Leukemia aggressiveness is driven by chromatin remodeling and expression changes of core regulators

Gracia Bonilla[1,2], Alexander Morris[1], Sharmistha Kundu[1,2], Anthony Ducasse[1], Nathan E. Jeffries[3,4,5], Kashish Chetal[1], Emma E. Yvanovich[3,4,5], Rana Barghout[3,4,5], David Scadden[3,4,5], Michael K. Mansour[4,6,7], Robert E. Kingston[1,2], David B. Sykes[3,4,5], Francois E. Mercier[8], Ruslan I. Sadreyev[1,9*]

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston MA, USA; [2]Department of Genetics, Harvard Medical School, Boston MA, USA; [3]Center for Regenerative Medicine, Massachusetts General Hospital, Boston MA, USA; [4]Harvard Stem Cell Institute, Cambridge MA, USA; [5]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge MA, USA; [6]Division of Infectious Diseases, Massachusetts General Hospital, Boston MA, USA; [7]Department of Medicine, Harvard Medical School, Boston MA, USA; [8]Department of Medicine, McGill University, Montreal, Quebec, Canada; [9]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston MA, USA

* Correspondence should be addressed to:
Ruslan I. Sadreyev
Massachusetts General Hospital
Department of Molecular Biology
Richard B. Simches Research Center
185 Cambridge Street CPZN-7250
Boston MA 02114
sadreyev@molbio.mgh.harvard.edu

ABSTRACT

Molecular mechanisms driving clonal aggressiveness in leukemia are not fully understood. We tracked and analyzed two mouse MLL-rearranged leukemic clones independently evolving towards higher aggressiveness. More aggressive subclones lost their growth differential ex vivo but restored it upon secondary transplantation, suggesting molecular memory of aggressiveness. Development of aggressiveness was associated with clone-specific gradual modulation of chromatin states and expression levels across the genome, with a surprising preferential trend of reversing the earlier changes between normal and leukemic progenitors. To focus on the core aggressiveness program, we identified genes with consistent changes of expression and chromatin marks that were maintained in vivo and ex vivo in both clones. Overexpressing selected core genes (Smad1 as aggressiveness driver, Irx5 and Plag1 as suppressors) affected leukemic progenitor growth in the predicted way and had convergent downstream effects on central transcription factors and repressive epigenetic modifiers, suggesting a broader regulatory network of leukemic aggressiveness.

INTRODUCTION

Therapeutic challenges in treating acute myeloid leukemia (AML) often arise from disease heterogeneity and clonal evolution during disease development [1,2] and post-treatment relapse [3,4]. Clonal evolution in leukemia is driven by both genetic and epigenetic changes in hematopoietic stem cells and leukemic progenitors [1,5]. Epigenetic dysregulation is a major component of AML origin and development. Genome-wide dynamics of DNA methylation [6] linked to methylases [7] and demethylase proteins [8] are associated with AML progression and outcome. COMPASS and SWI-SNF chromatin modifiers[9] that include MLL and other protein components [10,11], COMPASS-associated Menin-LEDGF complex [12], and COMPASS antagonists, Polycomb repressive complexes [13–15] are involved both in the origin and further development of AML. Another COMPASS antagonist, histone demethylase Kdm5b is a strong suppressor of AML aggressiveness[15]. Despite clinical and functional evidence of a central role that DNA and chromatin modifiers play in the evolution and maintenance of leukemia, understanding of epigenetic evolution among individual AML clones has remained elusive, in large part due to the heterogeneity of both molecular lesions and phenotypes across individual subclones [1,2,5,16,17].

Addressing these limitations, we analyzed a tractable clonal model to understand the dynamics of chromatin regulation and gene expression during the development of aggressiveness and to identify candidate drivers and suppressors of this process. We used a well-established mouse model of AML driven by the MLL-AF9 fusion [18], a prevalent MLL rearrangement observed in AML patients [19]. We performed genome-wide profiling of chromatin modifications and gene expression in independent MLL-AF9 leukemic clones and characterized epigenomic and transcriptional changes as the cells became more aggressive. Compared to the changes between normal progenitors and the early stage of leukemic development, these aggressiveness-associated

changes affected fewer genes and enhancers and were more clone-specific, consistent with heterogeneity of clonal evolution. These changes often involved gradual shifts of expression levels and chromatin states rather than on/off switching between full silencing and activity. Surprisingly, most of these shifts reversed the prior promoter and enhancer changes observed at the early leukemic stage.

To identify candidate regulators of aggressiveness and its memorization throughout long-term ex vivo culture, we focused on the core set of genes whose in vivo expression and ex vivo chromatin changes were consistent between individual clones. Three genes were further validated, one as a causal driver (Smad1) and two as suppressors of clonal growth (Irx5, Plag1). Mechanistically, we found an unexpected convergence of their downstream transcriptional effects on a group of transcription and epigenetic factors that may represent a broader regulatory network of clonal aggressiveness. Our results show how a dynamic epigenome can enable clonal selection through a shared program conferring growth advantage to aggressive leukemic clones.

RESULTS

Chromatin and gene expression profiling at two stages of leukemia aggressiveness

We employed a cellular model of developing leukemic aggressiveness using the mouse AML model driven by the human MLL-AF9 (KMT2A-MLLT3) fusion protein [18], a recurrent MLL rearrangement [19]. In this model, the leukemic initiation is mainly attributed to MLL fusion driving the dysregulated reprogramming of committed granulocyte macrophage progenitors (GMPs) into leukemic stem cells (LSCs) [18,20]. To track individual leukemic clones, GMPs isolated from CD45.1 congenic mouse were transduced with four individual fluorescent proteins. An equal mix of labeled clones was transplanted into irradiated host mice, followed by longitudinal tracking in bone marrow and blood

for 120 days (Figure 1A). Following engraftment, these clones evolved in bone marrow, resulting in cell-autonomous differences in growth rate. The more aggressive clones 'won' the competition and dominated the bone marrow and peripheral blood.

Winning clones dramatically increased growth rate in vivo as they transitioned from early to aggressive stage (Figures 1A, S1A,D). This difference between the early and aggressive subclones was significantly diminished ex vivo (Figures S1B,E), suggesting the essential role of the bone marrow microenvironment in maintaining aggressive growth. This difference, however, was restored upon secondary re-transplantation (Figures 1A, S1C,F), suggesting an intrinsic cellular memory of aggressive status that was maintained over multiple cell divisions ex vivo. To understand the epigenetic and transcriptional regulation of clonal aggressiveness, we focused on two independent leukemias that evolved into aggressive subclones (Figure 1A). We performed RNA-seq profiling of gene expression in LSC-enriched cell populations at the early and aggressive stages of the same clone both in vivo and ex vivo, as well as ChIP-seq profiling of H3K4me1, H3K4me3, H3K27ac, and H3K27me3 chromatin marks ex vivo. LSC-enriched progenitors from healthy mice were profiled as a normal comparator (Figure 1A).

Early changes of promoter chromatin marks and gene expression are consistent between leukemic clones

As an example of a major driver of MLL-rearranged (MLLr) leukemias[21], *Meis1* gene was strongly activated in all leukemic subclones when compared to normal progenitors, as evidenced by transcriptional upregulation concomitant with the loss of repressive chromatin mark H3K27me3 and gain of active marks H3K27ac and H3K4me3 at the promoter (Figure 1B, top of Figure 1C). However, this activation of *Meis1* and other major leukemic drivers did not change in aggressive

subclones (Figures 1C, S2, S3), suggesting that their further upregulation is not required for clonal aggressiveness.

Genome-wide, most changes of promoter chromatin marks and gene expression between normal progenitors and the early leukemic stage were also consistent between clones and maintained throughout leukemia progression (Figures 1D, S4A). When compared to normal progenitors, the two early-stage leukemic clones shared the vast majority (~85%) of differentially expressed genes (DEGs, Supplementary Data 1) and large fractions of promoters with H3K27ac, H3K4me3, and H3K27me3 changes (Figure 1E). According to EnrichR [22] pathway analysis, DEGs shared between clones were enriched in IL2/STAT5 and EMT pathways and targets of central transcription factors (Myc, Hif1a, Sox2, Tcf3, Mecom, Af4, Runx1, Cebpa, Irf8, p53, Smad3) and epigenetic factors (p300, Brd4, Polycomb proteins, Dnmt1) implicated in hematopoiesis and leukemogenesis (Figure 1F).

Compared to normal progenitors, promoter H3K27ac was preferentially increased and H3K7me3 was preferentially decreased, whereas H3K4me3 changes were more balanced (Figure 1G). Global gene expression was modestly skewed towards upregulation (Figure 1G). The combination of chromatin marks was robustly predictive of expression changes, with active marks H3K27ac and H3K4me3 being the strongest predictors (Figure S4B), despite a substantial part of promoter chromatin remodeling did not manifest as expression changes, for example at transcriptionally silent promoters that underwent H3K27me3 decrease without increase of active marks (Figure 1D, middle part).

When compared to normal progenitors, both stages of aggressiveness had similar differences in promoter chromatin marks (Figure 1H, points near diagonal); the minority of promoters that deviated between early and aggressive subclones (Figure 1H, points away from diagonal) are the focus of our analyses.

Promoter H3K4me3 and H3K27me3 levels define three distinct chromatin states

To understand global patterns of promoter remodeling, we analyzed genome-wide relationships between promoter chromatin marks. We focused on H3K4me3 and H3K27me3, given their roles in promoter regulation [23] and the importance of the corresponding COMPASS/MLL [10,11] and Polycomb protein complexes [13–15] in leukemia. The scatterplot in Figure 1I represents the combination of H3K27me3 (x axis) and H3K4me3 levels (y axis) at each promoter in normal progenitor cells as a point colored by the level of gene expression. The two marks followed a distinct pattern previously observed elsewhere [24]. They were anticorrelated among active promoters (the upper branch), with the fully active chromatin state [25] (highest H3K4me3 and no H3K27me3) at the tip of this branch (Figure 1I, green circle). They were positively correlated and co-depleted among silent promoters (the lower branch), down to the lowest tip with a full H3K4me3 and H3K27me3 depletion (Figure 1I, grey circle), which typically corresponds to constitutive heterochromatin and DNA methylation [25]. The “nexus” between two branches included promoters with the highest H3K27me3 and moderate H3K4me3 (Figure 1I, red circle), corresponding to Polycomb-repressed bivalent chromatin state of developmental factors with the potential for activation [25,26]. Many major leukemic drivers were bivalent in normal progenitors but activated along the active branch in leukemia (Figures 1B,C, S2, S3). This pattern of three states connected by two branches was similar in leukemic clones (Figures S5A,B).

Genome-wide trends of promoter chromatin remodeling in early leukemic progenitors

This pattern was used to analyze the chromatin state shifts between normal and early leukemic progenitors, depicted as arrows with the approximate numbers of promoters following each direction (Figure 1I). Different directions were associated with targets of different transcription factors and chromatin modifiers, suggesting specific roles of these factors in promoter regulation (Figure S5C). The strongest trend involved preferential movements from active and bivalent states (Figure 1I, green and red circles) towards the silent state [25] (grey circle), consistent with a major leukemic role of de novo DNA methylation [7]. More detailed effects on individual promoters and gene expression are shown in Figure S6. The second trend involved movements from bivalent state (Figure 1I, red circle) towards active state (green circle) via H3K4me3 increase and H3K27me3 decrease accompanied by transcriptional upregulation (Figure S6H). Notably, it included major leukemic drivers (Hoxa9, Meis1, Six1, Six4, Eya1, etc., Figures 1B,C, S2, S3). The third, unexpected trend involved H3K4me3 increase at bivalent promoters while maintaining high H3K27me3 (Figure 1I, upward arrow on the right) and transcriptional silencing (Figure S6F), suggesting H3K4 methylation uncoupled from Polycomb repression. Taken together, we identified recurrent shifts between promoter chromatin states driven by specific changes in the global activity of repressive and activating epigenetic factors.

Enhancer chromatin remodeling in early leukemic progenitors was consistent between clones

We identified ~40,000 enhancer elements based on the density of H3K27ac or H3K4me1 and the absence of strong H3K4me3 enrichment, and analyzed the shifts of enhancer chromatin states between normal and early leukemic progenitors. As an example, two enhancers in the proximity of the *Des* gene showed consistent increases in H3K27ac and H3K4me1 in both clones (Figure 1J).

These changes are quantified in a scatterplot (Figure 1K, top) where individual points correspond to the combination of two marks in normal (grey), early leukemic (blue), and aggressive cells (red), with proximal gene expression shown as a barplot. *Pitx2* (Figure 1K, middle) and *Lmo1* proximal enhancers (Figure 1K, bottom) also underwent consistent activation and deactivation of chromatin state (Figure S7). Across the genome, 17,338 enhancers changed H3K27ac or H3K4me1 >2-fold in at least one clone (Figures 1L, S8). These changes correlated between the two marks and were consistent between two clones (Figure 1L), which shared the majority of enhancers with >2-fold change of H3K27ac and a large fraction of enhancers with >2-fold change of H3K4me1 (Figure 1M). The magnitude of these changes was largely maintained at the aggressiveness stage (Figure 1N); the minority of enhancers that showed stage-specific differences are the focus of our analyses below.

H3K27ac and H3K4me1 levels across enhancers were largely correlated (Figure 1O). At the early leukemic stage, global changes of H3K27ac were balanced between increases and decreases, and this balance was a result of two opposite trends: concomitant H3K27ac and H3K4me1 changes skewed towards increase (Figures 1O and S9, diagonal arrows), and H3K27ac changes alone (decoupled from H3K4me1) which were skewed towards decrease (Figures 1O and S9, vertical arrows). Interestingly, H3K4me1 changes were skewed towards increase regardless of whether they were coupled to H3K27ac changes (diagonal arrows) or not (horizontal arrows), suggesting a global increase of enhancer H3K4 methylation independent from H3K27 acetylation.

Clonal aggressiveness is accompanied by promoter chromatin and gene expression changes

We focused on chromatin and expression changes associated with clonal aggressiveness. The *Fhdc* promoter, for example, was in a repressed bivalent state in normal progenitors, with no H3K27ac, high H3K27me3, and modest H3K4me3. At the early leukemic stage in both clones, it switched to a transcriptionally active state with high H3K4me3 and H3K27ac and no H3K27me3 (Figure 2A). However, at the aggressive stage this change was partially reversed by shifting back the levels of chromatin marks and expression. The *Fyb* promoter was active in normal progenitors, with high H3K27ac and H3K4me3 and no H3K27me3. At the early leukemic stage in both clones, it shifted to a repressed bivalent state with low or no expression. At the aggressive stage, this strong repression was partially reversed via decreasing H3K27me3 and increasing H3K4me3 and H3K27ac (Figure 2B). These changes are quantified in the scatterplots of H3K27me3 vs H3K4me3 (Figure 2C), with additional examples in Figures S10, S11. Collectively, these changes suggest that developing clonal aggressiveness largely involved modulation of promoter activity on a continuous scale rather than on/off switching between extreme chromatin states.

