

## HEMATOPOIESIS AND STEM CELLS

# Disruption of a GATA2-TAL1-ERG regulatory circuit promotes erythroid transition in healthy and leukemic stem cells

Julie A. I. Thoms,<sup>1,2</sup> Peter Truong,<sup>2,3</sup> Shruthi Subramanian,<sup>2,3</sup> Kathy Knezevic,<sup>2,3</sup> Gregory Harvey,<sup>2,3</sup> Yizhou Huang,<sup>2,3,4</sup> Janith A. Seneviratne,<sup>5</sup> Daniel R. Carter,<sup>4,5</sup> Swapna Joshi,<sup>2,3</sup> Joanna Skhinas,<sup>2,3</sup> Diego Chacon,<sup>4</sup> Anushi Shah,<sup>2,3</sup> Ineke de Jong,<sup>6</sup> Dominik Beck,<sup>2,3,4</sup> Berthold Göttgens,<sup>7</sup> Jonas Larsson,<sup>6</sup> Jason W. H. Wong,<sup>8</sup> Fabio Zanini,<sup>2,3,9,\*</sup> and John E. Pimanda<sup>1,2,3,10,\*</sup>

<sup>1</sup>School of Medical Sciences, <sup>2</sup>Adult Cancer Program, and <sup>3</sup>Prince of Wales Clinical School, Faculty of Medicine, UNSW Sydney, Sydney, Australia; <sup>4</sup>School of Biomedical Engineering, University of Technology Sydney, Sydney, NSW, Australia; <sup>5</sup>Children's Cancer Institute Australia for Medical Research, Lowy Cancer Research Centre, UNSW Sydney, Kensington, NSW, Australia; <sup>6</sup>Division of Molecular Medicine and Gene Therapy, Lund Stem Cell Center, Lund University, Lund, Sweden; <sup>7</sup>Wellcome and Medical Research Council (MRC) Cambridge Stem Cell Institute, Cambridge, United Kingdom; <sup>8</sup>School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region; <sup>9</sup>Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia; and <sup>10</sup>Department of Haematology, Prince of Wales Hospital, Randwick, NSW, Australia

## KEY POINTS

- Chromatin accessibility patterns at key heptad regulatory elements can predict cell identity in healthy progenitors and leukemic cells.
- A subcircuit comprising GATA2, TAL1, and ERG regulates the stem cell to erythroid transition in both healthy and leukemic cells.

**Changes in gene regulation and expression govern orderly transitions from hematopoietic stem cells to terminally differentiated blood cell types. These transitions are disrupted during leukemic transformation, but knowledge of the gene regulatory changes underpinning this process is elusive. We hypothesized that identifying core gene regulatory networks in healthy hematopoietic and leukemic cells could provide insights into network alterations that perturb cell state transitions. A heptad of transcription factors (LYL1, TAL1, LMO2, FLI1, ERG, GATA2, and RUNX1) bind key hematopoietic genes in human CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) and have prognostic significance in acute myeloid leukemia (AML). These factors also form a densely interconnected circuit by binding combinatorially at their own, and each other's, regulatory elements. However, their mutual regulation during normal hematopoiesis and in AML cells, and how perturbation of their expression levels influences cell fate decisions remains unclear. In this study, we integrated bulk and single-cell data and found that the fully connected heptad circuit identified in healthy HSPCs persists, with only minor alterations in AML, and that chromatin accessibility at key heptad regulatory elements was predictive of cell identity in both healthy progenitors and leukemic cells. The heptad factors GATA2, TAL1, and ERG formed an integrated subcircuit that regulates stem cell-to-erythroid transition in both healthy and leukemic cells. Components of this triad could be manipulated to facilitate erythroid transition providing a proof of concept that such regulatory circuits can be harnessed to promote specific cell-type transitions and overcome dysregulated hematopoiesis.**

## Introduction

Hematopoietic stem cells (HSCs) reside in the bone marrow (BM) niche where they are mostly quiescent but retain the capacity to self-renew and replace terminal blood cell types throughout life.<sup>1</sup> Hematopoiesis is a hierarchical process, with HSCs at the apex giving rise to a range of progenitor cells with increasing lineage restriction.<sup>1</sup> Although single-cell transcriptomic data suggest a continuous differentiation process,<sup>2-7</sup> relatively pure progenitor populations corresponding to intermediate differentiation stages can be prospectively isolated based on surface marker expression.<sup>3</sup> Cell type transitions are controlled by intrinsic and extrinsic cellular factors, and loss of control can lead to inappropriate proliferation and leukemic transformation.<sup>8-13</sup>

