.
NTC
Non-Cas9 scores
Nfrkb
Phf5a
Smc2
Chmp2a
Prmt5
Cbx3
Setdb1
Kmt2d
Kdm6a
Ss18
Ash1l
Spi1
Kmt2a
L3mbtl3
Brd4
Carm1
Dot1l
Men1
Cebpa
Prmt1
Med12
Chd1
Rcor1
Rbbp4
Pds5a
Med1
Wdr82
Cxxc1
Actl6a
Hdac3
Smarca5
Kti12
Rad21
Kdm1a
Chd4
Hcfc1
Wdr5
Tada1
Ash2l
Rbbp5
d
%Sig
Lineage scores
Δ cKit-LSK
ΔPop1-Pop2
(1) Diff. vs Self-renewal
(2) Mye vs Mega-ery lin. priming
(3) Mye Diff vs Mega-ery mature
(4) Late Mye vs Mye Prog
Cohesin
Stag2
Rad21
Pds5a
Mau2
Mediator
Med8
Med7
Med20
Med1
RNA Elongation
Supt16
Ssrp1
Phf5a
Ash1l
COMPASS
MLL1
Men1
Kmt2b
Kmt2a
MLL4
Ncoa6
Kmt2d
Kdm6a
SET1
Wdr82
Setd1b
Setd1a
Cxxc1
BAF
cBAF
Ss18
Smarcd2
Smarcc1
Smarca4
Dpf2
Arid1b
Arid1a
ncBAF
Smarcd1
Brd9
Smarce1
Bcl11a
Actl6a
ISWI
Smarca5
Baz1b
SRCAP
Yeats4
Actr6
Znhit1
INO80
Nfrkb
CAF-1
Chaf1b
Chaf1a
Chd1
NuRD
Rbbp4
Mbd2
Hdac1
Chd4
Zmynd8
Ncor
Hdac3
HDACs
Hdac7
Hdac4
Gatad1
Hetrochr
Setdb1
Mbd1
Ehmt1
Cbx3
Atf7ip
coREST
Rcor2
Rcor1
Kdm1a
PRC1
Phc3
Cbx6
PRC2
Mtf2
Ezh2
PRMTs
Prmt5
Prmt1
Carm1
TF
Cebpa
e
NTC
Kmt2d -KO
Kmt2a -KO
Setdb1 -KO
Mega-ery Prog
Mye Prog
18.7%
47.9%
88.7%
1.9%
5.6%
76.8%
5.4%
81%
SSC-A
PE/Cy7-CD11b
PercP/Cy.5.5-FcgR-III
Mye
Mega-ery
32.2%
2.7%
5.6%
35.5%
69.5%
1%
27.1%
0.3%