MLL-AF9 direct targets are stable during aggressiveness development

To address whether clonal aggressiveness could be ascribed to increased activity of direct MLL-AF9 targets, we analyzed target promoters with the strongest MLL-AF9 peaks (Homer score ≥ 13) based on published ChIP-seq data in mouse MLL-AF9 HSPCs [27]. Most targets were active in both normal and leukemic progenitors. At the early leukemic stage 35-40% of MLL-AF9 targets (depending on the clone) were further upregulated, while only 5-6% were downregulated, summarized as heatmap (Figure S12) and trends of combined H3K27me3 and H3K4me3 changes (Figure S13A,B). However, there was no preferential activity shift at the aggressive stage, where only a small minority of MLL-AF9 targets were differentially expressed (12-15% depending on the

clone, Figure S12) and were mostly downregulated. Only 3.5-12% of these targets had H3K4me3, H3K27ac, or H3K27me3 changes at the aggressive stage (Figure S13C).

Remarkably, virtually all MLL-AF9 targets that were repressed by Polycomb and had high H3K27me3 in normal progenitors underwent a dramatic H3K27me3 depletion and H3K4me3 increase at the early leukemic stage (Figure S13A,B). These targets included major leukemogenic drivers Hoxa9, Meis1, Six1, Six4, etc (Figures 1B, S2). However, their expression or chromatin state did not further change at the aggressive stage (Figures 1B,C, S2, S3), with the possible exception of Eya1 and Jmjd1c that changed expression in one of the clones (Figure S3E,F). This absence of clone-consistent changes suggested that direct regulation by MLL-AF9 or by major leukemogenic drivers was unlikely to be a primary driver of clonal aggressiveness.

Clone specific changes during development of aggressiveness preferentially reversed early leukemic changes

We then analyzed global patterns of aggressiveness-associated changes in gene expression and promoter chromatin marks (Figures 2D, S14). These changes affected many promoters across the genome, albeit fewer than at the early leukemic stage (Figure 1D). This smaller but substantial magnitude of aggressiveness-associated changes was also evident from principal component analysis (PCA) of gene expression (Figure 2E). The difference between normal progenitors (gray points) and all leukemic subclones regardless of aggressiveness (blue and red) corresponded to the first principal component (x axis) that accounted for the majority (64%) of overall expression variation. The difference between two stages of aggressiveness (blue vs red) corresponded to principal component 2 (y axis) and accounted for a smaller but substantial fraction of overall variation (20%).

Compared to the highly consistent changes at the early leukemic stage (Figure 1D), changes at the more aggressive stage (Figure 2D) were more clone-specific, suggesting distinct paths to the acquisition of aggressiveness. Clones 1 and 2 shared 54% and 67% (n=1388) of aggressiveness-associated DEGs, respectively (Figure 2F, Supplementary Data 1), compared to ~85% at the early stage (Figure 1E). This clonal specificity was even more pronounced among chromatin marks (Figure 2F). Consistent with a strong influence of the bone marrow environment, there were substantial differences between aggressiveness-associated changes observed in vivo and ex vivo in the same clone, which shared only 20% of upregulated and 13% of downregulated genes (Figure S15). We postulated that these shared genes included potential mediators of cellular memory that was maintained despite the diminished growth difference between early and aggressive subclones ex vivo and manifested upon secondary re-transplantation (Figure S1).

Genes with consistent expression changes at the aggressive stage in both clones (Figure 2G) were enriched in central regulatory pathways (EMT, interferon and inflammatory responses, IL6/JAK/STAT3, IL2/STA5, KRAS signaling), targets of transcription factors (Satb1, Cebpa, Myc, E2f8, Mecom, Gata1/2, Stat1, Jun, Irf4/8, Smad3/4, p53, Sox2, Oct4/Pou5f1, etc), chromatin modifying proteins (Polycomb components, Dot1l, p300, Arid3b, Setdb1), and DNA methylation and demethylation machinery (DNMT1, Tet1). These factors included general cell cycle regulators (e.g. E2f8), pluripotency factors (Sox2, Oct4, etc), and regulators of leukemic growth (Cebpa, Myc, Dot1l, etc). Among chromatin modifiers, there was a notable enrichment across Polycomb group proteins (Rnf2, Kdm2b, Suz12, Eed, Ezh2, Jarid2, MTF2), suggesting a major role of Polycomb in transcriptional regulation of aggressiveness.

The majority of DEGs associated with aggressiveness (~70-75% depending on the clone) were already differentially expressed at the prior early leukemic stage (Figure 2H). Promoters affected by chromatin remodeling also had a strong overlap between two stages (~50%, Figure 2H). Surprisingly, most of these changes at the aggressive stage were not a further continuation of the early leukemic changes but their reversal (Figure 2I, large arrows depicting the early changes compared to smaller arrows depicting changes at the aggressive stage).

Genome-wide trends of promoter chromatin remodeling in aggressive progenitors

Trends of concomitant H3K4me3 and H3K27me3 changes at the aggressive stage (Figures 2J, S16) were different from the early leukemic changes (Figure 1I). Despite a stronger clone specificity, three preferential directions were consistent between clones (Figure S16A,B, yellow arrows in Figure 2J). Promoters with high H3K27me3 preferentially increased H3K4me3 (Figures 2J, S16, vertical upward arrow on the right), similar to the early stage (Figure 1I). However, among promoters depleted in H3K27me3, the early skew towards H3K4me3 decrease (Figure 1I, vertical downward arrow on the left) was dramatically reversed to a preferential H3K4me3 increase (Figures 2J, S16, vertical arrow on the left). Finally, bivalent promoters (Figures 2J, S16, red circle) shifted preferentially towards active state (green circle) via H3K4me3 increase and H3K27me3 decrease (upward diagonal arrow), consistent with a strong enrichment of Polycomb targets among aggressiveness associated DEGs (Figure 2F). Although similar to the early leukemic stage (Figure 1I), this activation trend involved virtually no MLL-AF9 targets, which were stably active. Collectively, all these trends involved H3K4me3 increase in various chromatin contexts, consistent with the recent report of H3K4me3 demethylase Kdm5b as a potent suppressor of leukemic aggressiveness [15].

Chromatin changes were correlated with gene expression; the combination of promoter H3K4me3, H3K4me1, H3K27ac, and H3K27me3 differences ex vivo was robustly predictive of expression differences (R=0.5, Figure S17A). Active marks H3K27ac and H3K4me3 were most predictive (Figure S17A). Despite the difference between expression changes observed in vivo and ex vivo, a subset of DEGs was maintained throughout ex vivo culturing (Figure S15). Accordingly, despite ex vivo chromatin mark changes were less predictive of expression in vivo, H3K27ac and H3K4me3 were still the most predictive (Figure S17B).

Aggressive subclones preferentially activated chromatin state at enhancers and reversed early leukemic changes

Aggressiveness was associated with widespread modulation of enhancer chromatin marks. As an example, the Net1 enhancer increased both H3K27ac and H3K4me1 at the aggressive stage (Figure 3A), whereas the enhancer upstream of gene Zfp516 decreased both marks (Figure 3B, see additional examples in Figure S18). These changes are quantified in H3K4me1 vs H3K27ac scatterplots in Figures 3C and S19. Like promoters (Figures 2A-C, S10, S11), chromatin mark levels at these enhancers were gradually modulated, as opposed to switching between extreme chromatin states. Globally, aggressiveness-associated changes of enhancer chromatin state showed a strong skew towards activation (Figures 3D, S20). The consistency of H3K27ac and H3K4me1 changes between two clones was significant (20-22% for H3K27ac and 11-15% for H3K4me1 depending on the clone, Figure 3E) but lower than the clonal consistency of chromatin mark changes at promoters (Figure 2F).

Like promoters (Figure 2H), a large proportion of enhancers affected by the development of aggressiveness already underwent changes at the early leukemic stage. This proportion was

especially large for H3K27ac (Figure 3F). Surprisingly, most of these enhancers reversed their initial H3K27ac changes (Figure 3G, large up and down arrows depicting changes at the early stage compared to smaller arrows at the aggressive stage).

Taken together, these aggressiveness-associated changes of gene expression and chromatin marks at promoters (Figures 2A,B,H,I) and enhancers (Figures 3A,B,F,G) suggest a few important observations. First, most of these changes were gradual quantitative shifts rather than binary switches between silent and active states. These gradual shifts were often consistent between clones to a considerable degree of quantitative detail. Second, most of these shifts were preceded by the prior changes of the same promoter or enhancer at early leukemic stage. When compared to normal progenitors, these genes most often maintained the early direction of shift, either activation or repression, but altered the magnitude of this change (Figure 1H,N). This suggests that increasing clonal aggressiveness involved quantitative adjustment of initial dysregulation that occurred during leukemogenesis. Unexpectedly, this quantitative adjustment often dampened the initial perturbation. As one possible explanation of clonal fitness advantages conferred by this dampening, some of the extremely strong downstream effects of initial leukemogenic lesion (MLL rearrangement) may require fine tuning or even toning down for further increase of cell growth and survival. This fine tuning might be mediated via additional mutagenesis, changes in expression of key transcription factors, chromatin remodeling by epigenetic factors, or other mechanisms.

Genome-wide trends of enhancer chromatin remodeling in aggressive progenitors

In leukemic cells (Figures 3H and S21), like in normal progenitors (Figure 1O), H3K27ac and H3K4me1 levels were correlated across all enhancers. The trends of concomitant H3K27ac and H3K4me1 changes were consistent between two independent aggressive clones (Figure S21).

H3K27ac changes were skewed towards increase (Figures 3H, S21) in contrast with more balanced changes at the early stage (Figure 1O), mostly due to the reversal of preferential H3K27ac changes uncoupled from H3K4me1 change (vertical yellow arrow). These changes were skewed towards decrease at the early leukemic stage (Figure 1O, vertical arrows) but switched towards increase at the aggressive stage (Figures 3H and S21). This suggests that in addition to continued global increase of enhancer H3K4me1, aggressiveness was associated with increase of H3K27 acetylation often uncoupled from H3K4 methylation.

These changes of enhancer chromatin marks were associated with various transcription factor binding motifs according to MEME[28] (Figure 3I). Transcription factors associated with the changes of both H3K27ac and H3K4me1 included ETS family proteins and related PU.1 factor that have been implicated in hematopoiesis and leukemogenesis [29,30]; CEBPB/D/E/G involved in myeloid hematopoiesis and aggressive leukemia [31–33]; and ZNF384 with reported leukemic fusions and regulation of hematopoietic stem cell factors[34]. Motifs associated with only H3K27ac changes involved major regulatory factors STAT2, BCL11B, IKZF1, NR5A1/2, and KLF9, whereas H3K4me1 changes were associated with leukemic regulators CEBPA, MYB, RUNX1, pluripotency factor KLF4, AP-1 factor JUN, and other factors (Figure 3I). Taken together, enhancer chromatin remodeling at the aggressive stage was associated with motifs for ETS, CEBP, AP-1 factors and other differentiation and leukemic regulators.

For each enhancer, we calculated correlation between its H3K27ac level and the expression of the proximal gene across the five profiled cell types (normal progenitors and two stages of aggressiveness in two independent leukemic clones). As an example, Figure 3J quantifies this correlation for two enhancers depicted in Figures 3A,B. Among all enhancers with aggressiveness-

associated H3K27ac changes, there was a remarkable skew towards high correlation (histogram in Figure 3K), suggestive of transcriptional regulation. We focused on 406 enhancers with the highest correlation to gene expression (Pearson R>0.7, right of the dotted line in Figure 3K). EnrichR analysis of the proximal 307 genes (Figure 3L) revealed strong enrichment in the targets of major differentiation, hematopoiesis, and leukemogenesis regulators (Wt1, Esr1, Lmo2, Gf1, Myb, Tal1, Cebpd, Stat3, Meis1), components of super-elongation complex also known as MLL fusion partners (Af4/Mllt3 and Af9/Enl), pluripotency factors (Sox2, Nanog, Pou5f1, Prdm14), and chromatin modifiers (Brd4, p300, Smarcd1, Setdb1) often implicated in hematopoiesis and leukemia. The enrichment of Nfkb1 and Smad3/4 targets suggested the involvement of NFKB and TGFβ pathways. Although many enriched categories were similar to the ones among the whole population of aggressiveness-associated genes (Figure 2G), there were notable differences. Polycomb components were not as strongly enriched (Figure 3L), suggesting lesser Polycomb control; whereas targets of Kdm5b and Smad1 were enriched exclusively among enhancer-correlated promoters, suggesting that Kdm5b, H3K4 histone demethylase suppressing AML aggressiveness[15] and Smad1, a BMP pathway effector, are involved in enhancer-mediated regulation.

Core genes defined by consistent aggressiveness-associated expression and chromatin changes in vivo and ex vivo

Despite substantial variability by clone and condition, some aggressiveness-associated DEGs were consistent between clones (Figure 2F) and maintained their differential status both in vivo and ex vivo (Figure S15). To identify potential drivers or suppressors of aggressiveness that persisted ex vivo, we defined a core set of genes with consistent expression and promoter chromatin changes persisting between ex vivo and in vivo conditions in a clone-independent manner (Figure 4A). We

identified 17 such genes consistently upregulated and 11 genes consistently downregulated at the aggressive stage (Figures 4B, S22A).

While none of these genes were strong MLLr targets, many were known regulators of hematopoiesis, leukemia and other cancers. Among upregulated genes, Rxra is a hormone receptor mutated and therapeutically targeted in an aggressive form of AML, acute promyelocytic leukemia [35]. Smad1 is a key effector in BMP signaling pathway playing an important role in normal hematopoiesis [36,37] and leukemia [38,39]. Peg10 is known to enhance tumor growth by driving cell cycle progression through the inhibition of TGFβ pathway [40,41]. Consistent with this mechanism, another upregulated gene Matn2 is activated by TGFβ signaling inhibition [42].