Acute myeloid leukemia (AML) is characterized by an abundance of relatively undifferentiated cells (blasts) of myeloid lineage.<sup>14</sup> AMLs most likely originate in the earliest HSC compartments or acquire stem cell-like transcriptional programs during leukemic transformation.<sup>15-19</sup> Although blast cells can comprise the bulk of the AML population, self-renewal is restricted to a smaller population of leukemic stem cells (LSCs) that can recapitulate the disease after ablation of the blast population.<sup>20-22</sup> LSCs drive relapse,<sup>23</sup> possibly because they possess stem cell transcriptional programs.<sup>24,25</sup> Thus, AML induces a parallel hierarchy of malignant cell types with LSCs at the top.<sup>26</sup> Therapies that induce LSC differentiation by targeting mutant proteins that block differentiation are effective but limited to a minority of AMLs.<sup>27-31</sup>

AML is a heterogeneous disease with numerous driver mutations,<sup>14,32-34</sup> many of which converge on corruption of the transcriptional networks that control normal hematopoiesis.<sup>13,35-37</sup> Transcriptional networks coordinate gene regulation and play a key role in establishing and maintaining cell identity throughout the life of an organism.<sup>12,38</sup> Such networks are cell type specific and therefore have to be rewired during embryonic development and differentiation, whereas disruption can lead to oncogenic transformation.<sup>8-13</sup> Indeed, transcriptional networks are altered across AMLs with a wide spectrum of mutational origins, such that AML cells assume a new epigenetic identity distinct from any type of normal blood cell.<sup>35</sup> Furthermore, epigenetic rewiring is increasingly recognized as a nongenetic cause of treatment resistance.<sup>39-41</sup> However, the specific molecular mechanisms underlying disruption of transcriptional networks in AML and whether these can be therapeutically targeted remain unknown.

We and others have previously described 7 transcriptional regulators (heptad; LYL1, TAL1, LMO2, FLI1, ERG, GATA2, and RUNX1) that bind to key hematopoietic genes in normal human CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) and in AML.<sup>42-44</sup> Heptad factors also bind combinatorially at their own, and each other's, regulatory elements, forming a densely interconnected circuit that plays a role in maintaining the stem cell state.<sup>42,44</sup> The heptad circuit appears to be established at the hemogenic endothelium stage of blood development,<sup>45</sup> and overexpression of all 7 factors in a mouse *in vitro* differentiation system leads to increased production of pre-HSPCs with capacity for multilineage differentiation.<sup>46</sup> All 7 factors are key hematopoietic regulators, and mutation or dysregulation is commonly associated with hematological or other malignancies.<sup>32,47-50</sup> Furthermore, the heptad circuit is maintained or reactivated in AML,<sup>43,51-53</sup> and heptad expression is predictive of patient outcome.<sup>43</sup> However, heptad circuitry and function have primarily been established using bulk chromatin immunoprecipitation-sequencing (ChIPseq) experiments in heterogeneous cell populations (ie, HSPCs), which may obscure underlying subcircuits or relationships that exist only in specific cell types and cellular contexts. Thus, key questions remain about the precise roles of the heptad throughout normal and leukemic hematopoiesis, including whether all 7 factors act together in single cells and whether heptad TFs contribute to cell fate decisions and maintain stemness.

We integrated bulk and single-cell data in normal human HSPCs and leukemic cells and find that chromatin conformation at key heptad regulatory elements is predictive of cell identity in normal and leukemic progenitors. The interconnected heptad circuit identified in normal HSPCs persists in AML, but single-cell transcriptomics suggest that there are specific heptad subcircuits in individual cells that play a key role in determining differentiation trajectories as cells exit the stem cell state.

## Methods

The supplemental Methods (available on the *Blood* Web site) detail the standard techniques.

### NGS data generation and processing

ChIP was performed as described<sup>43</sup> (antibodies in supplemental Table 1). Library construction/sequencing was performed by BGI

Genomics (China) or Novogene (Hong Kong). Single-cell RNA sequencing (scRNAseq) used the 10X Genomics pipeline. Aligned sequencing data were displayed in BigWig format, and read counts covering enhancers (supplemental Table 2) were extracted using deepTools pyBigWig<sup>54</sup> and plotted.

Replicate assay for transposase-accessible chromatin-sequencing (ATACseq) counts were added. Profiles were encoded as unit vectors by dividing by total counts across all heptad peaks. City block distances on the multidimensional unit sphere between each sample and each average profile were used to compute the heat map and predict cell types.

### scRNAseq analysis

Analyses for Figures 1 and 4 are at [https://github.com/ionosofabio/heptad\\_paper](https://github.com/ionosofabio/heptad_paper). Healthy hematopoietic cell data were downloaded as described <https://github.com/dpeerlab/Palantir/blob/master/readme.md>, Rep1. Embedding coordinates, colors, cluster metadata, and smoothed counts data were extracted from the h5ad file and plotted using singlet (<https://github.com/ionosofabio/singlet>).

Count and metadata tables from Cell Ranger (10X Genomics) were converted to loom format (<http://loompy.org/>) and normalized to "counts per 10 000 (uniquely mapped) reads." The symmetric correlation matrix was ordered by hierarchical (average linkage) clustering on L2 distance with optimal leaf ordering. Conditional distributions of gene expression were computed via quantiles followed by kernel density estimate in logarithmic space.