Fig. 2

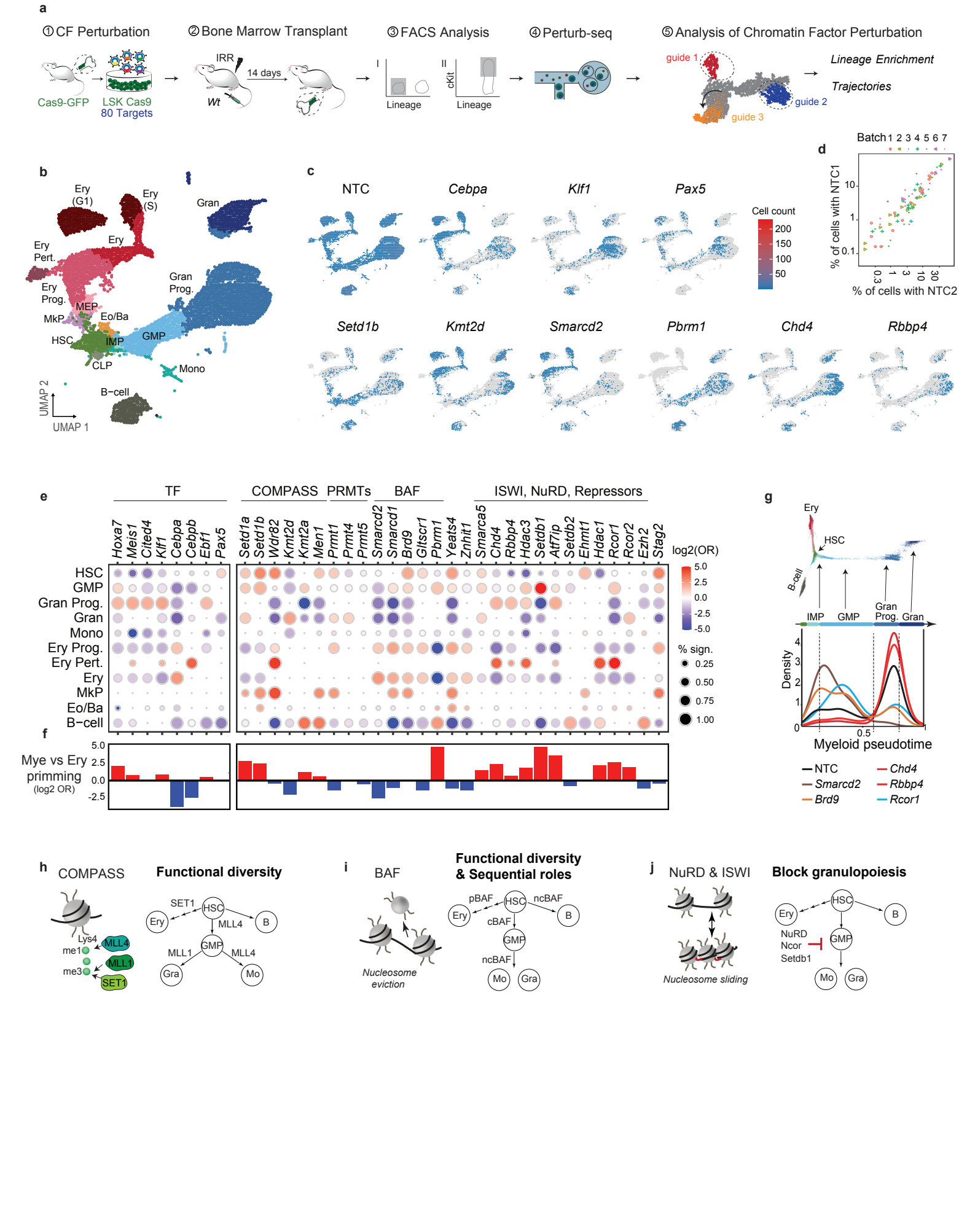

a
① CF Perturbation
② Bone Marrow Transplant
③ FACS Analysis
④ Perturb-seq
⑤ Analysis of Chromatin Factor Perturbation
Cas9-GFP
LSK Cas9
80 Targets
IRR
14 days
Wt
I
II
cKit
Lineage
Lineage
guide 1
guide 2
guide 3
Lineage Enrichment
Trajectories
b
Ery (G1)
Ery (S)
Gran
Ery
Ery Pert.
Gran Prog.
Ery Prog.
MEP
MkP
Eo/Ba
HSC
IMP
GMP
CLP
Mono
B−cell
UMAP 2
UMAP 1
c
NTC
Cebpa
Klf1
Pax5
Cell count
200
150
100
50
Setd1b
Kmt2d
Smarcd2
Pbrm1
Chd4
Rbbp4
d
Batch 1 2 3 4 5 6 7
% of cells with NTC1
% of cells with NTC2
10
1
0.1
0.3
1
3
10
30
e
TF
COMPASS
PRMTs
BAF
ISWI, NuRD, Repressors
Hoxa7
Meis1
Cited4
Klf1
Cebpa
Cebpb
Ebf1
Pax5
Setd1a
Setd1b
Wdr82
Kmt2d
Kmt2a
Men1
Prmt1
Prmt4
Prmt5
Smarcd2
Smarcd1
Brd9
Gltscr1
Pbrm1
Yeats4
Znhit1
Smarca5
Chd4
Rbbp4
Hdac3
Setdb1
Atf7ip
Setdb2
Ehmt1
Hdac1
Rcor1
Rcor2
Ezh2
Stag2
HSC
GMP
Gran Prog.
Gran
Mono
Ery Prog.
Ery Pert.
Ery
MkP
Eo/Ba
B−cell
log2(OR)
5.0
2.5
0.0
-2.5
-5.0
% sign.
0.25
0.50
0.75
1.00
f
Mye vs Ery primming (log2 OR)
5.0
2.5
0.0
-2.5
g
Ery
HSC
B-cell
IMP
GMP
Gran Prog.
Gran
Density
Myeloid pseudotime
NTC
Chd4
Smarcd2
Rbbp4
Brd9
Rcor1
h
COMPASS
Functional diversity
Lys4
me1
me3
MLL4
MLL1
SET1
SET1
HSC
MLL4
Ery
B
GMP
MLL1
MLL4
Gra
Mo
i
BAF
Functional diversity & Sequential roles
Nucleosome eviction
pBAF
ncBAF
HSC
Ery
B
cBAF
GMP
ncBAF
Mo
Gra
j
NuRD & ISWI
Block granulopoiesis
Nucleosome sliding
HSC
Ery
B
NuRD
Ncor
Setdb1
GMP
Mo
Gra

Fig. 3

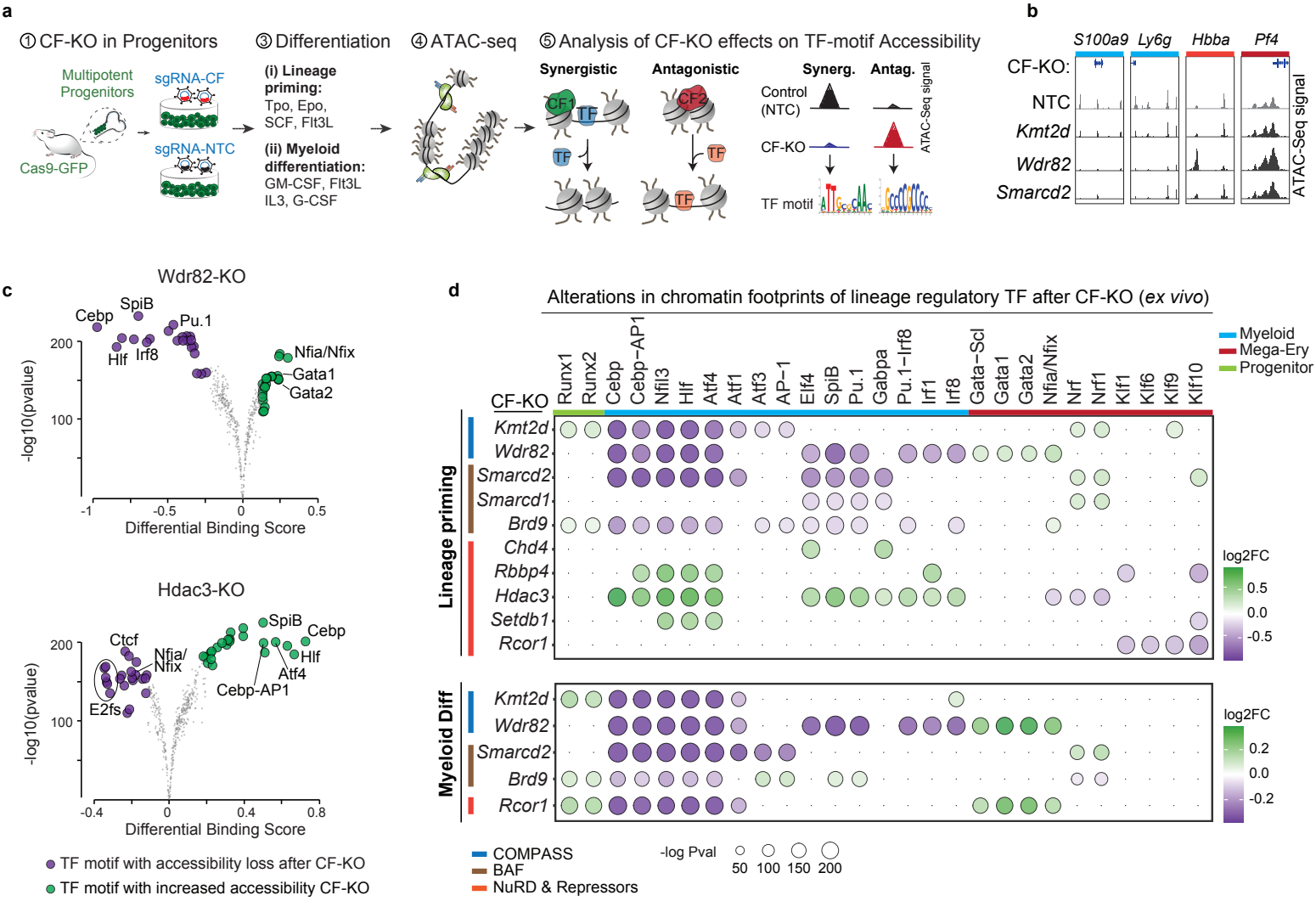

a
① CF-KO in Progenitors
Multipotent Progenitors
sgRNA-CF
sgRNA-NTC
Cas9-GFP
③ Differentiation
(i) Lineage priming:
Tpo, Epo, SCF, Flt3L
(ii) Myeloid differentiation:
GM-CSF, Flt3L IL3, G-CSF
④ ATAC-seq
⑤ Analysis of CF-KO effects on TF-motif Accessibility
Synergistic
Antagonistic
CF1
CF2
TF
Synerg.
Antag.
Control (NTC)
CF-KO
TF motif
ATAC-Seq signal
b
S100a9
Ly6g
Hbba
Pf4
CF-KO:
NTC
Kmt2d
Wdr82
Smarcd2
ATAC-Seq signal
c
Wdr82-KO
Cebp
SpiB
Pu.1
Hlf
Irf8
Nfia/Nfix
Gata1
Gata2
-log10(pvalue)
Differential Binding Score
Hdac3-KO
Ctcf
Nfia/
Nfix
E2fs
SpiB
Cebp
Hlf
Atf4
Cebp-AP1
TF motif with accessibility loss after CF-KO
TF motif with increased accessibility CF-KO
d
Alterations in chromatin footprints of lineage regulatory TF after CF-KO (ex vivo)
CF-KO
Runx1
Runx2
Cebp
Cebp−AP1
Nfil3
Hlf
Atf4
Atf1
Atf3
AP−1
Elf4
SpiB
Pu.1
Gabpa
Pu.1−Irf8
Irf1
Irf8
Gata−Scl
Gata1
Gata2
Nfia/Nfix
Nrf
Nrf1
Klf1
Klf6
Klf9
Klf10
Myeloid
Mega-Ery
Progenitor
Lineage priming
Kmt2d
Wdr82
Smarcd2
Smarcd1
Brd9
Chd4
Rbbp4
Hdac3
Setdb1
Rcor1
log2FC
0.5
0.0
-0.5
Myeloid Diff
Kmt2d
Wdr82
Smarcd2
Brd9
Rcor1
log2FC
0.2
0.0
-0.2
COMPASS
BAF
NuRD & Repressors
-log Pval
50
100
150
200