Among downregulated genes, homeobox factor Irx5 plays major roles in early development [43,44], metabolism [45] and cancers [46,47], and is upregulated in AML patients and cell lines [48], but no mechanistic role in AML has been reported. Irx5 is a strong target of MLL-AF9 binding and transcriptional regulation [49] and was dramatically activated at the early leukemic stage (Figures S22A, 4C,D). However, in contrast to most MLL-AF9 targets (Figures S12, S13), Irx5 reversed this activation and was downregulated in aggressive cells (Figure S22A, 4B-D). Plag1 is an oncogene involved in the origin of leukemia [14,50] and was upregulated at the early leukemic stage (Figures S22A,B, 4D). However, similar to Irx5, Plag1 reversed this activation and was downregulated at the aggressive stage (Figures S22A,B, 4B,D). To our knowledge, Plag1 has not been previously implicated in the development of AML aggressiveness. Protein tyrosine phosphatase Ptpn3 is a potential tumor suppressor facilitating the activity of TGFβ pathway by stabilizing TGFβ receptor [51]. Ptpn3 downregulation was in line with the upregulation of Peg10 and Matn2 associated with TGFβ inhibition (Figure 4B). Finally, Igf1r is a major regulator of cell transformation and oncogenesis,

including AML [52], consistent with Igf1r upregulation at the early leukemic stage (Figure S22A). However, once MLLr leukemia is established, Igf1r is not anymore required for clonal propagation [52], in line with the observed Igf1r reversal towards downregulation at the aggressive stage (Figures S22A, 4B).

Gradual modulation of core genes involved reversal of early leukemic changes

Compared to normal progenitors, most core genes maintained their activation or repression at both leukemic stages, but its magnitude was stage-specific (Figure S22A). Of note, enhancers proximal to these genes often had similar gradual shifts of chromatin state. As an example, strong Polycomb repression of *Irx5* promoter in normal progenitors was completely removed at the early leukemic stage but partially reversed at the aggressive stage by increasing H3K27me3 to a moderate level while partially decreasing H3K4me3 and H3K27ac (Figure 4C, left). This partial reversal of early activation was paralleled by H3K27ac and H3K4me1 changes at a proximal enhancer (Figure 4C, right). *Plag1* had a similar reversal of early promoter and enhancer activation (Figure S22B). This reversal is quantified in Figure 4D as scatterplots of promoter H3K27me3 vs H3K4me3 (left) and enhancer H3K4me1 vs H3K27ac (middle), along with gene expression changes (barplots on the right). Conversely, promoters and proximal enhancers of *Smad1* (Figure 4E) and *Rxra* (Figure S22C) were active in normal progenitors, followed by strong repression at the early leukemic stage and its partial reversal at the aggressive stage. These changes are quantified in Figure 4F. As RXRA is a well-known leukemogenic driver in a different genetic background of PML-RARA fusion [35], its behavior in aggressive MLL-AF9 clones may suggest a more general role in driving leukemic aggressiveness, uncoupled from its leukemogenic role.

The fraction of the core genes that were already affected at the early stage (Figure 4G) was even larger than among the broader set of all aggressiveness-associated promoters (Figure 2H). Most of these promoters reversed their early leukemic changes of gene expression and chromatin marks (Figure 4H), contrary to the naïve expectation that these changes would be further amplified.

Functional validation of aggressiveness-associated gene candidates

Three core aggressiveness-associated genes: Irx5, Plag1, and Smad1 were selected for validation in leukemic progenitors. This selection was based on the degree and clonal consistency of chromatin mark and expression changes during clonal progression (Figures 4B,C-F). Since transcription factors Irx5 and Plag1 were consistently repressed at the aggressive stage (Figures 4C,D, S22B), we hypothesized that they were suppressors of aggressive leukemic growth. Since BMP pathway effector Smad1 was consistently activated (Figures 4E,F), we hypothesized that Smad1 is a driver of aggressive leukemic growth.

To validate phenotypic effects of Irx5, Plag1, and Smad1, we used lentiviral transduction to overexpress these genes in leukemic progenitor clones. Irx5 and Plag1 were overexpressed in aggressive leukemic progenitors to test their suppressive effect while Smad1 was overexpressed in leukemic progenitors at the early leukemic stage to test its activating effect. As a positive control, the known oncogenic driver, Kras G12V mutant [53], was overexpressed in leukemic progenitors at the early stage. Expression was confirmed at the RNA level (Figure 5A).

The size and morphology of leukemic cells overexpressing either of these genes were similar to non-transduced control cells by Wright-Giemsa staining (Figure 5B). The HSPC marker c-Kit, the early progenitor marker IL-3R (CD123) and markers of mature myeloid cells Ly6g and CD11b,

remained stable when compared to non-transduced leukemic progenitors (Figures 5C, S23). However, overexpression of these genes had a strong effect on cell growth (Figure 5D). Notably, Smad1 expression increased growth rate by 45% (decreased the doubling time from 13.7 to 9.5 hours), even more than Kras G12V expression (decreased the doubling time from 13.7 to 11.7 hours). The expression of Irx5 and Plag1 in an aggressive clone had strong effects in the opposite direction, suppressing growth rate by 62% (increase of doubling time from 10.1 to 26.2 hours) and 48% (increase of doubling time from 10.1 to 19.3 hours), respectively (Figure 5D).

AML patient data supports the clinical relevance of SMAD1, IRX5, and PLAG1

Survival analysis of public TCGA (https://www.cancer.gov/tcga) AML patient dataset implemented in BloodSpot[54] revealed significant association of higher IRX5 expression with longer overall survival (Figure 6A), consistent with Irx5 downregulation in aggressive mouse progenitors (Figure 4C,D) and cell growth suppression upon Irx5 overexpression (Figure 5D). Consistent with Smad1 upregulation in aggressive mouse progenitors (Figure 4E,F) and its effect on cell growth (Figure 5D), higher SMAD1 expression tended to correspond to shorter overall survival (Figure S24A), albeit the difference was not statistically significant. PLAG1 expression did not show significant survival associations (Figure S24B).

SMAD1, IRX5, and PLAG1 expression in leukemic progenitors from MLLr AML patients profiled with single-cell RNA-seq [16] was also consistent with our results. SMAD1 was downregulated while IRX5 and PLAG1 are upregulated in leukemic compared to normal GMPs (Figure 6B). The magnitude of these differences was somewhat lower but on a similar scale to the dysregulation of a major leukemic driver HOXA9 (Figure 6B).

According to the BloodSpot database [54], in normal human hematopoiesis SMAD1 is strongly expressed in HSCs and hematopoietic progenitors but is dramatically downregulated during further myeloid differentiation (Figure 6C). SMAD1 is also downregulated in various AML subtypes, especially MLLr AML (Figure 6C), consistent with Smad1 repression in mouse leukemic progenitors (Figures 4C,E, 6B). IRX5 has low expression levels across human HSCs, progenitors, and differentiated cells and in most AML subtypes except for t(15;17) PML-RARA fusion leukemia where IRX5 is strongly upregulated. PLAG1 is strongly expressed in human HSCs and gradually downregulated through stages of myeloid differentiation. Its expression is relatively low in all AML subtypes profiled in BloodSpot (Figure 6C).

Convergence of Smad1, Irx5, and Plag1 downstream transcriptional effects suggests a common regulatory network

To understand transcriptional effects of Smad1, Irx5 and Plag1 overexpression, we performed RNA-seq and compared genome-wide differential expression patterns (Figure 5, Supplementary Data 2). Consistent with their opposite effects on cell growth (Figure 5D), Smad1 effects were largely opposite to the effects of Plag1 and Irx5 (Figure 6D). These opposite effects were associated with targets of key regulatory proteins (Figure 6E): Polycomb components Kdm2b, Ring1b, Suz12, Ezh2 and Mtf2, central chromatin factors Brd4, Setdb1 and Arid3a, DNA methylation and demethylation components Dnmt1 and Tet1, and major transcription factors (Myc, Myb, Irf8, Ikzf1, Wt1, Rxr, Tal1, Sox2, Nanog, p53, Smad3 etc). There was a significant overlap between individual DEGs affected in the same direction by Irx5 and Plag1 (Figure 6F); for example, both upregulated Bcorl1, a Polycomb component strongly involved in leukemia [55,56] and downregulated Lin28b, which regulates hematopoiesis by transcriptionally inhibiting Polycomb component Cbx2 [57] (heatmap in Figure 6G and schematic in Figure 6H). A significant fraction of Irx5 and Plag1 targets was affected

by Smad1 in the opposite direction (Figure 6F), consistent with direct Smad1 downregulation by Plag1 overexpression (Figure 6G,H). For example, Kdm5b, Glis2, Smad3 and DNAjc6 were upregulated by Irx5 and Plag1 and downregulated by Smad1 (Figure 6H, upper left). H3K4me2/3 demethylase Kdm5b is a potent suppressor of AML aggressiveness [15]. Glis2 is a developmental transcription factor implicated in the control of AML cell differentiation [58]. Smad3 is a central mediator of signals from TGFβ-like proteins widely implicated in leukemia [59,60], whereas Smad1 mediates signals from BMP proteins [38]. Downregulation of Smad3 by Smad1 suggests an interesting negative crosstalk between BMP and TGFβ signaling, reminiscent of reported TGFβ mediated repression of BMP targets in various contexts [61]. Together with Smad3 upregulation by Irx5 and Plag1, this implies opposite roles of BMP and TGFβ pathways in controlling aggressiveness. Meis2, oncogenic factor in MN1-induced leukemia and AML1-ETO AML [62], and Bmx, a Tec kinase contributing to drug resistance in AML [63], were downregulated by Irx5 and Plag1 and upregulated by Smad1 (Figure 6G,H).

Smad1 overexpression downregulated central DNA methylation and chromatin remodeling factors, which were often co-regulated by Irx5 or Plag1 (Figure 6H, lower row). For example, DNA methylase Dnmt3a and histone deacetylase Hdac11 were downregulated by Smad1 and upregulated by Plag1; whereas key DNA demethylation components Tet1 and Tet2 were downregulated by Irx5 and Plag1, respectively. DNA methylation and demethylation machineries, especially DNMT3A and TET2 are strongly involved in myeloid malignancies including AML [7,64,65], consistent with the enrichment of DNA methylation and demethylation targets among DEGs associated with aggressiveness (Figure 2G) and with Irx5, Plag1 and Smad1 overexpression (Figure 6E).

Collectively, downstream effects of Smad1, Irx5, and Plag1 converged on key transcription factors (Figure 6H, black), chromatin remodelers (Figure 6H, red), and components of DNA methylation machinery (Figure 6H, blue). Aggressiveness driver Smad1 tended to downregulate repressive epigenetic factors (Polycomb proteins, histone deacetylases and demethylases, de novo DNA methylases), whereas aggressiveness suppressors Irx5 and Plag1 tended to have opposite effects. Involvement of these epigenetic repressors in the origin and maintenance of leukemia has been emerging in the literature [64], and our results suggest their involvement in the development of leukemic aggressiveness. At least one of these modifiers, histone H3K4 demethylase Kdm5b, is a direct suppressor of AML aggressiveness[15], suggesting that the broader set of Irx5, Plag1, and Smad1 targets may include new candidate regulators of leukemic aggressiveness.

DISCUSSION

Clonal evolution in blood cancers is driven by the selection of advantageous genetic and epigenetic alterations [1,5]. Genomic analyses suggest that although the origination of myeloid malignancies involves a relatively small number of mutations [66], the clonal evolution of aggressiveness is heterogeneous and can be driven by various additional genetic lesions in individual subclones [1,2,5] during unperturbed clonal evolution or as adaptation to therapy [3,4]. Disease phenotype, prognosis, and therapy response are also affected by the temporal order of these lesions [67] and progenitor stage where they appear [68], recently studied at single-cell resolution [69,70]. This heterogeneity is also apparent in single-cell gene expression [16,17]. The dynamics of chromatin landscape in clonal evolution have not been studied to the same extent, despite strong evidence of the essential role of chromatin modifiers.

Here we analyzed genome-wide dynamics of transcription and chromatin marks during the development of clonal aggressiveness and showed how these dynamics can reinforce a shared cellular program that enables clonal growth advantage. These dynamics were more clone-specific than upon initial leukemogenesis, consistent with transcriptional heterogeneity among human AML clones [16,17]. However, there was a considerable overlap between these clonal changes. The global skew towards increasing H3K4 methylation at promoters and enhancers (Figures 2J and 3H) was consistent with the reported role of H3K4me2/3 demethylase Kdm5b as a suppressor of clonal AML aggressiveness [15]. The global skew towards increasing H3K27 acetylation at enhancers, which could be uncoupled from H3K4 methylation (Figure 3H), was consistent with the role of HATs and HDACs in AML.

Unexpectedly, most of the involved promoters and enhancers were already affected at the early leukemic stage of the same clone (Figures 2H, 3F). More surprisingly, rather than developing further, these early changes were most often reversed at the aggressive stage (Figures 2I, 3G). This reversal is consistent with the reported changes in the roles of some central leukemic regulators at later stages of AML; Polycomb component Ezh2 is a strong tumor suppressor during AML induction and is later switched to an oncogenic role during AML maintenance [14], and Igf1r is a major oncogene during AML induction but is not required during later clonal growth [52]. The biological role of this reversal is an important question. As one possible explanation, the initial leukemogenic lesion (MLL rearrangement) causes perturbations across a wide variety of direct and indirect genomic targets (Figure 1D,L), some of which (activation of Hoxa9, Meis1, etc) cause oncogenic transformation, while others may be non-essential and even maladaptive for later clonal fitness and growth. In this case, scaling back maladaptive initial effects (Figure 2C, 3C, 4D,F) while keeping strong initial activation of leukemogenic drivers (Figure 1C) may provide a growth

advantage. This reversal may be achieved through the regulation of promoters and enhancers by transcription factors and chromatin regulators.

Focusing on consistent changes of expression and chromatin marks in both clones in vivo and ex vivo revealed the core gene set (Figures 4A,B), which strongly followed this reversal trend, often paralleled by nearby enhancers (Figures 4G,H). We selected three of these genes for experimental validation and confirmed their causal roles as leukemic cell growth suppressors (Irx5 and Plag1) and activator (Smad1). These roles are consistent with public data in human hematopoiesis and AML. Irx5 is a curious example of a strong MLL-AF9 target [49] that was released from Polycomb repression and dramatically activated at the early leukemic stage, like known major leukemogenic drivers (Figure 4D). However, unlike these drivers, at the aggressive stage Irx5 reversed this activation by increased Polycomb repression (Figure 4D). Plag1 is an oncogene derepressed at the induction of AML [14,50], but it's surprising opposite role in suppressing clonal aggressiveness has not been previously reported. SMAD1 has not been reported as a major leukemic regulator, although it is a central effector of the BMP pathway implicated in leukemia [38,39]. Smad1 was the only Smad family member with this behavior, which positions BMP pathway as a possible therapeutic target distinct from the related TGFβ pathway.

Downstream transcriptional effects of these three genes upon their overexpression converge on a network of transcription factors, chromatin modifiers, and components of DNA methylation machinery, many strongly implicated in leukemia (Figure 6H). The effects on Smad3, a key effector of TGFβ pathway (Figure 6H), may suggest that Smad3-mediated TGFβ signaling is inhibited and replaced by Smad1-mediated BMP signaling in aggressive subclones. This inhibition is in line with

aggressivenss-associated regulation of other TGFβ related genes (Peg10, Matn2, Ptpn3, Figure 4B).