Palantir data were subsampled to 40 cells per type. Northstar's subsample method<sup>55</sup> was used to infer cell states within ME-1 guided by Palantir data.<sup>6</sup> For graph construction, 10 external (non-mutual) neighbors were allowed to compensate for the fact that ME-1 cells are distant from actual hematopoietic cells. RNA velocity<sup>56</sup> was computed using scVelo<sup>57</sup> and projected onto Northstar's embedding. Gene expression was plotted in the same embedding after iterative nearest-neighbor smoothing. For predicting the ME-1 cell state, we trained a random forest classifier using scikit-learn and evaluated its performance via train/test splits.

## Results

### Heptad expression during hematopoiesis

To understand heptad expression patterns during hematopoiesis, we interrogated existing scRNAseq data (Palantir) from BM cells<sup>6</sup> (Figure 1A). Diverging patterns of heptad transcription factor (TF) expression were observed across developmental time (Figure 1B). All 7 TFs were expressed in HSCs, with increasing divergence during differentiation. For example, *GATA2*, *TAL1*, *LYL1*, and *LMO2* are upregulated along the erythroid lineage, whereas *RUNX1* is upregulated along the granulocytic/monocytic lineage.

### Heptad regulatory region accessibility during normal hematopoiesis

Heptad TFs form a densely interconnected circuit in bulk CD34<sup>+</sup> HSPCs, with each corresponding gene having regulatory regions bound by most of the heptad.<sup>42</sup> Because heptad expression patterns are heterogeneous in single cells, we asked whether there is evidence for changes in heptad regulation at any of the combinatorially bound regions over developmental time. Although