Fig. 4

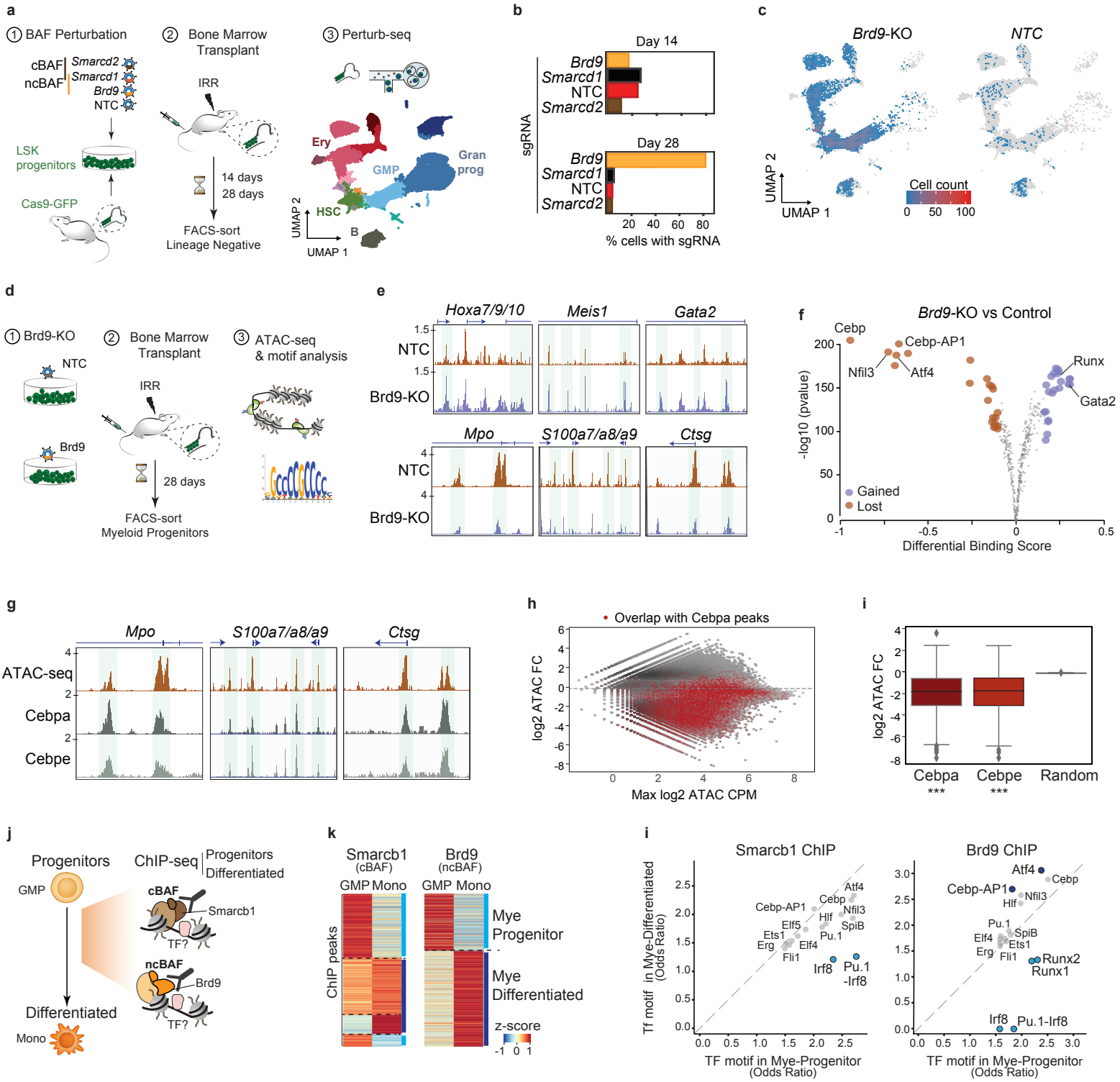

a
① BAF Perturbation
cBAF
ncBAF
Smarcd2
Smarcd1
Brd9
NTC
LSK progenitors
Cas9-GFP
② Bone Marrow Transplant
IRR
14 days
28 days
FACS-sort Lineage Negative
③ Perturb-seq
Ery
GMP
Gran prog
HSC
B
UMAP 2
UMAP 1
b
Day 14
Day 28
sgRNA
Brd9
Smarcd1
NTC
Smarcd2
20 40 60 80
% cells with sgRNA
c
Brd9-KO
NTC
UMAP 2
UMAP 1
Cell count
0 50 100
d
① Brd9-KO
NTC
Brd9
② Bone Marrow Transplant
IRR
28 days
FACS-sort Myeloid Progenitors
③ ATAC-seq & motif analysis
e
Hoxa7/9/10
Meis1
Gata2
NTC
Brd9-KO
Mpo
S100a7/a8/a9
Ctsg
f
Brd9-KO vs Control
Cebp
Cebp-AP1
Nfil3
Atf4
Runx
Gata2
-log10 (pvalue)
Gained
Lost
Differential Binding Score
g
Mpo
S100a7/a8/a9
Ctsg
ATAC-seq
Cebpa
Cebpe
h
Overlap with Cebpa peaks
log2 ATAC FC
Max log2 ATAC CPM
i
log2 ATAC FC
Cebpa
Cebpe
Random
***
***
j
Progenitors
GMP
Differentiated
Mono
ChIP-seq
Progenitors
Differentiated
cBAF
Smarcb1
TF?
ncBAF
Brd9
TF?
k
Smarcb1 (cBAF)
Brd9 (ncBAF)
GMP Mono
GMP Mono
ChIP peaks
Mye Progenitor
Mye Differentiated
z-score
-1 0 1
i
Smarcb1 ChIP
Brd9 ChIP
Tf motif in Mye-Differentiated (Odds Ratio)
TF motif in Mye-Progenitor (Odds Ratio)
Atf4
Cebp
Cebp-AP1
Nfil3
Hlf
Elf5
SpiB
Ets1
Pu.1
Erg
Elf4
Fli1
Irf8
Pu.1 -Irf8
Runx2
Runx1