Coordinated downstream regulation also affects major epigenetic factors: H3K4me2/3 demethylase Kdm5b, Polycomb component Bcorl1 [55,56], as well as DNA methylase Dnmt3a [7] and DNA demethylation component Tet2 [8], highlighting the role of repressive Polycomb, H3K4 demethylation, and DNA methylation machineries. Modulation of Polycomb repression was a major mechanism of promoter regulation at both early and aggressive leukemic stages, for example at *Smad1*, *Irx5*, and *Plag1* (Figure 4C-F). The reported role of Kdm5b as aggressiveness suppressor in leukemia [15] was consistent with its transcriptional regulation by Smad1, Irx5, and Plag1 (Figure 6H) and the global skew towards increased promoter (Figures 2J) and enhancer H3K4 methylation (Figure 3H) at the aggressive stage. DNA methylation and demethylation components with known strong AML involvement[7,8] were regulated by Smad1 and Plag1 in opposite directions (Figure 6H). This collective involvement of epigenetic repressors is consistent with their known functional crosstalk. In addition to antagonism between Polycomb and H3K4 methylation [23], as well as direct Polycomb repression of Kdm5b [15], there is growing evidence of direct Polycomb-DNMT3A interactions [71] contributing to leukemia [72].

Taken together, our results suggest that development of aggressiveness is a clone-specific process often mediated through gradual adjustment of chromatin state and gene expression, as opposed to on/off switch behavior typical for major leukemogenic drivers. This adjustment often involves a surprising reversal of prior expression and chromatin state changes at the early leukemic stage. We identified a set of core genes with aggressiveness-associated changes consistent between clones and between in vivo and ex vivo conditions, and successfully validated three genes as causal

regulators of leukemic cell growth with convergent downstream transcriptional effects, suggesting new candidate pathways in regulating leukemic aggressiveness.

ACKNOWLEDGEMENTS

R.I.S. was supported by NIH NIDDK P30 DK040561. DBS was supported by grants from the St. Baldrick's Foundation and the American Society of Hematology.

COMPETING INTERESTS

DBS is a co-founder and holds equity in Clear Creek Bio.

MATERIALS AND METHODS

Murine models and cell lines

Experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of the Massachusetts General Hospital. C57BL/6J, B6.SJL-Ptprca Pepcb/BoyJ (CD45.1) were purchased from Jackson Laboratories. All MLL-AF9 leukemic cell lines were obtained as previously described [73] and maintained in RPMI1640 (10% FBS), penicillin/streptomycin, 6 ng/ml rmIL-3, and approximately 100 ng/ml SCF. Briefly, a MLL-AF9 CD45.1 parental line was established by sorting, using a FACS AriaII (BD Biosciences), granulocyte-macrophage progenitors (GMP; c-Kit^{high}Sca-1^{-} Gr-1^{low}CD16/32^{+}CD34^{+}) from a male CD45.1 donor. Retrovirally transfection with a MSCV-MLL/AF9-IRES-NeoR construct in sorted progenitor cells was performed a day after incubation in RPMI1640 (Lonza) transduction medium containing 10% (FBS; Gibco), 100 I.U./ml penicillin, 100 mg/ml streptomycin (Gibco), 10 ng/ml recombinant mouse SCF (rmSCF), 6 ng/ml recombinant mouse interleukin 3 (rmIL-3), 5 ng/ml rmIL-6 (all from Peprotech). Progenitors were transferred to RetroNectin®-coated (Lonza) 6-well plates in a volume of 1 ml, combined with 1 ml of fresh retroviral supernatant packaged in 293FT cells (Invitrogen) with pCL-Eco (Addgene 12371) and 8 µg/ml final concentration of polybrene, then spinoculating (1,000g, 90 min, 22°). Immediately following the transduction, 4 ml of fresh transduction medium was added to each well to dilute the polybrene and 24 hours post-transduction, cells were washed by centrifugation (500g, 5 min, 4°C) then resuspended in fresh transduction medium. 48 hours post-transduction, neomycin resistant lines were selected using 50 µg/ml neomycin (Gibco).

Clonal fluorescent labeling

To generate MLL-AF9 clones expressing fluorescent proteins (FP), single MLL-AF9 CD45.1 clones were sorted using a FACS Aria II (BD Biosciences) into 96-well plates. A single clone was expanded and transduced with lentiviral constructs EF1a-FP-SV40-PuroR, generated by Gateway cloning of FP Entry plasmids into the pLEX_307 Destination construct (a gift from David Root, Addgene plasmid # 41392). From each FP transduction, single subclones were expanded in vitro and those that stably expressed FPs were preserved.

Syngeneic leukemia experiments

All leukemia transplantations were performed by injecting cells intravenously (tail vein) in a volume of 300 ul of phosphate-buffered saline (PBS) in recipient C57Bl/6J male mice (age 8-10 weeks) that were sublethally irradiated (4.5Gy) the day prior using a Cesium-137 source. Blood was collected through tail vein sampling and complete blood counts were done on a Heska Element HT5 analyzer. Bone marrow aspiration was done under isoflurane anaesthesia using a 25G needle attached to a 1 ml syringe rinsed with PBS, gently inserted into the femoral cavity through the articular surface of the knee; approximately 10 µl of bone marrow was gently aspirated in the syringe and immediately resuspended in 300 µl of PBS on ice. Hematopoietic cell isolation was done as follows: after euthanasia by asphyxiation with CO_2, vertebrae and the long bones of the arms and legs were dissected and crushed into PBS containing 2% FBS (FACS buffer). The bone marrow mononuclear cells were isolated over a Ficoll-Paque PLUS density gradient and counted using trypan blue and the Cellometer instrument.

Overexpression of gene candidates

Leukemic subclones were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum, 1% L-glutamine, 1% penicillin/streptomycin, 25 nG/mL SCF, and 5 nG/mL IL-3. ORFs of interest were overexpressed using second-generation lentiviral system that consisted of pCMV-VSV-G (Weinberg lab) envelope plasmid and PsPAX2 (Trono lab) packaging plasmid acquired from

Addgene. HEK293t cells were transfected with plasmid DNA at a ratio of 5:4:1 transfer plasmid:envelope plasmid:packaging plasmid. Transfection efficiency was aided by the addition of Promega's Fugene6 Reagent at a ratio of 3:1 to plasmid DNA. HEK293t cells were cultured in DMEM, high glucose, pyruvate supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. Lentiviral titer was assessed by FACS. Mouse leukemic progenitors were transduced by exposing them to viral particles and room temperature centrifugation at 1000g for 1 hr in 8uG/mL polybrene (Millipore Sigma). Spinfection plates were pre-coated using 500 uL of PBS with 10 uG/mL of purified human plasma fibronectin (Sigma-Aldrich) and incubated at 37°C several hours prior to transduction. Single cell FACS sorting was performed 48 hours after transduction to isolate individual GFP-positive clones. Overexpression of the gene of interest in selected clones was confirmed by RT-qPCR and RNA-seq. RNA was extracted using TRIzol LS Reagent, followed by cDNA synthesis using SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen).

Cell growth quantitation

Cell growth was quantitated using CellTrace Far Red Cell Proliferation Kit (Invitrogen), Initially, on day 1, cell lines were standardized to a density of 1,000,000 cells/mL in 1mL of PBS, labeled with CellTrace™ Far Red dye at a 1 μM concentration, and subsequently centrifuged at 500g for 5 minutes and resuspended in fresh media. The labeled cells were then cultured in 6-well plates containing 5 mL of RPMI 1640 media supplemented with 10% Fetal Bovine Serum. Starting on day 2, daily sampling was conducted by aspirating a fraction of the culture, centrifuging at 500g for 5 minutes, and resuspending the cell pellet in 400 μL FACS Buffer containing 0.8% formaldehyde for overnight fixation. The following day, formaldehyde fixation was neutralized with the addition of 100 μL FBS. These fixed samples were stored at 4°C until the final timepoint. Flow cytometric analysis was performed on a BD FACS Melody system, where the mean fluorescence intensity of each

sample was measured at every timepoint. The doubling time was inferred by calculating the time interval required for the fluorescence intensity to halve, reflecting cell proliferation rates.

Cell sorting to assess hematopoietic lineage

Cellular suspensions were prepared for flow cytometric analysis by harvesting 1 million cells and resuspending them in 100 µL of FACS Buffer, consisting of 2% fetal bovine serum, 1 mM EDTA, and phosphate-buffered saline. The cell suspensions were then labeled with lineage related antibodies: PE/Dazzle™ 594 anti-mouse/human CD11b (Biolegend 101256), APC anti-mouse Ly-6G/Ly-6C (Gr-1) (Biolegend 108412), PE anti-mouse CD123 (Biolegend 106005), and PE/Cyanine7 anti-mouse CD117 (c-kit) (Biolegend 135112), at the appropriate concentrations for the volume and cell count (1uL in 1 million cells), followed by incubation at room temperature for 30 minutes in the dark. Post-incubation, cells were centrifuged at 500g for 5 minutes, washed twice with FACS Buffer, and finally resuspended in 400 µL of the same buffer. The cell suspensions were then strained through a 70 µm mesh into FACS tubes to ensure a single-cell suspension. Single-color controls were established using OneComp eBeads™ (Thermofisher 01-1111-41). Analysis was performed in FlowJo, with doublet discrimination and fluorescent compensation.

RNA-seq

Total RNA was isolated from three biological replicates of each cell type using TRIzol (Invitrogen) and DNase treated. Sequencing libraries were prepared using polyA-selection followed by NEBNext Ultra Directional RNA library preparation protocol (New England Biolabs). Sequencing was performed on Illumina HiSeq 2500 instrument, resulting in approximately 30 million 50 bp reads per sample.

RNA-seq data analyses

Sequencing reads were mapped to the mouse transcriptome (mm9 EMSEMBL annotation) using STAR [74]. Gene expression counts were calculated using HTSeq v.0.6.0 [75]. Calculation of

expression values and differential expression analysis were performed using EgdeR package [76] after normalizing read counts and including only genes with count per million reads (CPM) > 1 for one or more samples. Differentially expressed genes were defined based on the criteria of > 2 fold change of expression and FDR < 0.05. Analysis of gene set enrichment was performed using Enrichr [22].

Chromatin immunoprecipitation sequencing (ChIP-seq)

Cells were crosslinked for 10 minutes in 1% formaldehyde at room temperature and quenched with 125mM Glycine. ChIP was performed as previously described [77]. Antibodies used: H3K4me1 (Abcam ab8895), H3K4me3 (Millipore, 04-745), H3K27me3 (Cell Signaling Technology, C36B11), H3K27ac (Abcam ab4729). Samples were further processed for high throughput sequencing. Sequencing was performed on Illumina HiSeq 2500 instrument resulting in approximately 30 million paired-end 50 bp reads per sample.

ChIP-seq data analysis

Reads were aligned against the mm9 reference genome using BWA [65]. Alignments were filtered for uniquely mapped reads and alignment duplicates were removed. Input-normalized and CPM normalized coverage tracks were generated using the bamcompare function of deepTools [78]. Enrichment over TSS-proximal regions (TSS ± 3 Kb) was calculated as the ratio of ChIP and input read counts scaled by library size. Reads of publicly available MLL-AF9 ChIP-seq (GEO accession number GSE29130) were aligned as described previously, and peaks were identified using Homer with the “factor” option. Strong targets genes of MLL-AF9 were identified by filtering the peak list using a peak enrichment cutoff >=13. Putative enhancer elements were identified using H3K27ac and H3K4me1 ChIP-Seq peaks. In brief, we analyzed tag counts in sliding 1 Kb windows with the step of 200 bp and estimated statistical significance of enrichment of ChIP vs input using negative binomial distribution with the parameter of mean based on input tag count and the size parameter

(s) selected based on manual inspection of peak calls. Regions of significant enrichment were generated by merging adjacent significantly enriched windows within 1 Kb. To define putative enhancer elements, peak sets of H3K27ac and H3K4me1 were merged, followed by removing regions within 3 Kb from annotated promoters and regions overlapping H3K4me3 peaks.

Analyses of public human single-cell RNA-seq data

Publicly available scRNA-seq read counts from GEO dataset GSE116256 [16] were analyzed using Seurat 3.2.3 package [79]. We used the patient samples with > 300 cells annotated as GMP or GMP-like by the authors. Cells with <200 expressed genes and genes expressed in <3 cells were filtered out, followed by the exclusion of cells with high content of mitochondrial transcripts (> 10% of total reads). Counts across all cells for each sample were normalized using NormalizeData function in Seurat. Individual sub-samples were integrated using Seurat canonical correlation analysis (CCA) method. Integration anchors were determined using FindIntegrationAnchors function and then used in IntegrateData function. We used cell annotations provided in the study to visualize expression of genes of interest in GMP and GMP-like cells using DotPlot function of Seurat.

FIGURE LEGENDS

Figure 1. Chromatin and gene expression changes between normal and early leukemic progenitors were consistent between independent clones and mostly maintained through the later aggressive stage.