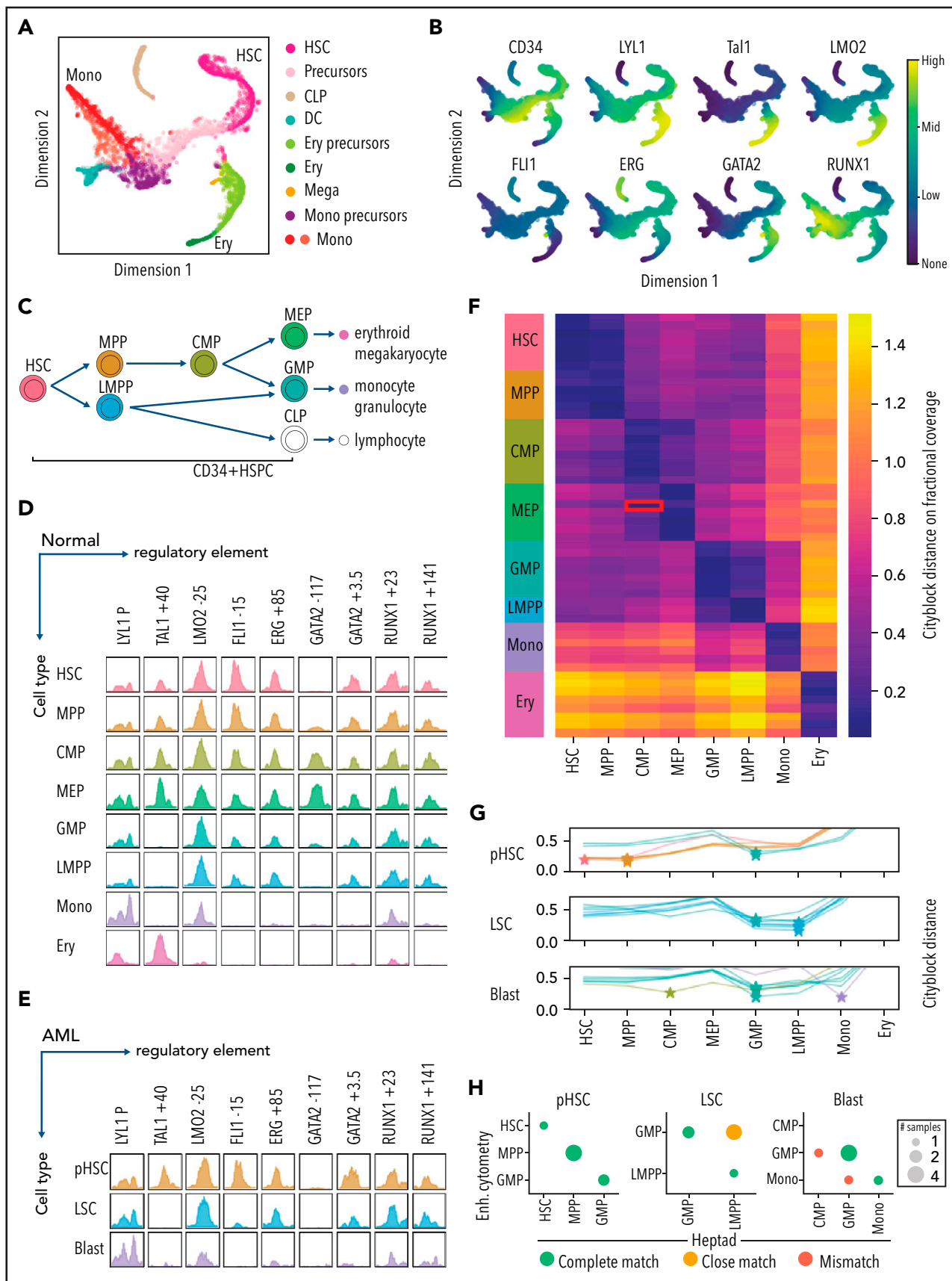


Figure 1.

hematopoiesis is a continuum (Figure 1A), functionally defined subpopulations representing various waypoints can be isolated based on cell surface marker expression (Figure 1C). We queried chromatin accessibility data from sorted BM subpopulations,<sup>4</sup> focusing on known heptad gene regulatory regions (*LYL1* promoter, *TAL1+40*, *LMO2-25*, *FLI1-16*, *ERG+85*, *GATA2+3.5*, and *RUNX1+23*<sup>42</sup>). We included 2 putative regulatory regions: *RUNX1+141*, an intragenic *RUNX1* region that was heptad-bound in HSPCs,<sup>42</sup> and *GATA2-117*, a distal regulatory element for *GATA2* that is dysregulated by translocation in the inv(3) AML subtype.<sup>58,59</sup> Strikingly, accessibility patterns differed throughout development, with some elements (*FLI1-15*, *ERG+85*, *GATA2+3.5*, and *RUNX1+141*) losing accessibility upon exiting the CD34<sup>+</sup> progenitor stage, suggesting that heptad connectivity is lost once cells commit to terminal differentiation (Figure 1D). Individual heptad regulatory elements remain accessible in more differentiated cells (*LYL1P*, *LMO2-25*, and *RUNX1+23* monocyte lineage, and *LYL1P* and *TAL1+40*; erythroid lineage) consistent with expression of the related TF in these cells, with some exceptions, such as the *LMO2-25* enhancer, which is inaccessible in erythroid cells, even though *LMO2* is highly expressed, presumably controlled by alternate regulatory regions. The *TAL1+40* and *GATA2-117* elements had the most restricted accessibility patterns with both biased toward the erythroid lineage in line with higher expression of *TAL1* and *GATA2* in these cells.

### Heptad regulatory region accessibility in AML

The heptad circuit can be active in AML,<sup>43,51-53</sup> and heptad expression can predict survival.<sup>43</sup> Data from 2 cohorts of AML cells showed that heptad regulatory regions were accessible in AMLs with diverse molecular lesions<sup>35</sup> (supplemental Figure 1A) and in preleukemic HSCs, LSCs, and leukemic blasts isolated from the same patient<sup>4</sup> (Figures 1E; supplemental Figure 1B). Notably, the *TAL1+40* enhancer was rarely accessible in AML, and the *GATA2-117* enhancer varied between patient samples.

### Heptad regulatory region accessibility can classify normal and leukemic cells

Genome-wide chromatin accessibility profiles reflect cell identity.<sup>4</sup> Because heptad expression and regulatory region accessibility are heterogenous throughout development, we asked whether the pattern of chromatin accessibility at heptad regulatory regions is sufficient to predict cell type. Using a classifier based on 9 regulatory regions, we correctly identified normal cells across the hematopoietic spectrum (Figure 1F). Furthermore, this classifier could assign a “closest normal” type to AML samples sorted into preleukemic HSC (pHSC), LSC, and blast populations (Figure 1G). Consistent with known AML biology, pHSCs were predominantly classified as HSCs or multipotent progenitors (MPPs), LSCs as lymphoid-primed MPPs (LMPPs) or granulocyte-macrophage

progenitors (GMPs), and blasts as more variable cell types. We compared our cell-type assignments to published classifications of these samples based on whole-genome accessibility patterns<sup>4</sup> and found a high concordance in pHSC and LSC populations (Figure 1H; supplemental Figure 1C). Consistent with lost heptad connectivity in more differentiated cells, the heptad-based classifier had reduced concordance with genome-wide classification in blast populations. Overall, our analysis indicates that heptad expression and accessibility are associated with cell identity in healthy hematopoietic progenitors and leukemic cells.

### The heptad network persists in AML, with altered connectivity

We extended our analysis and asked which heptad TFs were bound at each regulatory region in normal and AML contexts, looking first at heptad binding patterns at the 9 regulatory regions in CD34<sup>+</sup> HSPCs<sup>42</sup> (Figure 2A, left; supplemental Figure 2). Combinatorial binding was observed, with *LYL1*, *FLI1*, *GATA2*, and *RUNX1* bound at all regions, and *FLI1*, *ERG*, *GATA2*, and *RUNX1* each having at least 1 regulatory element bound by all 7 TFs. Binding patterns were then used to infer the connectivity map of heptad autoregulation in HSPCs (Figure 2A, right).