Fig. 5

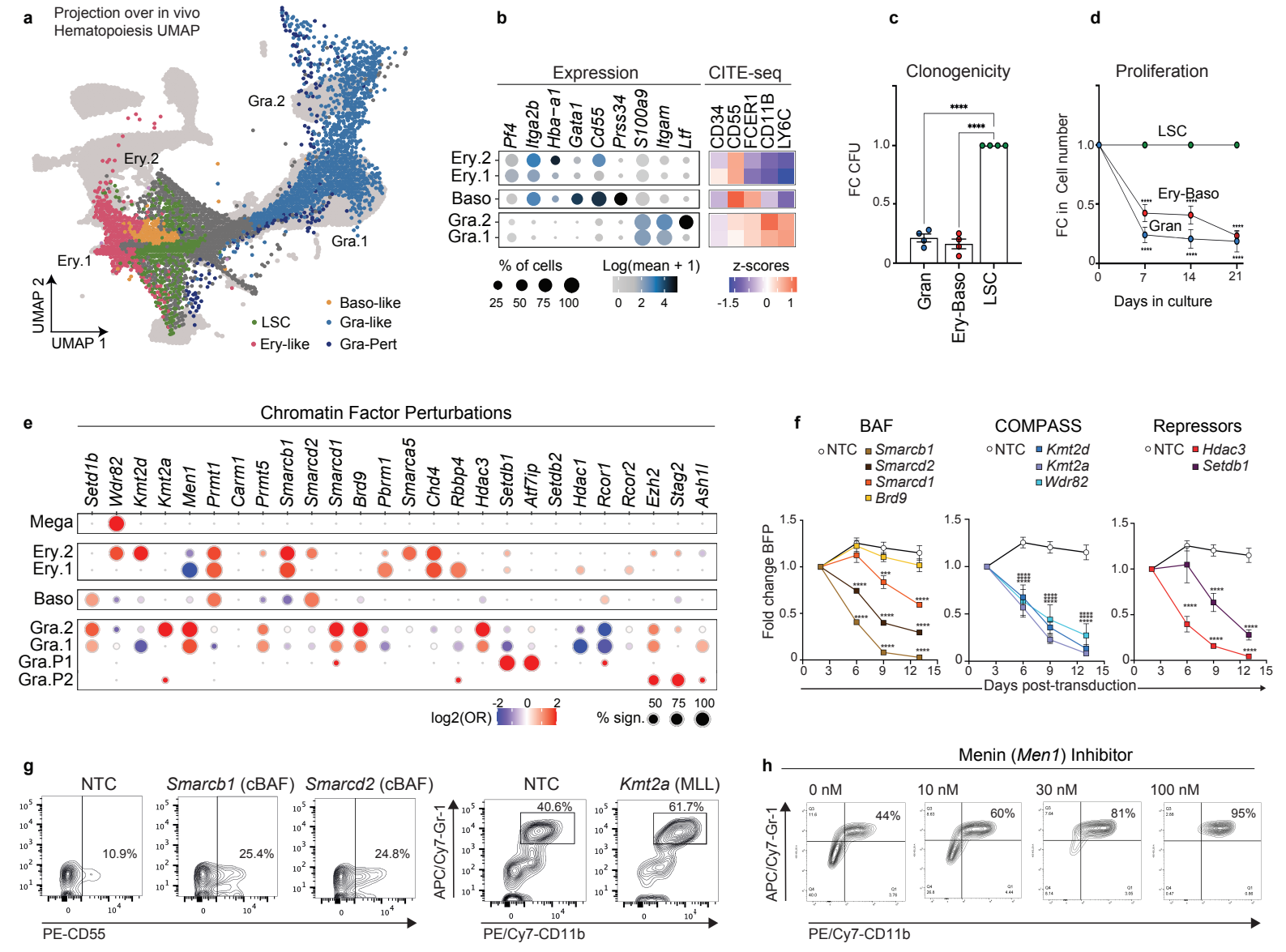

a
Projection over in vivo Hematopoiesis UMAP
Gra.2
Ery.2
Gra.1
Ery.1
UMAP 2
UMAP 1
LSC
Ery-like
Baso-like
Gra-like
Gra-Pert
b
Expression
CITE-seq
Pf4
Itga2b
Hba−a1
Gata1
Cd55
Prss34
S100a9
Itgam
Ltf
CD34
CD55
FCER1
CD11B
LY6C
Ery.2
Ery.1
Baso
Gra.2
Gra.1
% of cells
Log(mean + 1)
z-scores
c
Clonogenicity
FC CFU
Gran
Ery-Baso
LSC
d
Proliferation
FC in Cell number
LSC
Ery-Baso
Gran
Days in culture
e
Chromatin Factor Perturbations
Setd1b
Wdr82
Kmt2d
Kmt2a
Men1
Prmt1
Carm1
Prmt5
Smarcb1
Smarcd2
Smarcd1
Brd9
Pbrm1
Smarca5
Chd4
Rbbp4
Hdac3
Setdb1
Atf7ip
Setdb2
Hdac1
Rcor1
Rcor2
Ezh2
Stag2
Ash1l
Mega
Ery.2
Ery.1
Baso
Gra.2
Gra.1
Gra.P1
Gra.P2
log2(OR)
% sign.
f
BAF
COMPASS
Repressors
NTC
Smarcb1
Smarcd2
Smarcd1
Brd9
Kmt2d
Kmt2a
Wdr82
Hdac3
Setdb1
Fold change BFP
Days post-transduction
g
NTC
Smarcb1 (cBAF)
Smarcd2 (cBAF)
NTC
Kmt2a (MLL)
10.9%
25.4%
24.8%
40.6%
61.7%
PE-CD55
APC/Cy7-Gr-1
PE/Cy7-CD11b
h
Menin (Men1) Inhibitor
0 nM
10 nM
30 nM
100 nM
44%
60%
81%
95%
APC/Cy7-Gr-1
PE/Cy7-CD11b