A. A mix of leukemic clones individually labeled with different flurorescent proteins was transplanted in a primary mouse recipient, followed by clonal evolution for 120 days and collection of aggressive subclones, which maintained their aggressiveness upon secondary transplantation. The early and aggressive stages of two leukemic clones were profiled by RNA-seq in vivo and ex vivo, and by ChIP-seq of histone marks ex vivo, and compared to normal progenitors. **B.** Promoter chromatin state of a major leukemogenic driver Meis1 was dramatically activated in leukemic cells compared to normal but showed no strong difference between two stages of aggressiveness. ChIP-seq genomic tracks of H3K27me3 (red), H3K4me3 (blue), and H3K27ac (green) in normal progenitors (top) compared to two independent leukemic clones 1 and 2 at early and aggressive stages. **C.** Known drivers of MLLr leukemia were activated at the early stage but did not show further changes at the aggressive stage. Promoter chromatin state progression for Meis1 (tracks in **B**), HoxA9, and Six1 (tracks in Figure S2) in the space of H3K27me3 (x axis) vs H3K4me3 density (y axis) at TSS-proximal region. Normal progenitors and early and aggressive stages of two leukemic clones (1 and 2) are represented as grey, blue, and red points, respectively. Gene expression levels (RNA-seq RPKM) are shown as barplots. Early loss of repressive H3K27me3 mark accompanied by strong gain of active H3K4me3 mark (arrows) and upregulation of expression were later maintained at the aggressive stage. **D.** Heatmap of early changes (log2 fold change) of promoter chromatin marks across all promoters with a substantive absolute level and > 2-fold change of at least one mark, juxtaposed with the changes of gene expression in vivo (log2 fold change, right). Changes

were strongly consistent in two independent clones (two columns for each mark and expression). **E.** Venn diagrams of a strong overlap between the early stages of two independent clones (Early 1, Early 2) among promoters with differential gene expression in vivo (purple) and chromatin mark density (green, blue, and red) compared to normal progenitors. **F.** Functional category enrichment among genes differentially expressed between normal and early leukemic progenitors in vivo consistently in both clones. Bars colored by category types represent adjusted P-value of enrichment according to EnrichR. **G.** In both clones, early-stage changes were skewed towards the increase of active H3K27ac and H3K4me3 promoter marks and gene expression and decrease of repressive mark H3K27me3. Barplots of promoter numbers with increase (red) and decrease (blue) of each individual mark and gene expression at the early stage in each leukemic clone (Early 1, Early 2). **H.** Most of the early changes of promoter chromatin marks and expression were later maintained at the same level at the aggressive stage. Scatterplots of chromatin mark changes (log2 fold change) at all promoters between normal cells and the early leukemic stage (y axis) compared to the changes between normal cells and the aggressive leukemic stage (x axis). Individual promoters are represented as points, color indicates point density. **I.** Genome-wide trends among the early leukemic changes of promoter chromatin states. All promoters across the genome are represented as points in the space of H3K27me3 (x axis) vs H3K4me3 density (y axis) at TSS-proximal region. Each point is colored by gene expression in normal progenitors, from blue (silent) to red (active). Upper left: active promoters enriched in H3K4me3 and depleted in H3K27me3. Middle right: bivalent promoters with high H3K27me3 and moderate H3K4me3. Lower left: silent promoters with both marks depleted. Arrows with corresponding promoter numbers indicate major directions of H3K4me3/ H3K27me3 changes. **J.** Example of ChIP-seq tracks at two enhancers that changed chromatin state at the early leukemic stage. H3K27ac (green) and H3K4me1 (dark green) densities in normal progenitors (top) compared to two independent leukemic clones 1 and 2 at the

early and aggressive stage. **K.** Chromatin state progression at enhancers proximal to Des (shown in **J**), Lmo1, and Pitx2 genes in the space of H3K4me1 (x axis) vs H3K27ac density (y axis). Points represent normal progenitors (grey) and the early (blue) and aggressive (red) stages of two leukemic clones (1 and 2). **L.** Heatmap of the early leukemic changes (log2 fold change) of H3K4me1 and H3K27ac densities across all enhancers with a substantive absolute level and > 2-fold change of at least one mark, juxtaposed with the changes of proximal gene expression in vivo (log2 fold change, right). Changes were largely consistent between marks and between two independent clones (two columns for each mark and expression). **M.** Venn diagrams of a strong overlap between between two clones (Early 1, Early 2) among enhancers with chromatin mark changes between normal progenitors and the early leukemic stage. **N.** Most of the early changes of enhancer chromatin marks were later maintained at the aggressive stage. Genome-wide scatterplots of chromatin mark changes (log2 fold change) at all enhancers between normal cells and the early leukemic stage (y axis) compared to the changes beteen normal cells and the aggressive leukemic stage (x axis). Individual enhancers are represented as points; color indicates point density. **O.** Genome-wide trends among the early leukemic changes of enhancer chromatin states. All enhancers across the genome are represented as points in the space of H3K4me1 (x axis) vs H3K27ac density (y axis). Each point is colored by proximal gene expression in normal progenitors, from blue (silent) to red (active). Arrows with corresponding enhancer numbers indicate major directions of H3K4me1/ H3K27ac changes.

Figure 2. Changes of promoter chromatin state and expression between the early and aggressive leukemic stages were more clone-specific, gradual, and often reversed the early leukemic changes.

A,B. Examples of ChIP-seq genomic tracks showing promoter chromatin mark and expression changes between two stages of aggressiveness. H3K27me3 (red), H3K4me3 (blue), and H3K27ac (green) densities in normal progenitors (top) compared to two independent leukemic clones 1 and 2 at early and aggressive stages. Compared to normal progenitors, Fhdc promoter was activated (**A**) and Fyb promoter was repressed (**B**) to a different degree between the early and aggressive stage of the same clone. **C.** Promoter chromatin state progression for Fhdc and Fyb promoters (tracks shown in **A, B**) in the space of H3K27me3 (x axis) vs H3K4me3 density (y axis) at TSS-proximal region. Points represent normal progenitors (grey) and early (blue) and aggressive stages (red) of two leukemic clones (1 and 2). Strong opposite shifts of H3K27me3 and H3K4me3 at the early stage of both clones (blue arrow) were later partially reversed at the aggressive stage (red arrow), concomitant with changes of gene expression (barplots of RNA-seq RPKM). **D.** Heatmap of changes (log2 fold change) of promoter chromatin marks between the early and aggressive stages across all promoters with a substantive absolute level and > 2-fold change of at least one mark, juxtaposed with the changes of gene expression in vivo (log2 fold change, right). Changes were only partially consistent between two independent clones (two columns for each mark and expression). **E.** Principal component analysis (PCA) plot of gene expression in normal progenitors (grey) compared to early (blue) and aggressive stage (red) of two leukemic clones (1 and 2). **F.** Venn diagrams of moderate overlap between two independent clones (Aggressive 1, Aggressive 2) among promoters with changes of gene expression in vivo (purple) and chromatin marks (green, blue, and red) at the aggressive stage. **G.** Functional category enrichment among genes differentially expressed between the early and aggressive stages in vivo consistently in both clones. Bars colored by category types represent adjusted P-value of enrichment according to EnrichR. **H.** Most promoters affected at the aggressive stage underwent prior changes at the early leukemic stage. Venn diagrams of promoters with changes of expression and individual chromatin marks at

the aggressive and early stages of the same clone. **I.** Among promoters that underwent consequent changes at both leukemic stages, the majority reversed their initial early changes. Large arrows (Early 1, Early 2): numbers of promoters with increase (red) and decrease (blue) of expression or a chromatin mark at the early leukemic stage of a clone (1 and 2). Adjacent smaller arrows (Aggressive 1, Aggressive 2): numbers of these promoters with the later increase (red) or decrease (blue) of expression or chromatin mark at the aggressive stage. **J.** Genome-wide trends among changes of promoter chromatin states at the aggressive stage (representative example of one clone). All promoters across the genome are shown as points in the space of TSS-proximal H3K27me3 (x axis) vs H3K4me3 densities (y axis) at the early leukemic stage, colored by the level of gene expression from blue (silent) to red (active). Arrows with corresponding promoter numbers indicate major directions of H3K4me3/ H3K27me3 changes at the aggressive stage. Yellow, preferential skews that were consistent between clones.

Figure 3: Enhancer chromatin remodeling at the aggressive stage was more clone-specific, skewed towards activation, and often reversed early leukemic changes.

A,B. Examples of ChIP-seq genomic tracks showing enhancer chromatin mark changes between two stages of aggressiveness. H3K27ac (green) and H3K4me1 (dark green) densities in normal progenitors (top) compared to two independent leukemic clones 1 and 2 at the early and aggressive leukemic stages. Chromatin state of an intronic enhancer in Net1 gene became more active (**A**) and chromatin state of an enhancer proximal to gene Zfp516 became less active at aggressive compared to early stage in both clones (**B**). **C.** Chromatin state progression in the space of H3K27ac vs H3K4me1 density at enhancers shown in **A, B.** Points represent normal progenitors (grey) and the early (blue) and aggressive stages (red) of two leukemic clones (1 and 2). Changes between the early and aggressive stages (red arrows) were accompanied by proximal gene expression

changes (barplots of RNA-seq RPKM). **D.** Heatmap of H3K27ac and H3K4me1 density changes (log2 fold change) at the aggressive stage across all enhancers with a substantive absolute level and > 2-fold change of at least one mark, juxtaposed with the changes of gene expression in vivo (log2 fold change, right). Changes were only partially consistent between two independent clones (two columns for each mark). **E.** Venn diagrams of moderate overlap between two independent clones (Aggressive 1, Aggressive 2) among enhancers with changes of H3K27ac (green) and H3K4me1 (dark green) between the early and aggressive leukemic stages. **F.** Most enhancers affected at the aggressive stage underwent prior changes at the early stage. Venn diagrams of enhancers with changes of H3K27ac and H3K4me1 at the aggressive and early stages of the same clone. **G.** Among enhancers that underwent consequent changes at both leukemic stages, the majority reversed their initial early changes. Large arrows (Early 1, Early 2): numbers of enhancers with the early increase (red) and decrease (blue) of H3K27ac in clones 1 and 2. Adjacent smaller arrows (Aggressive 1, Aggressive 2): numbers of these enhancers with the later increase (red) or decrease (blue) of H3K27ac at the aggressive stage. **H.** Genome-wide trends among changes of enhancer chromatin states at the aggressive stage (representative example of one clone). All enhancers across the genome are shown as points in the space of H3K27me3 (x axis) vs H3K4me3 density (y axis) at the early leukemic stage, colored by the level of proximal gene expression from blue (silent) to red (active). Arrows with corresponding enhancer numbers indicate major directions of H3K4me1/ H3K27ac changes between the early and aggressive stages. Yellow, preferential skew towards H3K27ac increase reversed the opposite early trend (Figure 1O). **I.** Enrichment of transcription factor binding motifs associated with enhancer mark changes at the aggressive stage. MEME enrichment E-values for transcription factor motifs (rows) shown for enhancers changing the level of a given chromatin mark in a given clone (columns). **J.** Enhancer H3K27ac level is often quantitatively correlated with the expression of the proximal gene. Correlation scatterplots of

enhancer H3K27ac density vs proximal gene expression across normal progenitors (grey), early (blue) and aggressive leukemic stages (red) for two enhancers shown in **A-C**. Pearson correlation coefficient and linear regression P-value are indicated. **K.** Enhancers increasing H3K27ac at the aggressive stage show preferential positive correlation of H3K27ac level with proximal gene expression. Histogram of all Pearson correlation coefficients between enhancer H3K27ac and gene expression (calculated as in **J**) is strongly skewed towards higher R and includes a distinct enhancer subset with strong correlation (R > 0.7, dotted line). **L.** Functional category enrichment among genes proximal to the enhancer subset shown in **K**. Bars colored by category types represent adjusted P-value of enrichment according to EnrichR.

Figure 4. Core genes with consistent aggressiveness-associated differences of promoter chromatin state ex vivo and expression in vivo in both independent clones.

A. Filtering genes by the consistency of aggressiveness-associated changes of gene expression in vivo (yellow) vs active H3K27ac and H3K4me3 (blue) and repressive H3K27me3 (red) promoter marks ex vivo in both clones. Numbers of promoters with expression or chromatin mark increase and decrease are indicated in red and blue, respectively. A core set of 28 candidate genes had fully consistent behavior across all comparisons. **B.** Heatmap of promoter H3K27me3, H3K4me3, H3K27ac and expression changes (fold change in log2 scale) for the core gene set between the early and aggressive stage of two clones. **C.** Examples of genomic ChIP-seq tracks at Irx5 (left) and its proximal enhancer (right) where the magnitude of the early leukemic activation was later adjusted at the aggressive stage. **D.** Examples of promoter chromatin state progression in the space of H3K27me3 (x axis) vs H3K4me3 (y axis) for core genes Irx5 (tracks in **C**) and Plag1 (tracks in Figure S22B). Points represent normal progenitors (grey) and early (blue) and aggressive stages (red) of two leukemic clones (1 and 2). H3K27me3 loss and H3K4me3 gain at the early leukemic

stage of both clones (left plots, blue arrows) were later partially reversed at the aggressive stage (left plots, red arrows), consistent with in vivo expression changes (right barplots). These changes were paralleled by H3K27ac and H3K4me1 at a proximal enhancer (middle plots). **E.** Examples of genomic ChIP-seq tracks at Smad1 (left) and its proximal enhancer (right) where the magnitude of the early leukemic repression was later adjusted at the aggressive stage. **F.** Examples of promoter chromatin state progression in the space of H3K27me3 vs H3K4me3 for core genes Smad1 (tracks shown in **E**) and Rxra (tracks in Figure S22C). H3K27me3 gain and H3K4me3 loss at the early leukemic stage of both clones (left plots, blue arrows) were later partially reversed at the aggressive stage (left plots, red arrows), consistent with in vivo expression changes (right barplots). These changes were paralleled by H3K27ac and H3K4me1 at a proximal enhancer (middle plots). **G.** Virtually all core aggressiveness-associated genes underwent prior changes at the early leukemic stage. Venn diagrams of promoters with aggressiveness-associated (Agg) and early leukemic changes in the same clone (Early) for expression and individual chromatin marks. **I.** Among the core genes that underwent consequent changes at both leukemic stages, the majority reversed their initial early changes. Large arrows (Early 1, Early 2): numbers of promoters with the early increase (red) and decrease (blue) of expression or a chromatin mark in clones 1 and 2. Adjacent smaller arrows (Agg 1, Agg 2): numbers of these promoters with later increase (red) or decrease (blue) of expression or chromatin mark at the aggressive stage.

Figure 5. Effects of overexpressing selected core genes validate their causal roles

A. Smad1 and Kras G12V mutant (used as a control) were overexpressed in the early stage subclone, whereas Irx5 and Plag1 were overexpressed in the aggressive stage subclone of the same leukemic clone. Barplots of RNA-seq gene expression values (RPKM) in transduced compared to control untransduced cells (n=3 biological replicates). **B.** Representative images of

Wright-Giemsa–stained cells transduced with each gene, compared to control. Scale bar, 1 micron. **C.** Flow cytometry analysis of c-Kit, Gr1, CD123, and CD11b expression in each transduced subclone compared to control suggest no changes of hematopoietic lineage. **D.** Overexpression of Smad1 increased, whereas overexpression of Irx5 and Plag1 decreased cell growth. Ex vivo cell growth curves and the corresponding doubling times for transduced cells compared to untransduced controls. Data are represented as mean +/- SEM.

Figure 6. The roles of Smad1, Irx5, and Plag1 are supported by AML patient data and convergent downstream effects on a common regulatory network.