We next compared heptad connectivity in 2 AML cell lines: ME-1, and KG-1. AML cell lines recapitulate properties of primary AML cells<sup>60</sup> and can be experimentally manipulated. ME-1 and KG-1 cells express all 7 heptad genes, although the pattern of individual TF expression varies both between cell lines and compared with HSPCs (supplemental Figure 3). Consistent with primary AML accessibility, heptad ChIPseq in ME-1 (Figures 2B; supplemental Figure 4) and KG-1 (Figure 2C; supplemental Figure 5) revealed that the densely interconnected circuit observed in HSPCs persists in AML cells, although the precise pattern of connectivity varies. For example, both ME-1 and KG-1 have prominent binding peaks at *LYL1P*, whereas at *TAL1+40*, ME-1 and KG-1 had fewer called peaks (4 of 7 and 2 of 7, respectively) than HSPCs (5 of 7), and these were generally small. Overall, heptad TFs remain highly connected in both AML cell lines, albeit with somewhat different circuit structures compared with HSPCs. Expression levels of individual TFs in HSPCs and AML cell lines were broadly in keeping with the number and binding intensities of TFs at the cognate regulatory element (Figure 2; supplemental Figure 3), except for *LMO2*, which had similar numbers and sizes of ChIPseq peaks across all cell types but was highly expressed in HSPCs.

### Heptad regulatory elements must contain ETS and GATA motifs

Having shown that heptad binding at regulatory regions persists in AML, we wanted to understand the role of specific TF binding motifs within these regulatory regions. *Cis*-regulatory elements integrate signals from multiple TFs that bind to specific DNA

**Figure 1. Heptad regulatory regions have dynamic accessibility profiles across normal and leukemic blood development, and accessibility patterns are sufficient to classify normal and leukemic cells.** (A) tSNE plot of scRNAseq in normal BM, with cells labeled by inferred identity as determined by Setty et al.<sup>6</sup> CLP, common lymphoid progenitor; DC, dendritic cell (B) Relative expression of CD34 and heptad genes projected on to the tSNE plot in panel A. (C) The branching hierarchy model of normal blood development showing relationships between the cell populations shown in panel D. (D) ATACseq peaks at heptad regulatory regions over developmental time. Plots show merged data from available replicates. (E) ATACseq peaks at heptad regulatory regions in 1 representative patient with AML, showing pHSCs, LSCs, and leukemic blasts (Blast). (F) Classification of normal cell types using only ATACseq signal at heptad regulatory regions. Heat map shows calculated distance between each sample and the training set. The red box indicates a single MEP replicate that was misclassified as a CMP. (G) Classification of AML nearest normal cell type using only ATACseq signal at heptad regulatory regions. Plots show distance from each normal cell type for preleukemic HSCs, LSCs, and leukemic blasts from 7 patients with AML. (H) Performance of the heptad regulatory region classifier compared with previous classification of these samples using genome wide enhancer (Enh.) cytometry. Panels D and H adapted from Corces et al<sup>4</sup> with permission. tSNE, t-distributed stochastic neighbor embedding.