Fig. 6

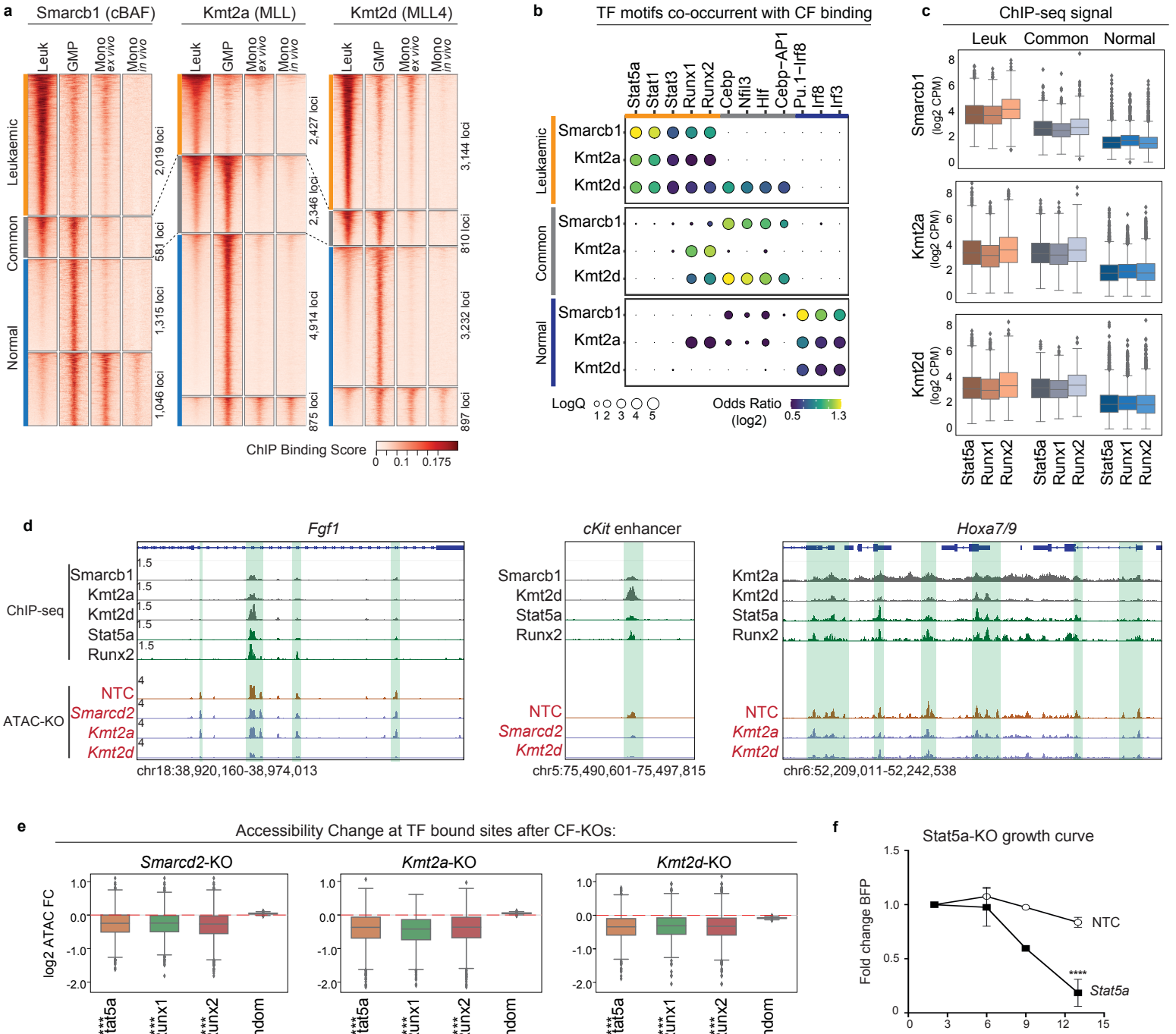
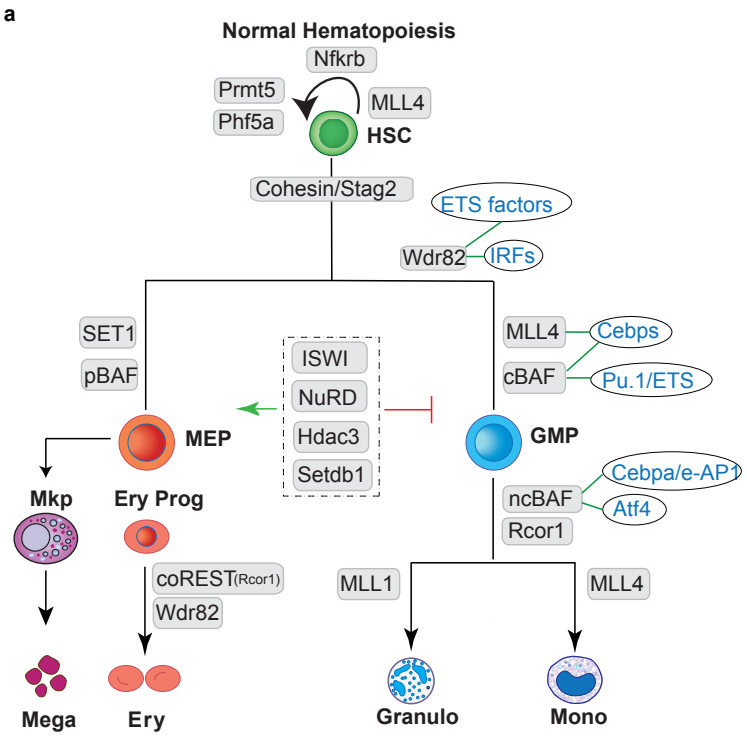
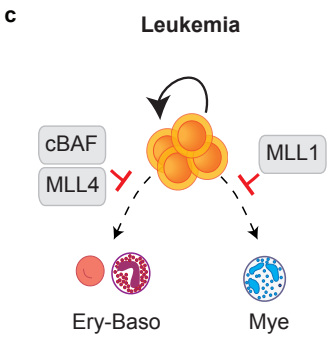

a
Smarcb1 (cBAF)
Kmt2a (MLL)
Kmt2d (MLL4)
Leuk
GMP
Mono ex vivo
Mono in vivo
Leukaemic
Common
Normal
2,019 loci
581 loci
1,315 loci
1,046 loci
2,427 loci
2,346 loci
4,914 loci
875 loci
3,144 loci
810 loci
3,232 loci
897 loci
ChIP Binding Score
0
0.1
0.175
b
TF motifs co-occurrent with CF binding
Stat5a
Stat1
Stat3
Runx1
Runx2
Cebp
Nfil3
Hlf
Cebp−AP1
Pu.1−Irf8
Irf8
Irf3
Smarcb1
Kmt2a
Kmt2d
LogQ
1 2 3 4 5
Odds Ratio (log2)
0.5
1.3
c
ChIP-seq signal
Leuk
Common
Normal
Smarcb1 (log2 CPM)
Kmt2a (log2 CPM)
Kmt2d (log2 CPM)
Stat5a
Runx1
Runx2
d
Fgf1
cKit enhancer
Hoxa7/9
ChIP-seq
ATAC-KO
Smarcb1
Kmt2a
Kmt2d
Stat5a
Runx2
NTC
Smarcd2
Kmt2a
Kmt2d
chr18:38,920,160-38,974,013
chr5:75,490,601-75,497,815
chr6:52,209,011-52,242,538
e
Accessibility Change at TF bound sites after CF-KOs:
Smarcd2-KO
Kmt2a-KO
Kmt2d-KO
log2 ATAC FC
Stat5a
Runx1
Runx2
Random
f
Stat5a-KO growth curve
Fold change BFP
NTC
Stat5a
****
Days post-transduction

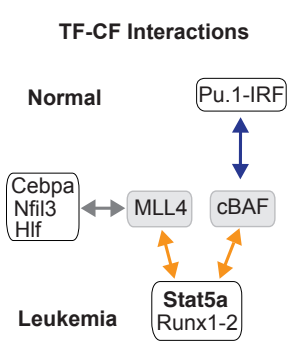
Fig. 7



b

Complex	Required for	Expanded after CF-KO	Interacting TF
cBAF	Mye branching	Ery	Cebpa/e, Pu.1
ncBAF	Mye maturation, B-cell	Prog	Cebpa/e, AP1 (Mye), Ebf1 (Bcell)
pBAF	Ery	Mye & B-cells	
SET1	Ery	Mega, Mye & Prog	
Wdr82	Mye	Mega	Cebpa/e, Pu.1
MLL1	Gran	B-cells	Cebp, Runx1
MLL2			
MLL3			
MLL4	Prog, Mono	Gran	Cebp





Contents

Supplementary Figures 1-7.