A. High IRX5 expression is significantly associated with longer overall survival in AML patients. Analysis of TCGA AML patient cohort implemented in BloodSpot. Kaplan-Meier curves for AML patients with IRX5 expression above (red) and below median (green). **B.** Analysis of previously published single-cell RNA-seq data in granulocyte/macrophage progenitors (GMPs) from AML patients [16] confirms lower expression of SMAD1 and higher expression of IRX5 and PLAG1 in human leukemic progenitors. These differences are smaller but similar in scale to the expression differences for a major leukemic driver HOXA9. **C.** SMAD1, IRX5, and PLAG1 expression levels in public data from normal human myeloid lineage (GEO GSE42519) and various AML types (GEO GSE13159), analyzed in BloodSpot. SMAD1 is highly expressed whereas IRX5 and PLAG1 are silent in normal human GMPs. SMAD1 is downregulated whereas IRX5 is upregulated specifically in MLLr AML t(15;17). PLAG1 expression is low across AML types. **D.** Genome-wide transcriptional changes caused by Smad1 overexpression in mouse leukemic progenitors are largely opposite to the changes caused by Irx5 and Plag1 overexpression. Heatmap of RNA-seq expression changes compared to control (log2 fold change) among genes differentially expressed (> 2 fold change, FDR < 0.05) in any of the three transduction experiments. **E.** Overlap of functional categories enriched

among genes up- and downregulated (bars to the right and left of dotted line, respectively) by Irx5, Plag1, and Smad1 overexpression. Bars colored by category types represent 10log10 of adjusted P-value of enrichment according to EnrichR. **F.** Overlaps between downstream transcriptional effects of the three overexpressed genes. Venn diagrams of the genes downregulated by Plag1 and Irx5 and upregulated by Smad1 (top), and of the genes upregulated by Plag1 and Irx5 and downregulated by Smad1 (top). Hypergeometric P-values for pairwise overlaps are indicated. **G.** Heatmap of RNA-seq expression changes compared to control (log2 fold change) for selected downstream targets consistently affected by Smad1, Irx5, and Plag1 overexpression. **H.** Smad1, Irx5, and Plag1 effects suggest a common downstream regulatory network (schematic).

REFERENCES

1. Ferrando, A. A. & López-Otín, C. Clonal evolution in leukemia. *Nat. Med.* **23**, 1135–1145 (2017).

2. Klco, J. M. *et al.* Functional Heterogeneity of Genetically Defined Subclones in Acute Myeloid Leukemia. *Cancer Cell* **25**, 379–392 (2014).

3. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).

4. Quek, L. *et al.* Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. *Nat. Med.* **24**, 1167–1177 (2018).

5. Takahashi, K. & Tanaka, T. Clonal evolution and hierarchy in myeloid malignancies. *Trends cancer* (2023) doi:10.1016/j.trecan.2023.05.004.

6. Li, S. *et al.* Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat. Med.* **22**, 792–799 (2016).

7. Brunetti, L., Gundry, M. C. & Goodell, M. A. DNMT3A in Leukemia. *Csh Perspect Med* **7**, a030320 (2017).

8. Bacher, U. *et al.* Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Ann. Hematol.* **89**, 643–652 (2010).

9. Cenik, B. K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and disease. *Nat. Rev. Genet.* **22**, 38–58 (2021).

10. Chen, Y. *et al.* MLL2, Not MLL1, Plays a Major Role in Sustaining MLL-Rearranged Acute Myeloid Leukemia. *Cancer Cell* **31**, 755-770.e6 (2017).

11. Cao, F. *et al.* Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Mol. Cell* **53**, 247–261 (2014).

12. Hughes, C. M. *et al.* Menin Associates with a Trithorax Family Histone Methyltransferase Complex and with the Hoxc8 Locus. *Mol. Cell* **13**, 587–597 (2004).

13. Neff, T. *et al.* Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc National Acad Sci* **109**, 5028–5033 (2012).

14. Basheer, F. *et al.* Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J Exp Medicine* **216**, 966–981 (2019).

15. Ren, Z. *et al.* A PRC2-Kdm5b axis sustains tumorigenicity of acute myeloid leukemia. *P Natl Acad Sci Usa* **119**, e2122940119 (2022).

16. Galen, P. van *et al.* Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281.e24 (2019).

17. Zeng, A. G. X. *et al.* A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med* **28**, 1212–1223 (2022).

18. Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. *Nature* **442**, 818–822 (2006).

19. Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273–284 (2018).

20. Somervaille, T. C. P. & Cleary, M. L. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* **10**, 257–268 (2006).

21. Krivtsov, A. V., Hoshii, T. & Armstrong, S. A. Mixed-Lineage Leukemia Fusions and Chromatin in Leukemia. *Csh Perspect Med* **7**, a026658 (2017).

22. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).

23. Schuettengruber, B., Bourbon, H.-M., Croce, L. D. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).

24. Sadreyev, R. I., Yildirim, E., Pinter, S. F. & Lee, J. T. Bimodal quantitative relationships between histone modifications for X-linked and autosomal loci. *Proc. Natl. Acad. Sci.* **110**, 6949–6954 (2013).

25. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).

26. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

27. Bernt, K. M. *et al.* MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).

28. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

29. Hock, H. & Shimamura, A. ETV6 in hematopoiesis and leukemia predisposition. *Semin. Hematol.* **54**, 98–104 (2017).

30. Endo, A. *et al.* EWSR1/ELF5 induces acute myeloid leukemia by inhibiting p53/p21 pathway. *Cancer Sci.* **107**, 1745–1754 (2016).

31. Duprez, E., Wagner, K., Koch, H. & Tenen, D. G. C/EBPbeta: a major PML-RARA-responsive gene in retinoic acid-induced differentiation of APL cells. *EMBO J.* **22**, 5806–16 (2003).

32. Wang, Q. & Friedman, A. D. CCAAT/enhancer-binding proteins are required for granulopoiesis independent of their induction of the granulocyte colony-stimulating factor receptor. *Blood* **99**, 2776–2785 (2002).

33. Agrawal, S. *et al.* The C/EBPδ tumor suppressor is silenced by hypermethylation in acute myeloid leukemia. *Blood* **109**, 3895–3905 (2007).

34. Dickerson, K. M. *et al.* ZNF384 Fusion Oncoproteins Drive Lineage Aberrancy in Acute LeukemiaRole of ZNF384 Fusions in Lineage Aberrant Leukemia. *Blood Cancer Discov.* **3**, OF1–OF24 (2022).

35. Martino, O. di & Welch, J. S. Retinoic Acid Receptors in Acute Myeloid Leukemia Therapy. *Cancers* **11**, 1915 (2019).

36. Toofan, P. & Wheadon, H. Role of the bone morphogenic protein pathway in developmental haemopoiesis and leukaemogenesis. *Biochem. Soc. Trans.* **44**, 1455–1463 (2016).

37. Cook, B. D., Liu, S. & Evans, T. Smad1 signaling restricts hematopoietic potential after promoting hemangioblast commitment. *Blood* **117**, 6489–97 (2011).

38. Zylbersztejn, F. *et al.* The BMP pathway: A unique tool to decode the origin and progression of leukemia. *Exp Hematol* **61**, 36–44 (2018).

39. Lefort, S. & Maguer-Satta, V. Targeting BMP signaling in the bone marrow microenvironment of myeloid leukemia. *Biochem Soc T* **48**, 411–418 (2020).

40. Yahiro, Y. *et al.* PEG10 counteracts signaling pathways of TGF-β and BMP to regulate growth, motility and invasion of SW1353 chondrosarcoma cells. *J. Bone Miner. Metab.* **37**, 441–454 (2019).

41. Akamatsu, S. *et al.* The Placental Gene PEG10 Promotes Progression of Neuroendocrine Prostate Cancer. *Cell Rep.* **12**, 922–936 (2015).

42. Tang, H. *et al.* Lysyl oxidase drives tumour progression by trapping EGF receptors at the cell surface. *Nat. Commun.* **8**, 14909 (2017).

43. Koth, M. L. *et al.* Canonical Wnt/β-catenin activity and differential epigenetic marks direct sexually dimorphic regulation of Irx3 and Irx5 in developing mouse gonads. *Dev Camb Engl* **147**, (2019).

44. Bonnard, C. *et al.* Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. *Nat. Genet.* **44**, 709–713 (2012).

45. Sobreira, D. R. *et al.* Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. *Science* **372**, 1085–1091 (2021).

46. Zhu, Z. *et al.* Iroquois Homeobox 5 Negatively Regulated by miRNA-147 Promotes the Proliferation, Metastasis, and Invasion by Oral Squamous Cell Carcinoma. *J. Biomed. Nanotechnol.* **17**, 1098–1108 (2021).

47. Zhu, Q. *et al.* IRX5 promotes colorectal cancer metastasis by negatively regulating the core components of the RHOA pathway. *Mol. Carcinog.* **58**, 2065–2076 (2019).

48. Nagel, S., Pommerenke, C., Meyer, C. & MacLeod, R. A. F. The Hematopoietic TALE-Code Shows Normal Activity of IRX1 in Myeloid Progenitors and Reveals Ectopic Expression of IRX3 and IRX5 in Acute Myeloid Leukemia. *Int. J. Mol. Sci.* **23**, 3192 (2022).

49. Olsen, S. N. *et al.* MLL::AF9 degradation induces rapid changes in transcriptional elongation and subsequent loss of an active chromatin landscape. *Mol. Cell* **82**, 1140-1155.e11 (2022).

50. Landrette, S. F. *et al.* Plag1 and Plagl2 are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. *Blood* **105**, 2900–2907 (2005).

51. Yuan, B. *et al.* PTPN3 acts as a tumor suppressor and boosts TGF-β signaling independent of its phosphatase activity. *EMBO J.* **38**, e99945 (2019).

52. Jenkins, C. R. *et al.* IGF signaling contributes to malignant transformation of hematopoietic progenitors by the MLL-AF9 oncoprotein. *Exp. Hematol.* **40**, 715-723.e6 (2012).

53. Fatrai, S. *et al.* KRASG12V Enhances Proliferation and Initiates Myelomonocytic Differentiation in Human Stem/Progenitor Cells via Intrinsic and Extrinsic Pathways*. *J. Biol. Chem.* **286**, 6061–6070 (2011).

54. Bagger, F. O., Kinalis, S. & Rapin, N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res.* **47**, D881–D885 (2019).

55. Schaefer, E. J. *et al.* BCOR and BCORL1 mutations drive epigenetic reprogramming and oncogenic signaling by unlinking PRC1.1 from target genes. *Blood Cancer Discov.* **3**, bloodcandisc.0115.2021 (2021).

56. Li, M. *et al.* Somatic mutations in the transcriptional corepressor gene BCORL1 in adult acute myelogenous leukemia. *Blood* **118**, 5914–5917 (2011).

57. Wang, D. *et al.* Developmental maturation of the hematopoietic system controlled by a Lin28b-let-7-Cbx2 axis. *Cell Reports* **39**, 110587 (2022).

58. Shima, H. *et al.* Ring1A and Ring1B inhibit expression of Glis2 to maintain murine MOZ-TIF2 AML stem cells. *Blood* **131**, 1833–1845 (2018).

59. Wolfraim, L. A. *et al.* Loss of Smad3 in Acute T-Cell Lymphoblastic Leukemia. *N. Engl. J. Med.* **351**, 552–559 (2004).

60. Le, B. V. *et al.* TGFβR-SMAD3 Signaling Induces Resistance to PARP Inhibitors in the Bone Marrow Microenvironment. *Cell Rep.* **33**, 108221 (2020).

61. Grönroos, E. *et al.* Transforming Growth Factor β Inhibits Bone Morphogenetic Protein-Induced Transcription through Novel Phosphorylated Smad1/5-Smad3 Complexes. *Mol Cell Biol* **32**, 2904–2916 (2012).

62. Vegi, N. M. *et al.* MEIS2 Is an Oncogenic Partner in AML1-ETO-Positive AML. *Cell Rep.* **16**, 498–507 (2016).

63. Oosterwijk, J. G. van *et al.* Hypoxia-induced upregulation of BMX kinase mediates therapeutic resistance in acute myeloid leukemia. *J. Clin. Investig.* **128**, 369–380 (2017).

64. Kishtagari, A., Levine, R. L. & Viny, A. D. Driver mutations in acute myeloid leukemia. *Curr Opin Hematol* **27**, 49–57 (2020).

65. Yang, X., Wong, M. P. M. & Ng, R. K. Aberrant DNA Methylation in Acute Myeloid Leukemia and Its Clinical Implications. *Int J Mol Sci* **20**, 4576 (2019).

66. Network, C. G. A. R. *et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).

67. Benard, B. A. *et al.* Clonal architecture predicts clinical outcomes and drug sensitivity in acute myeloid leukemia. *Nat. Commun.* **12**, 7244 (2021).

68. Pei, S. *et al.* Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia. *Cancer Discov.* **10**, 536–551 (2020).

69. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat. Commun.* **11**, 5327 (2020).

70. Miles, L. A. *et al.* Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* **587**, 477–482 (2020).

71. Gu, T. *et al.* The disordered N-terminal domain of DNMT3A recognizes H2AK119ub and is required for postnatal development. *Nat Genet* **54**, 625–636 (2022).

72. Koya, J. *et al.* DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. *Nat. Commun.* **7**, 10924 (2016).

73. Sykes, D. B. *et al.* Inhibition of Dihydroorotate Dehydrogenase Overcomes Differentiation Blockade in Acute Myeloid Leukemia. *Cell* **167**, 171-186.e15 (2016).

74. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

75. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

76. 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).

77. Carey, M. F., Peterson, C. L. & Smale, S. T. Chromatin Immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.* **2009**, pdb.prot5279 (2009).

78. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).

79. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

FIGURES

Figure 1

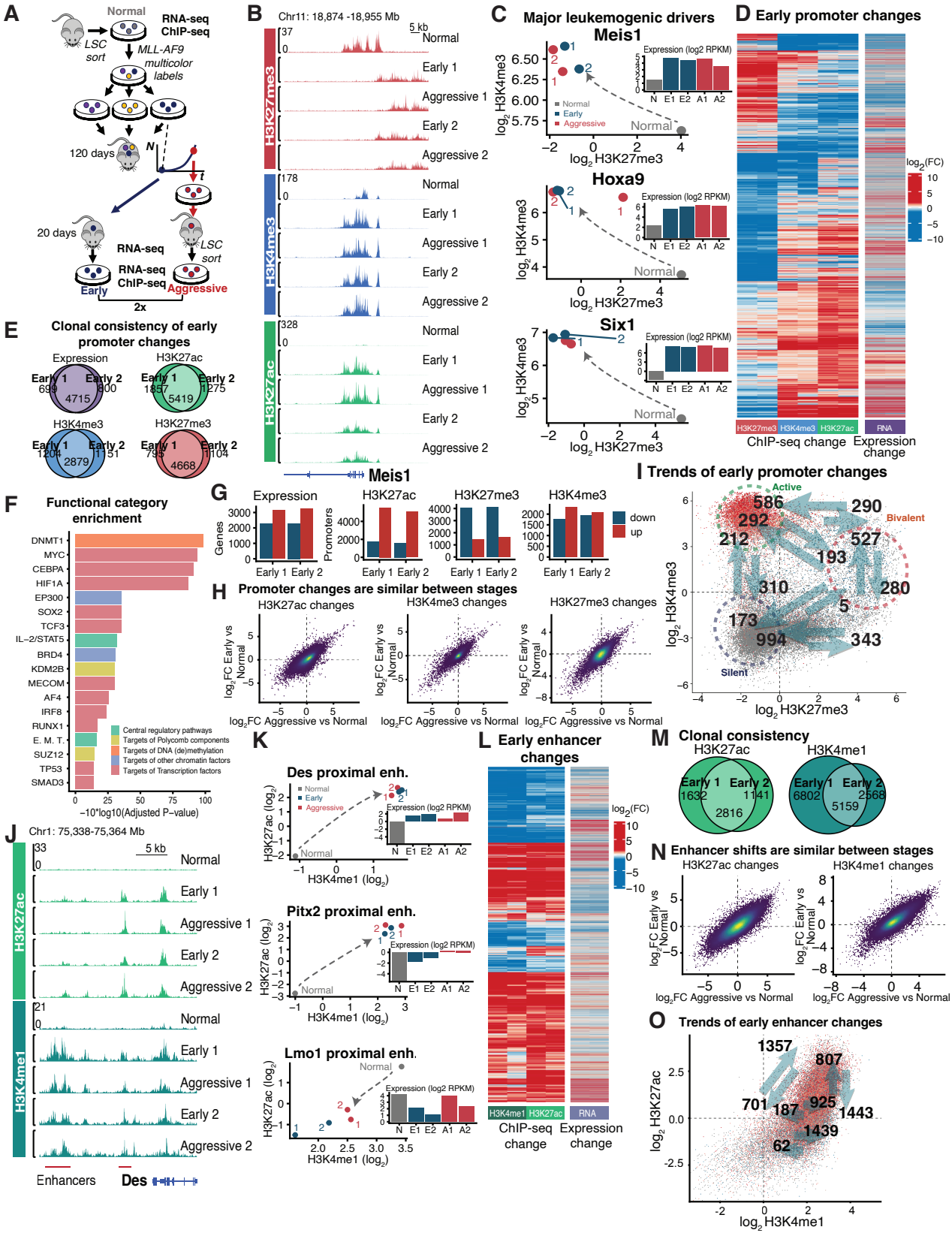


Figure 2

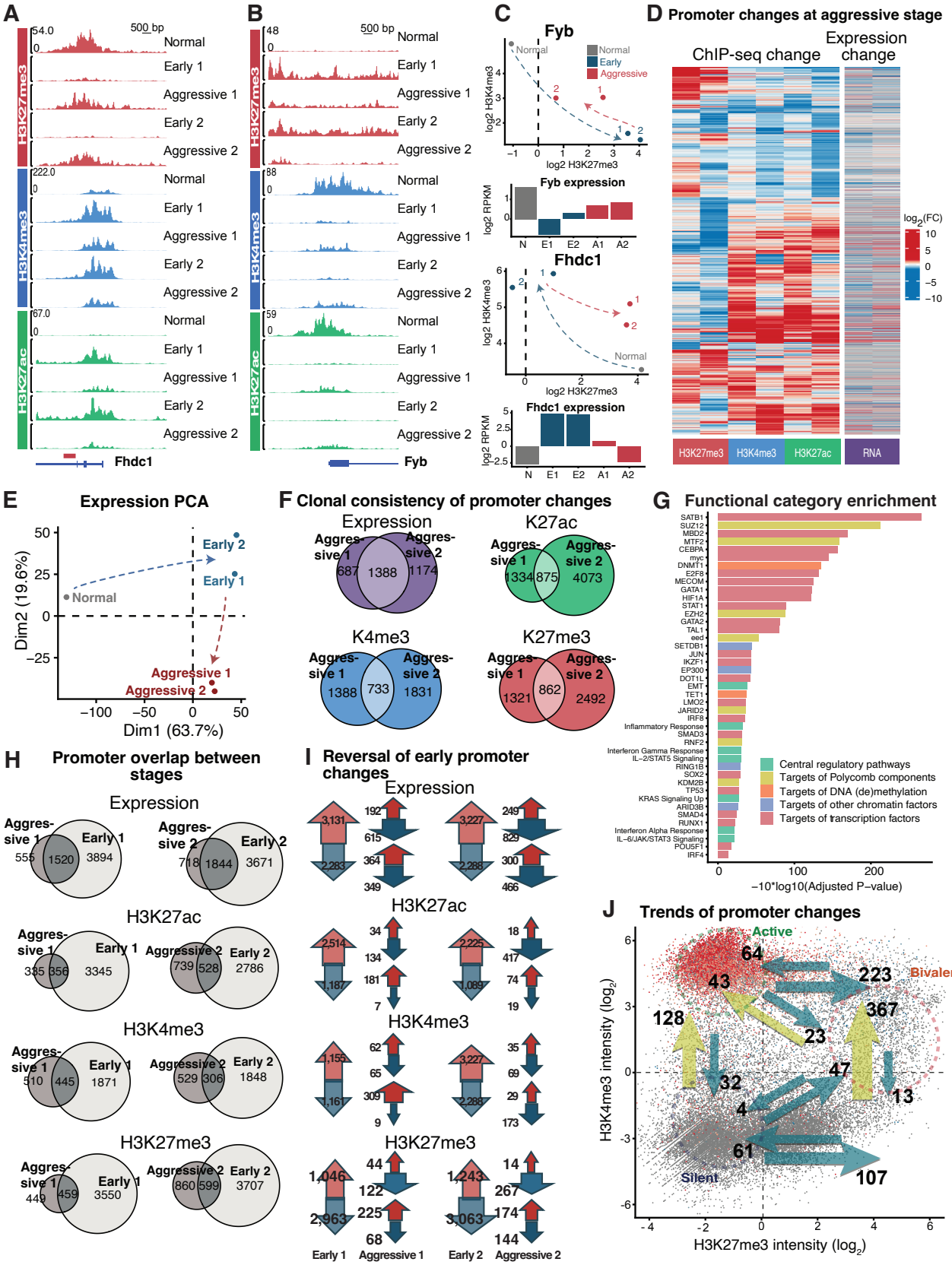

A
B
C
D Promoter changes at aggressive stage
ChIP-seq change
Expression change
H3K27me3
H3K4me3
H3K27ac
RNA
Normal
Early 1
Aggressive 1
Early 2
Aggressive 2
500 bp
Fhdc1
Fyb
Fyb expression
Fhdc1 expression
log2 H3K4me3
log2 H3K27me3
log2 RPKM
E Expression PCA
Dim1 (63.7%)
Dim2 (19.6%)
F Clonal consistency of promoter changes
Expression
K27ac
K4me3
K27me3
G Functional category enrichment
Central regulatory pathways
Targets of Polycomb components
Targets of DNA (de)methylation
Targets of other chromatin factors
Targets of transcription factors
−10*log10(Adjusted P−value)
H Promoter overlap between stages
I Reversal of early promoter changes
J Trends of promoter changes
Active
Bivalent
Silent
H3K4me3 intensity (log2)
H3K27me3 intensity (log2)

Figure 3

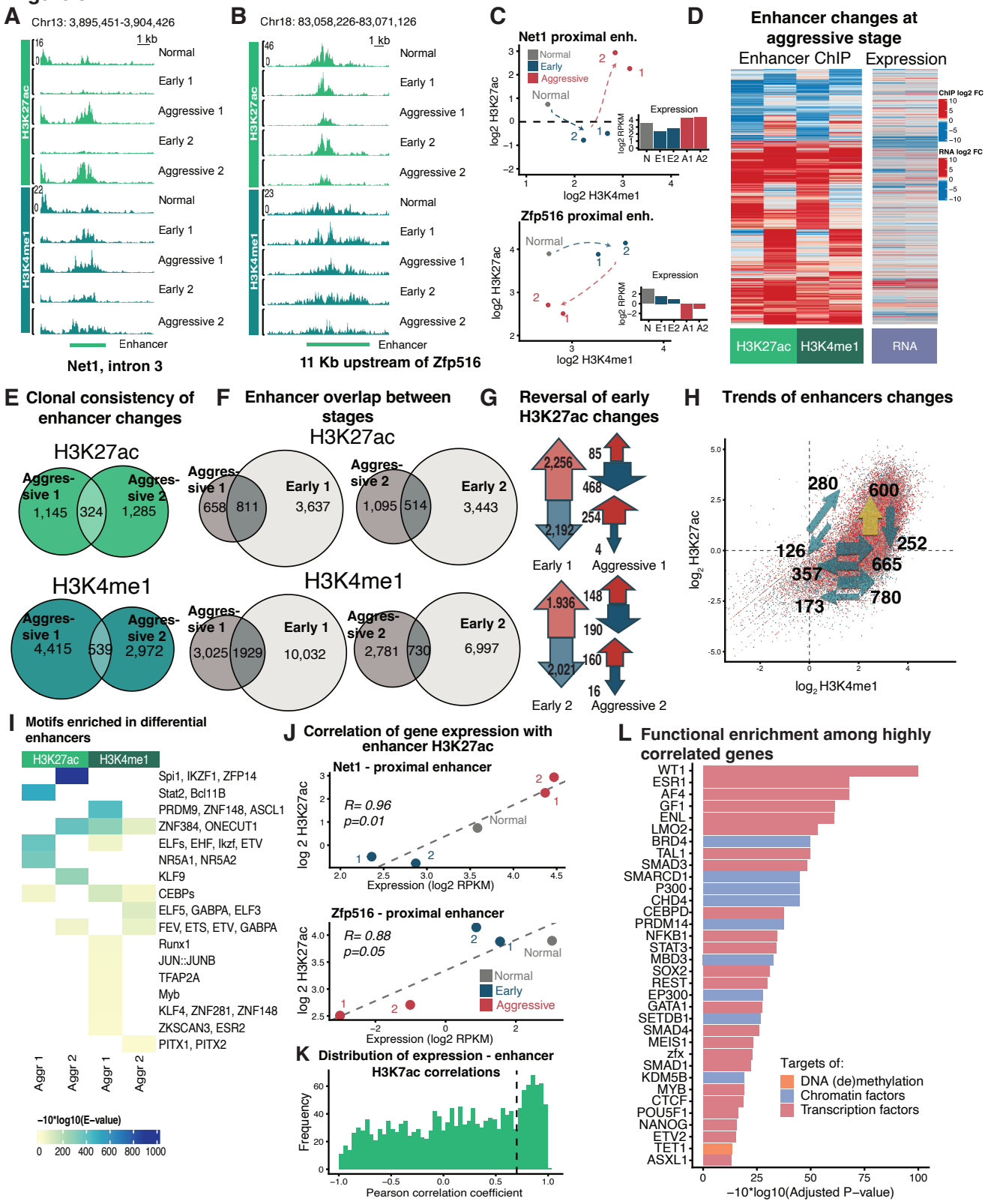

A
Chr13: 3,895,451-3,904,426
1 kb
H3K27ac
H3K4me1
Normal
Early 1
Aggressive 1
Early 2
Aggressive 2
Enhancer
Net1, intron 3
B
Chr18: 83,058,226-83,071,126
11 Kb upstream of Zfp516
C
Net1 proximal enh.
Normal
Early
Aggressive
log2 H3K27ac
log2 H3K4me1
Expression
log2 RPKM
N E1 E2 A1 A2
Zfp516 proximal enh.
D
Enhancer changes at aggressive stage
Enhancer ChIP
Expression
ChIP log2 FC
RNA log2 FC
H3K27ac
H3K4me1
RNA
E Clonal consistency of enhancer changes
H3K27ac
Aggres-sive 1
1,145
324
Aggres-sive 2
1,285
H3K4me1
Aggres-sive 1
4,415
539
Aggres-sive 2
2,972
F Enhancer overlap between stages
H3K27ac
Aggres-sive 1
658
811
Early 1
3,637
Aggres-sive 2
1,095
514
Early 2
3,443
H3K4me1
Aggres-sive 1
3,025
1929
Early 1
10,032
Aggres-sive 2
2,781
730
Early 2
6,997
G Reversal of early H3K27ac changes
2,256
85
468
254
2,192
4
Early 1
Aggressive 1
1,936
148
190
160
2,021
16
Early 2
Aggressive 2
H Trends of enhancers changes
280
600
252
126
665
357
780
173
log2 H3K27ac
log2 H3K4me1
I Motifs enriched in differential enhancers
H3K27ac
H3K4me1
Spi1, IKZF1, ZFP14
Stat2, Bcl11B
PRDM9, ZNF148, ASCL1
ZNF384, ONECUT1
ELFs, EHF, Ikzf, ETV
NR5A1, NR5A2
KLF9
CEBPs
ELF5, GABPA, ELF3
FEV, ETS, ETV, GABPA
Runx1
JUN::JUNB
TFAP2A
Myb
KLF4, ZNF281, ZNF148
ZKSCAN3, ESR2
PITX1, PITX2
Aggr 1
Aggr 2
Aggr 1
Aggr 2
−10*log10(E−value)
0 200 400 600 800 1000
J Correlation of gene expression with enhancer H3K27ac
Net1 - proximal enhancer
R= 0.96
p=0.01
Normal
log 2 H3K27ac
Expression (log2 RPKM)
Zfp516 - proximal enhancer
R= 0.88
p=0.05
Normal
Early
Aggressive
K Distribution of expression - enhancer H3K7ac correlations
Frequency
Pearson correlation coefficient
L Functional enrichment among highly correlated genes
WT1
ESR1
AF4
GF1
ENL
LMO2
BRD4
TAL1
SMAD3
SMARCD1
P300
CHD4
CEBPD
PRDM14
NFKB1
STAT3
MBD3
SOX2
REST
EP300
GATA1
SETDB1
SMAD4
MEIS1
zfx
SMAD1
KDM5B
MYB
CTCF
POU5F1
NANOG
ETV2
TET1
ASXL1
Targets of:
DNA (de)methylation
Chromatin factors
Transcription factors
−10*log10(Adjusted P−value)