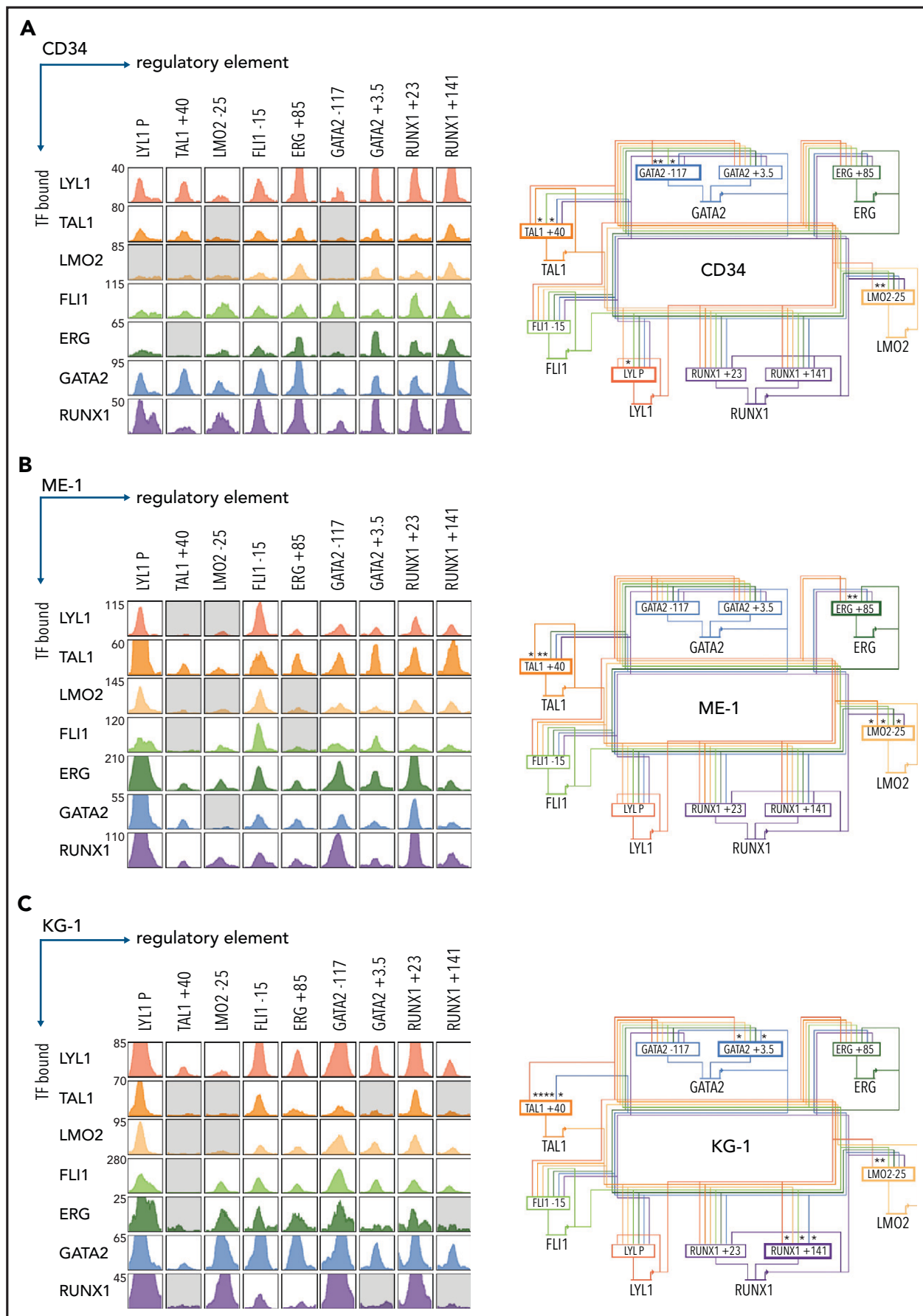


Figure 2.

sequences, with direct binding occurring at consensus binding motifs. The heptad TFs belong to 4 broad classes of TFs with different consensus binding motifs, E-box (CANNTG, bound directly by LYL1 and TAL1 and indirectly by LMO2), ETS (GGAW, bound by FLI1 and ERG), GATA (bound by GATA2), and RUNX (TGYGGT, bound by RUNX1). To identify consensus motifs that are likely to correspond to TF binding sites, we performed multiple sequence alignments using human, mouse, dog, and opossum genomes (Figure 3A). All regulatory elements contained conserved ETS and GATA motifs, whereas 7 of 9 contained a conserved E-Box motif and 6 of 9 a conserved RUNX motif. We mutated all conserved instances of each binding motif class (supplemental Table 4) and tested in luciferase reporter constructs in KG-1 and ME-1 cells.

Deletion of ETS consensus motifs was universally deleterious, leading to significant loss of activity for all elements tested (Figure 3B). Deletion of GATA consensus motifs had a significant negative impact for all regions in at least 1 cell line. Deletion of E-box or RUNX motifs reduced luciferase reporter activity; however, the effect was generally small compared with deletion of ETS or GATA motifs, and in 1 case (*LMO2-25*) deletion of the RUNX motif led to slightly increased activity. Overall, regulatory region activity was impaired by loss of any class of TF binding motif, with loss of ETS or GATA motifs dominating. Two WT reporter constructs, *TAL1+40* and *RUNX1+141*, showed minimal activity in 1 or both cell lines (Figure 3C) and were excluded from the mutation analysis. Consistent with its activity, *TAL1+40* had few heptad TF binding inputs in either cell line, and *RUNX1+141*, which was active in ME-1 but not KG-1, had fewer inputs in KG-1 than in ME-1.

## Single-cell transcriptomics reveal key regulators of the HSC–erythroid transition

Altered enhancer activity is read out as gene expression changes. Encouraged by our results indicating that removing specific consensus motifs altered activity of heptad regulatory regions, we proceeded to scRNAseq analysis of heptad expression in ME-1 cells that are amenable to downstream perturbation. We quantified heptad heterogeneity and observed that, for both high (eg, *LYL1*)– and low (eg, *ERG*)–expression genes, heterogeneity across the ME-1 population spanned an order of magnitude (Figure 4A). Furthermore, the highest gene expression (*LYL1*) corresponded to the highest heptad binding at an associated regulatory region, whereas lower gene expression (*TAL1* and *GATA2*) corresponded to lower heptad binding at their associated regulatory regions (Figure 2B).