Supplementary Table 1. List of Chromatin Factors studied in this article.

Supplementary Table 2: sgRNA sequences.

Supplementary Table 3: Lineage scores in bulk CRISPR screens

Supplementary Table 4: *In Vivo* Perturb-seq Enrichment scores

Supplementary Table 5: *In Vivo* Perturb-seq Differential Gene Expression analysis

Supplementary Table 6: Leukemic Perturb-seq Enrichment scores

Supplementary Table 7: Perturb-seq Differential Gene Expression analysis in Leukemia (comparisons of CF-KOs to NTCs).

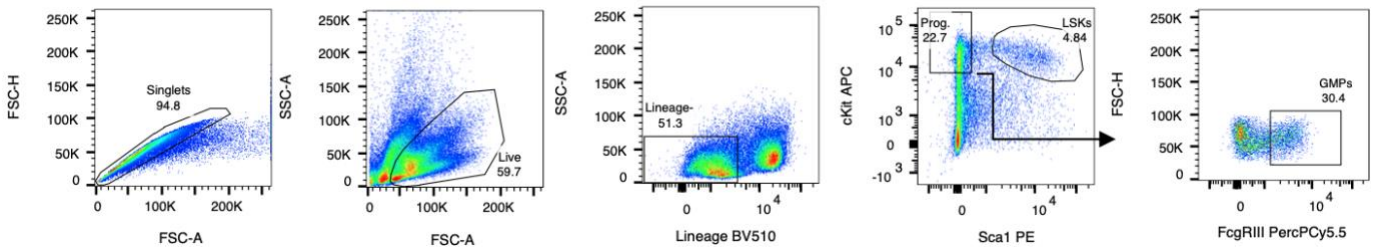
Supplementary Table 8: Primer list

Supplementary Table 9: FACS and CITE-seq Antibodies

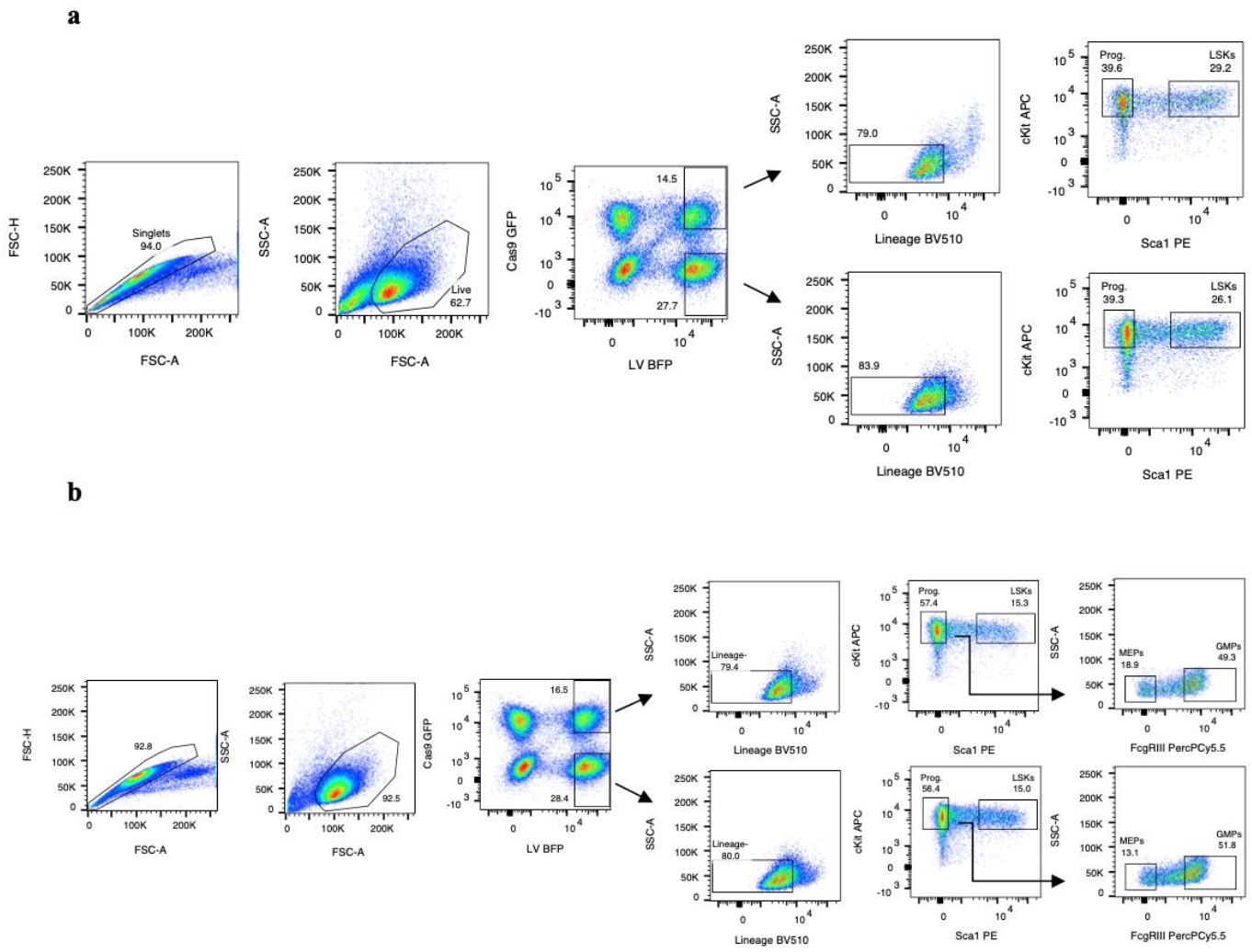
Supplementary Table 10: ChIP-seq Antibodies