Figure 4

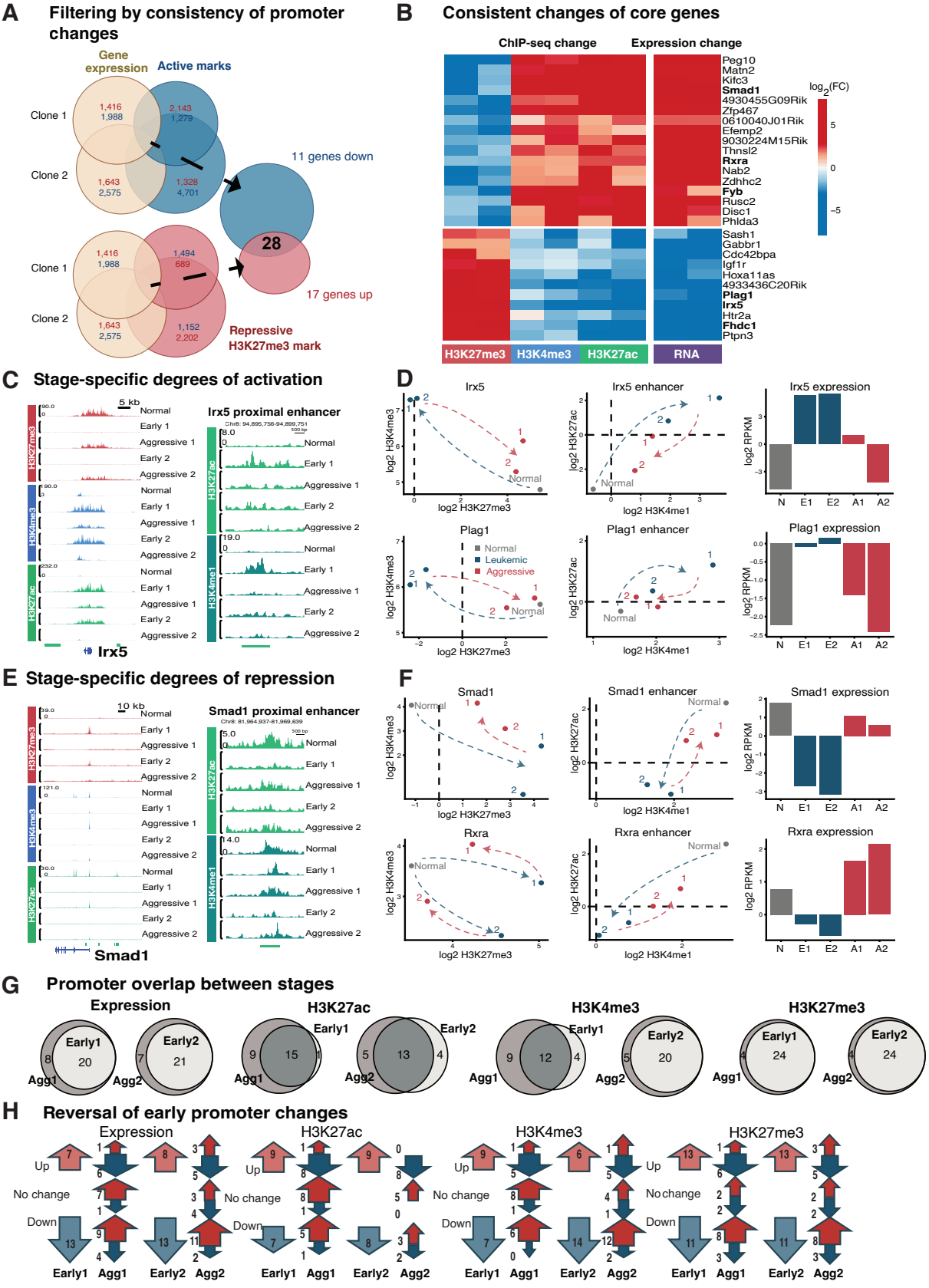
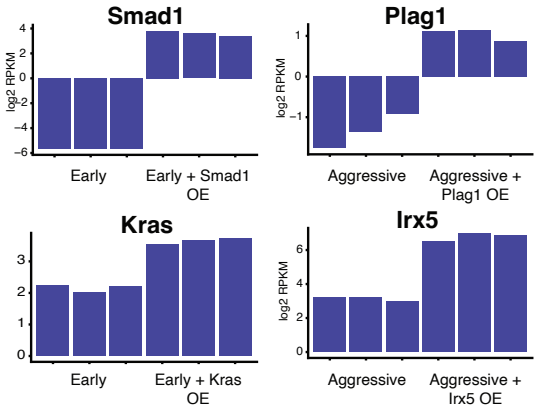
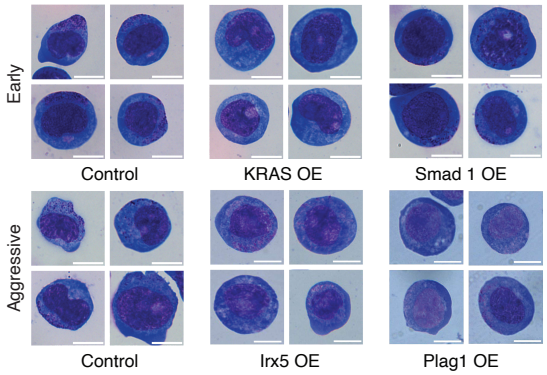

A
Filtering by consistency of promoter changes
Gene expression
Active marks
Clone 1
Clone 2
1,416
1,988
2,143
1,279
1,643
2,575
1,328
4,701
11 genes down
28
17 genes up
1,494
689
1,152
2,202
Repressive H3K27me3 mark
B
Consistent changes of core genes
ChIP-seq change
Expression change
Peg10
Matn2
Kifc3
Smad1
4930455G09Rik
Zfp467
0610040J01Rik
Efemp2
9030224M15Rik
Thnsl2
Rxra
Nab2
Zdhhc2
Fyb
Rusc2
Disc1
Phlda3
Sash1
Gabbr1
Cdc42bpa
Igf1r
Hoxa11as
4933436C20Rik
Plag1
Irx5
Htr2a
Fhdc1
Ptpn3
log2(FC)
H3K27me3
H3K4me3
H3K27ac
RNA
C
Stage-specific degrees of activation
5 kb
Irx5 proximal enhancer
Normal
Early 1
Aggressive 1
Early 2
Aggressive 2
Irx5
D
Irx5
Irx5 enhancer
Irx5 expression
Plag1
Plag1 enhancer
Plag1 expression
Normal
Leukemic
Aggressive
log2 H3K27me3
log2 H3K4me1
log2 H3K27ac
log2 H3K4me3
log2 RPKM
N
E1
E2
A1
A2
E
Stage-specific degrees of repression
10 kb
Smad1 proximal enhancer
Smad1
F
Smad1
Smad1 enhancer
Smad1 expression
Rxra
Rxra enhancer
Rxra expression
G
Promoter overlap between stages
Expression
H3K27ac
H3K4me3
H3K27me3
Early1
Early2
Agg1
Agg2
H
Reversal of early promoter changes
Up
No change
Down
Early1
Agg1
Early2
Agg2

Figure 5

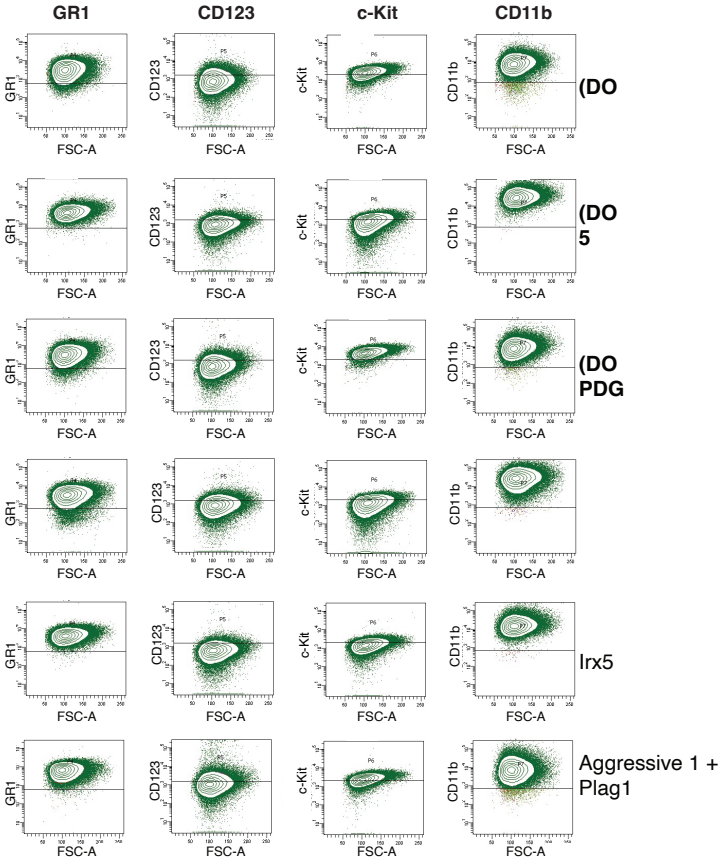
A Overexpression levels



B Cell morphology before and after overexpression of candidates



C



D

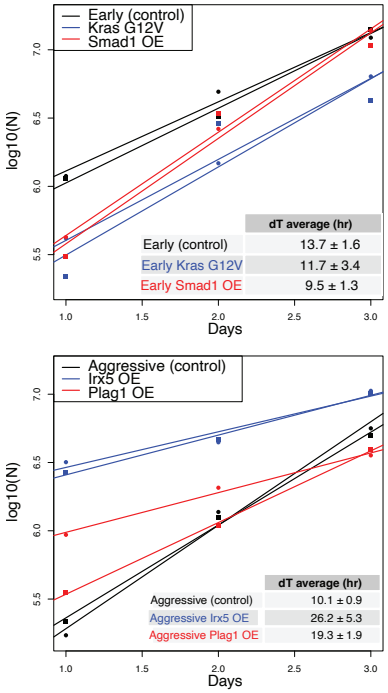
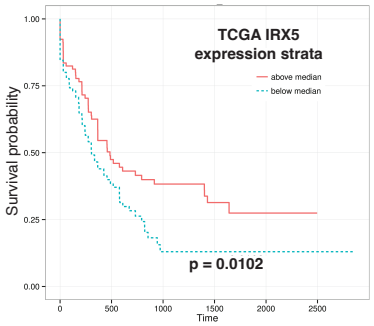
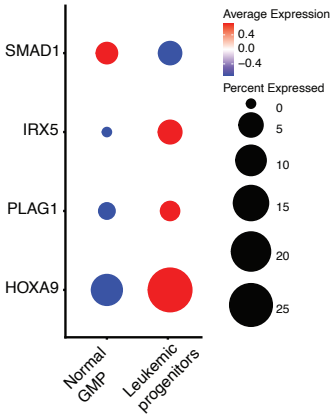


Figure 6

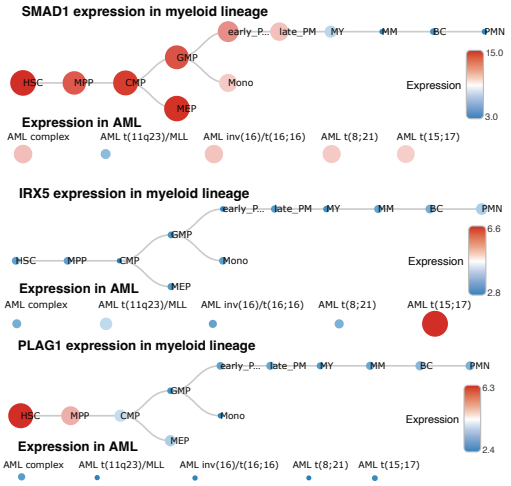
A IRX5 expression in AML is associated with patient survival



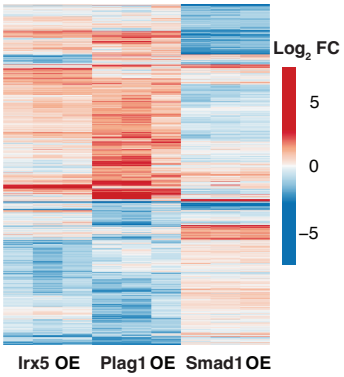
B Single-cell expression in AML



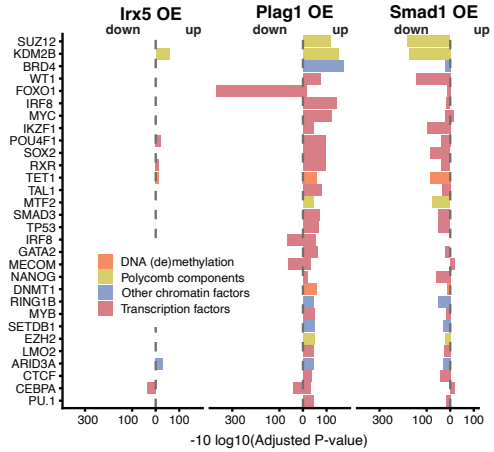
C Expression in human hematopoiesis and AML



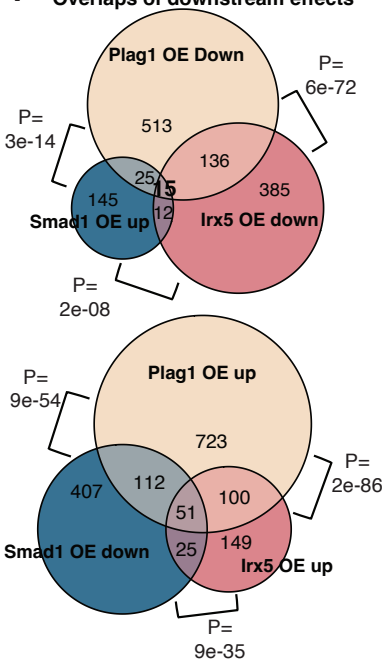
D Expression changes upon Irx5, Plag1 or Smad1 OE



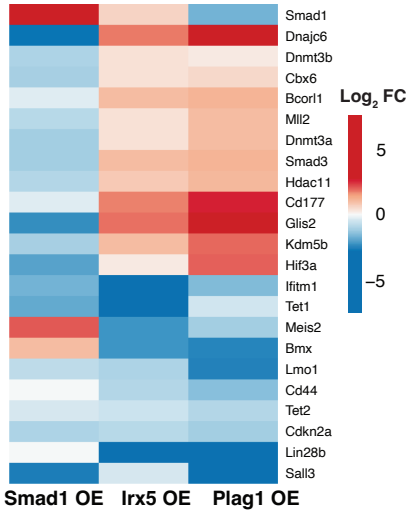
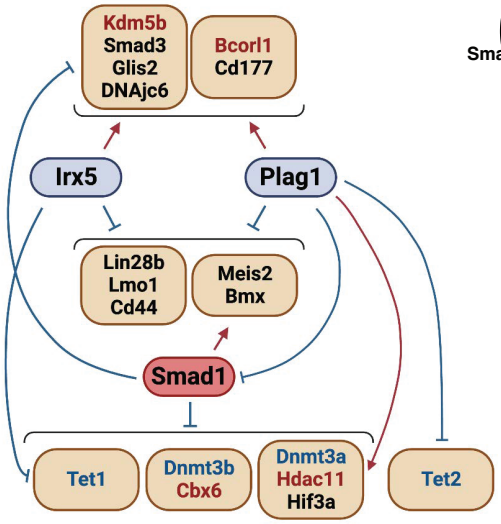
E Functional category enrichment



F Overlaps of downstream effects



G Representative shared targets



H