We next looked for pairwise expression correlations between TFs and found that *GATA2* correlated positively with *TAL1* and negatively with *ERG* and *LMO2* (Figure 4B). Because correlation

measures are insensitive to extreme phenotypes, we performed complementary analysis to evaluate whether this effect is also seen at the extreme of the distribution and plotted conditional gene expression distribution in the bottom and top quantiles of expressors of *GATA2* (Figure 4C). Given the observed heterogeneity in heptad expression in ME-1 cells and the strong association between heptad regulation and cell type, we asked whether we could identify subpopulations within the ME-1 scRNAseq data. A canonical, unsupervised, clustering approach based on overdispersed features did not result in distinct biological patterns beyond the cell cycle, as expected from a cell line. We reasoned that a more sophisticated feature selection, together with soft guidance from healthy marrow data would reveal additional hidden heterogeneity. We therefore switched from unsupervised clustering to Northstar, a semisupervised clustering algorithm that leverages information from training data to channel the axes of heterogeneity during feature selection, graph construction, and cell community detection.<sup>55</sup> Using healthy marrow transcriptomes<sup>6</sup> (Figure 1A) as training data, this analysis revealed 2 major subpopulations, HSC-like (pink) and mono-precursor-like (purple, 1136 and 277 of 1489 cells, respectively) plus a minor population that was more similar to ery-precursor cells (lime, 47 of 1489 cells) and 2 small groups of cells resembling megakaryocytes (18 cells) and monocytes (Figure 4D; 11 cells). RNA velocity analysis<sup>56</sup> (Figure 4D, arrows) revealed a major trajectory along the HSC-mono-precursor axis, and an alternate trajectory connecting the HSCs to the ery-precursor population. This flow diagram (independent of Northstar clustering) confirmed population structure reminiscent of healthy hematopoiesis (Figure 4D, inset). Primary AML cells also have population structures resembling normal hematopoiesis<sup>61</sup> and have differential heptad expression between subpopulations (supplemental Figure 6A). We projected expression levels of the 4 previously identified genes on embedded cell plots (Figure 4E) and, consistent with our correlation data and known biological functions, *GATA2* and *TAL1* expression were enriched in the ery-precursor population. Conversely, *ERG* and *LMO2* expression were enriched in the HSC-like and mono-precursor-like populations. We then computed the fold expression change in heptad genes between HSC and ery-precursor cells in both ME-1 and normal BM cells (Figure 4F; supplemental Figure 6B–C; supplemental Tables 5 and 6). In ME-1 cells, *ERG* expression was reduced (0.6 times) and *GATA2* and *TAL1* expression increased (11 and 3.5 times, respectively) in ery-precursor cells (Figure 4F, left). We observed a similar pattern in healthy cells, although *FLI1*, *RUNX1*, and *LMO2* also showed expression changes in this context (Figure 4F, right).

To better understand how heptad TFs influence cell-specific gene expression we interrogated TF binding in bulk HSPCs. As these cells are a mixture of progenitor types, we focused on ATACseq peaks uniquely accessible in HSCs or megakaryocyte erythrocyte

**Figure 2. A densely interconnected heptad autoregulatory circuit persists in AML cells with altered connectivity compared with CD34<sup>+</sup> HSPCs.** (A) ChIPseq binding pattern at heptad regulatory regions in CD34<sup>+</sup> HSPCs (left). Gray boxes indicate regulatory regions not computationally called as binding peaks for the indicated TF. Plots are scaled to 5 times the height of the smallest called peak for that TF to allow visualization of a wide range of peak heights. Corresponding inferred heptad autoregulatory circuit (right). Most regulatory elements have all 7 heptad TFs bound; asterisk and bold border indicate regions where binding of a particular TF is absent. (B) ChIPseq binding pattern at heptad regulatory regions in ME-1 AML cells (left). Gray boxes indicate regulatory regions not computationally called as binding peaks for the indicated TF. Plots are scaled to 5 times the height of the smallest called peak for that TF to allow visualization of a wide range of peak heights. Corresponding inferred heptad autoregulatory circuit (right). Most regulatory elements have all 7 heptad TFs bound; asterisk and bold border indicate regions where binding of a particular TF is absent. (C) ChIPseq binding pattern at heptad regulatory regions in KG-1 AML cells (left). Gray boxes indicate regulatory regions not computationally called as binding peaks for the indicated TF. Plots are scaled to 5 times the height of the smallest called peak for that TF to allow visualization of a wide range of peak heights. Corresponding inferred heptad autoregulatory circuit (right). Most regulatory elements have all 7 heptad TFs bound; asterisk and bold border indicate regions where binding of a particular TF is absent.