Supplementary Table 11: ChIP-seq sequencing statistics

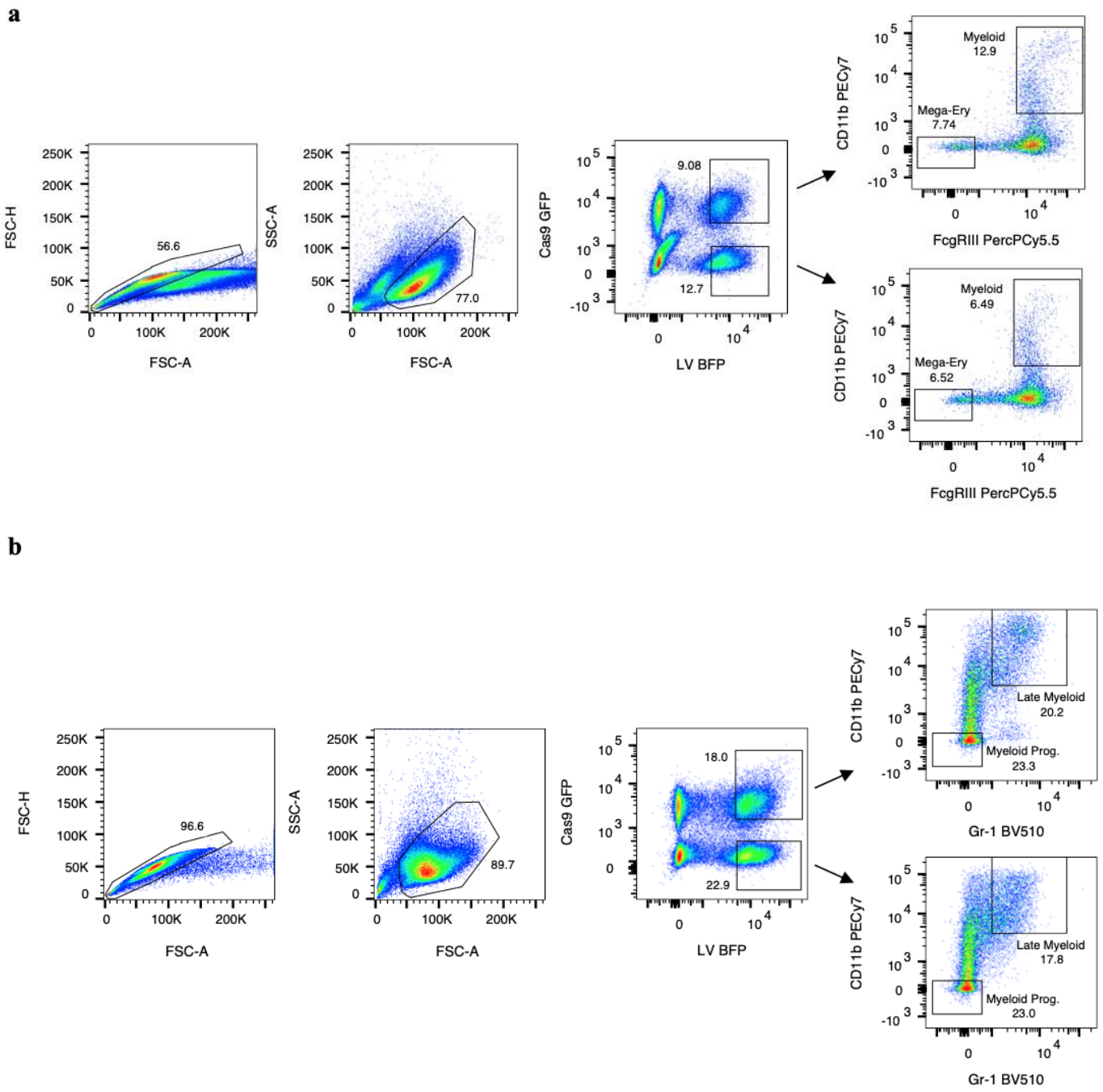
Supplementary Data 1-4: ChIP-seq peak counts (normalized) for Kmt2d, Kmt2a, Smarcb1 and Brd9.



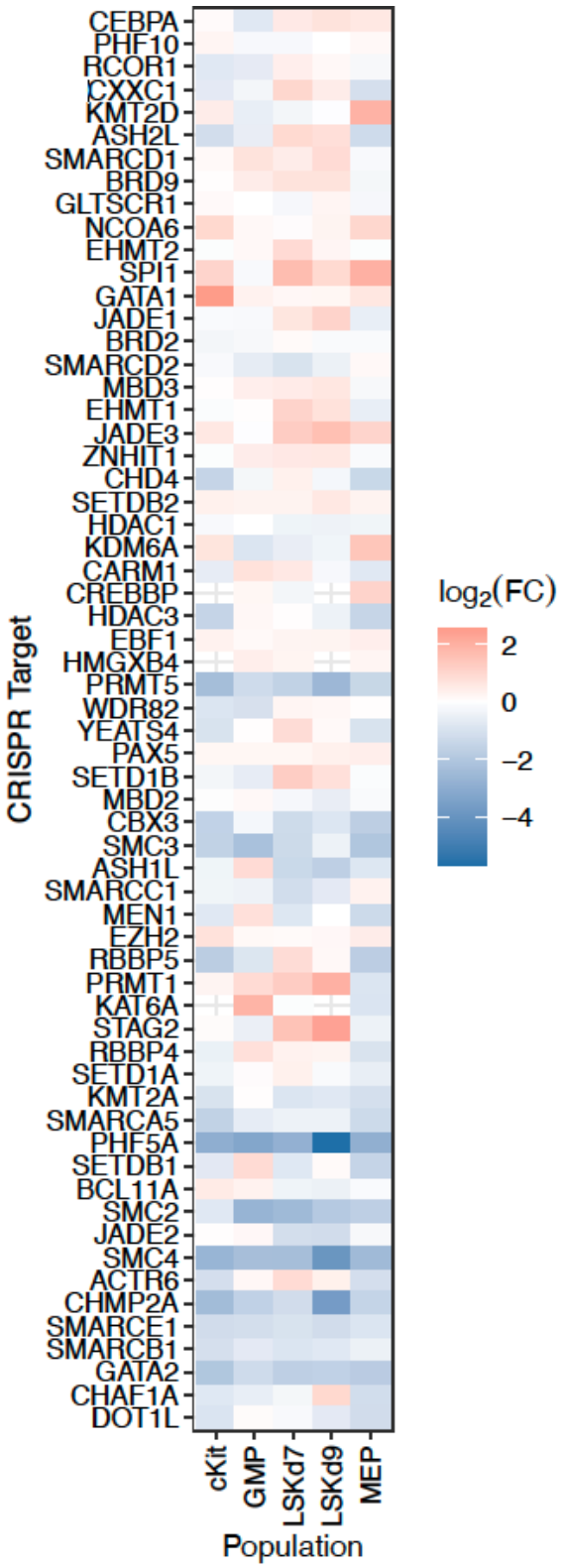
Supplementary Figure 1 | Gating strategy to isolate Haematopoietic progenitors (LSK) and Myeloid Progenitors (GMP)



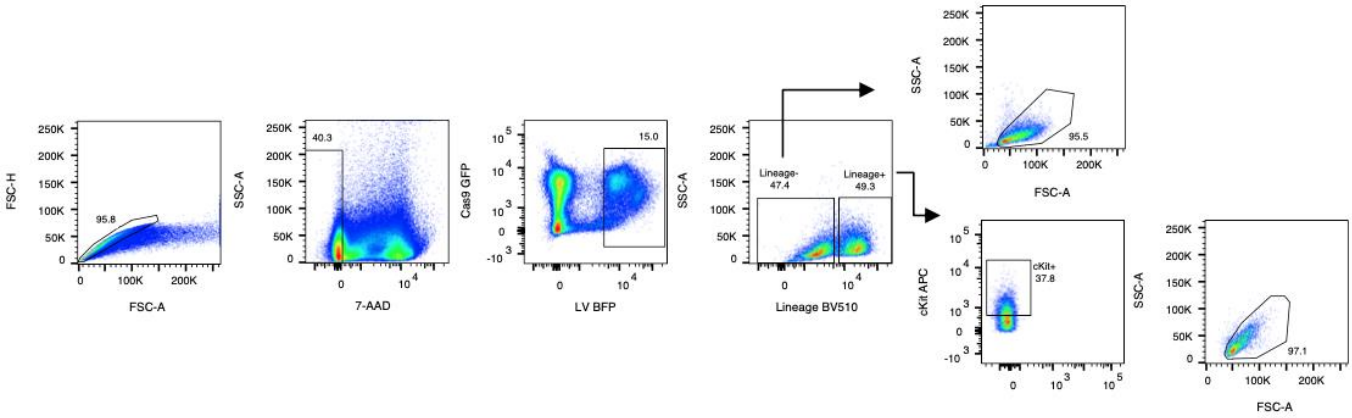
Supplementary Figure 2 | FACS Readouts for Bulk CRISPR screens. (a) Gating strategy for Self-renewal (LSK) vs Differentiation (Prog) Readout**. (b)** Gating strategy for Myeloid (GMP) vs Mega-erythroid (MEP)



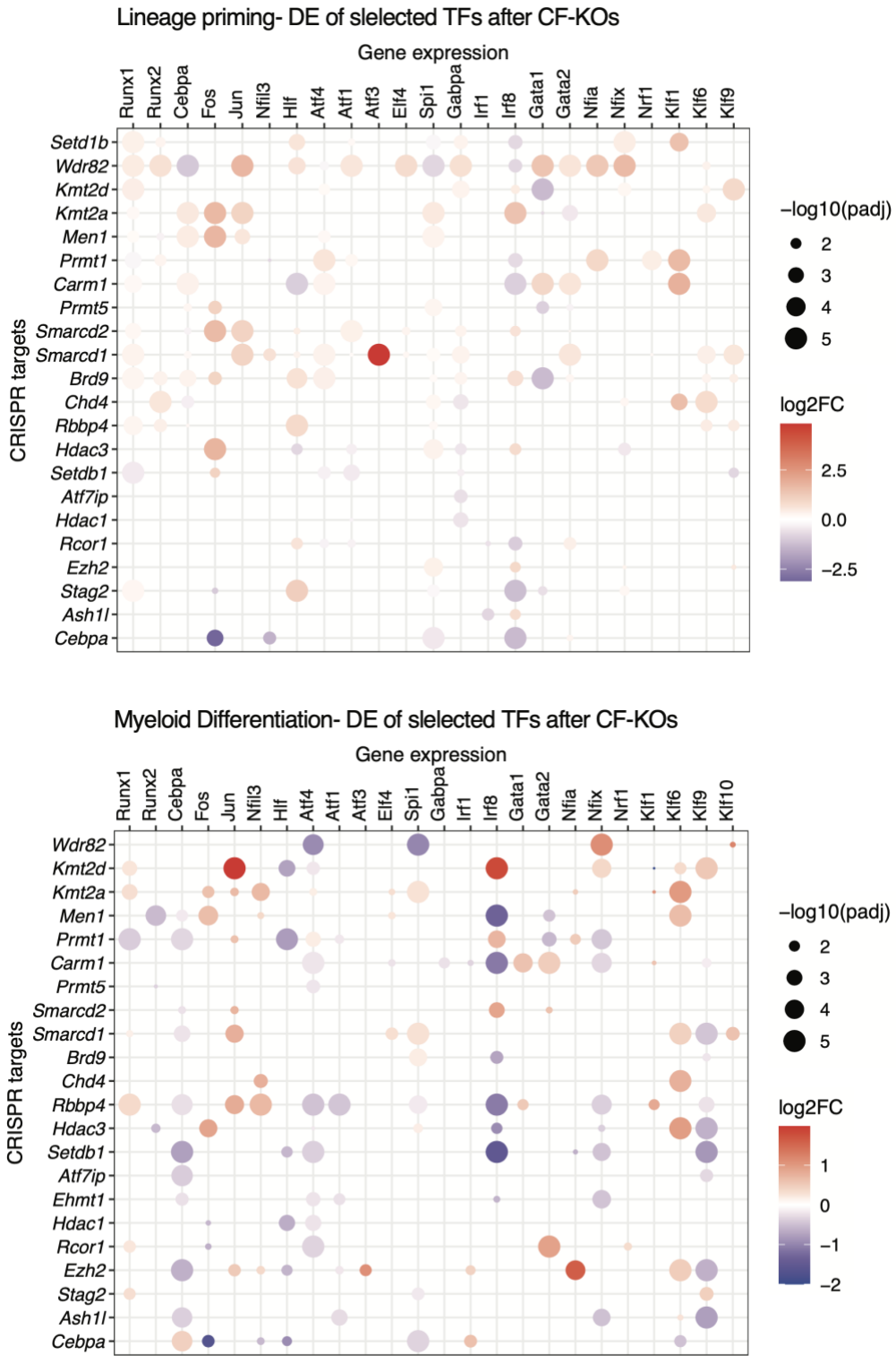
Supplementary Figure 3 | FACS Readouts for Bulk CRISPR screens. (**a**) Gating strategy for Myeloid Differentiation. (**b**) Gating strategy for Terminal Myeloid Maturation.



Supplementary Figure 4 | *Ex vivo* viability scores for CFs assessed in in vivo Perturb-seq The scores are calculated by comparing sgRNA abundances between Cas9 and Non-Cas9 fractions.

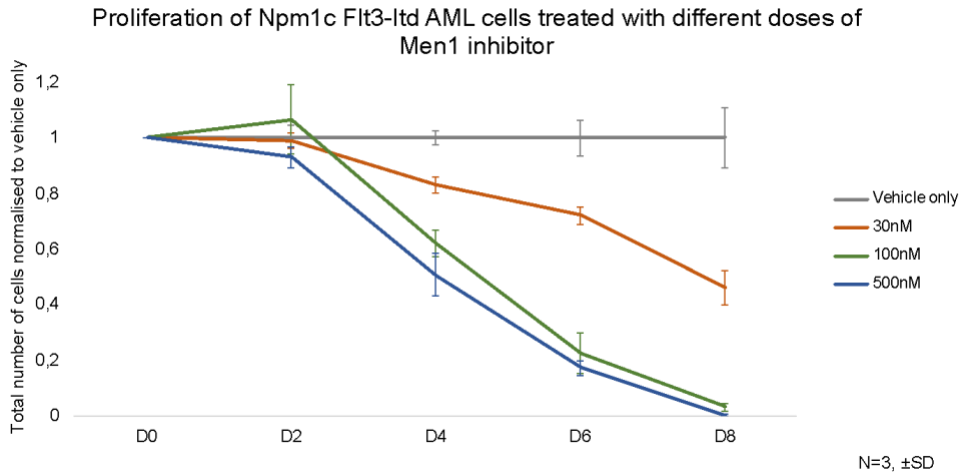


Supplementary Figure 5 | In vivo Perturb-seq gating strategy. We profiled cells derived from transplanted (GFP+), transduced (BFP+) LSK haematopoietic progenitors (gating strategy in Supplementary Figure 1) corresponding to Lineage- and Lineage+/ ckit+ gates.



Supplementary Figure 6 |In vivo Perturb-seq. Differential expression analysis of the Transcription Factors (examined in Figure 3) after CF-KOs under ex vivo lineage priming and

myeloid differentiation conditions. P-values were calculated using negative binomial mixed models from the nebula R package.



Supplementary Figure 7 | Effect of Menin inhibitor on the proliferation of *Npm1c/Flt3-ITD* leukemia cells. Cells were treated with increasing doses of the Menin Inhibitor SNDX-5613 (n=4). Error bars represent the Standard Deviation (SD).