progenitor (MEPs; supplemental Figure 7; supplemental Table 7). *ERG*, *FLI1*, and *RUNX1* had higher expression in HSCs than in ery-precursors and showed higher average binding at HSC-unique peaks, whereas *GATA2*, *TAL1*, and *LYL1* were more highly expressed in ery-precursors but had similar average binding at both MEP- and HSC-unique peaks (supplemental Figure 7). *LMO2* had higher expression in ery-precursors, but higher binding at HSC-unique peaks. TFs bind DNA directly via their cognate binding motifs, or indirectly via protein-protein interactions. HSC-unique peaks were highly enriched for ETS motifs (supplemental Table 8, significance value [sv] 5.50E-171), and enriched for RUNX motifs (supplemental Table 8, sv 5.70E-08), consistent with higher *ERG*, *FLI1*, and *RUNX1* binding at these peaks. MEP-unique peaks were bound by *GATA2* and highly enriched for GATA motifs (supplemental Table 8, sv 3.20E-111). *GATA2* was also bound at HSC-unique peaks, whereas GATA motifs were enriched in only a minor fraction of HSC-unique peaks (supplemental Table 8; 33 of 7396, sv 3.10E-02), suggesting that *GATA2* binding at these sites may be mediated by interactions with other TFs, rather than direct DNA binding.

Finally, we asked whether heptad expression was sufficient to classify ME-1 cells as HSC-like or ery-precursor-like (Figure 4G). Using a random forest classifier based on Palantir data, we found that heptad expression correctly classified cells with high accuracy (area under the receiver operating characteristic curve = 0.80), and that *GATA2* expression was the best performing gene in terms of model accuracy (area under the receiver operating characteristic curve = 0.84).

### Direct manipulation of *GATA2* and *ERG* promotes erythroid trajectory

We then evaluated the effects of perturbing heptad factors on (1) expression of other heptad factors, (2) global transcriptome of perturbed cells, and (3) cell function. Specifically, we predicted that high levels of *GATA2* or *TAL1* and low levels of *ERG* would promote transition along the HSC-ery-precursor axis (Figure 5A). We first knocked down key heptad genes in ME-1 cells (supplemental Figure 8A) and measured the response of other heptad genes. *GATA2* knockdown led to a decrease in *TAL1* and most other heptad genes, except for *ERG*, which was unaffected by *GATA2* knockdown (Figure 5B, left). Similarly, *TAL1* knockdown led to decreased *GATA2* and most other heptad genes, except for *ERG* (Figure 5B center). Conversely, *ERG* knockdown led to

decreased *LMO2* expression, but increased expression of *GATA2*, *FLI1*, and *TAL1* (Figure 5B, right). *RUNX1* expression showed inconsistent changes, possibly because of dysregulation via translocation of its essential binding partner *CBFb* in ME-1 cells.<sup>62</sup> Similar results were observed using additional short hairpin RNAs (shRNAs) targeting *GATA2* or *ERG* (supplemental Figure 8B). Heptad gene expression also changed after knockdown of *GATA2*, *TAL1*, or *ERG* in 2 additional AML cell lines (supplemental Figure 8C-D), although response patterns varied between cell lines, most likely reflecting the unique cell subpopulations in each.

Because the bulk of ME-1 cells were assigned as HSC-like, we reasoned that *ERG* knockdown or *GATA2* overexpression, would alter their trajectory away from the HSC-like and toward the ery-precursor-like state. *ERG* knockdown reduced ME-1 colony formation in methylcellulose (supplemental Figure 8E), consistent with a shift away from the HSC-like state. We also analyzed RNA-seq data from *GATA2* overexpression in ME-1 cells<sup>63</sup> and found that increased *GATA2* led to increased *TAL1* and *RUNX1* and reduced *ERG* and *LMO2*, similar to expression changes between ery-precursor-like and HSC-like ME-1 cells (Figure 5C, left; compare with Figure 4F, left). Gene Set Enrichment Analysis (GSEA) was used to compare *GATA2*-driven changes in global gene expression to expression differences between ery-precursors and HSCs. Globally, genes that were high in ery-precursors tended to increase after *GATA2* overexpression, whereas genes that were low in ery-precursors tended to decrease (Figure 5C right). *ERG* overexpression in HSPCs promotes progenitor expansion,<sup>64</sup> and we have now shown that *ERG* expression is reduced across the HSC to ery-precursor boundary in normal BM and ME-1 cells (Figure 4F). Furthermore, an independent method using scRNAseq landscapes as references predicts that perturbing *ERG* in mouse or human LMPPs would push cells toward an erythroid fate.<sup>65</sup> We therefore asked whether *ERG* knockdown in HSPCs promoted an ery-progenitor phenotype. *ERG* knockdown led to downregulation of *FLI1*, *LYL1*, and *LMO2*, and upregulation of *GATA2* and *TAL1* (Figure 5D, left), similar to expression changes across the HSC-ery-progenitor transition in Palantir data (Figure 4F, right). GSEA was used to compare *ERG* knockdown-driven changes in global gene expression to expression differences between ery-precursors and HSCs. Globally, genes that were high in ery-precursors tended to increase after *ERG* knockdown, whereas genes that were low in ery-precursors

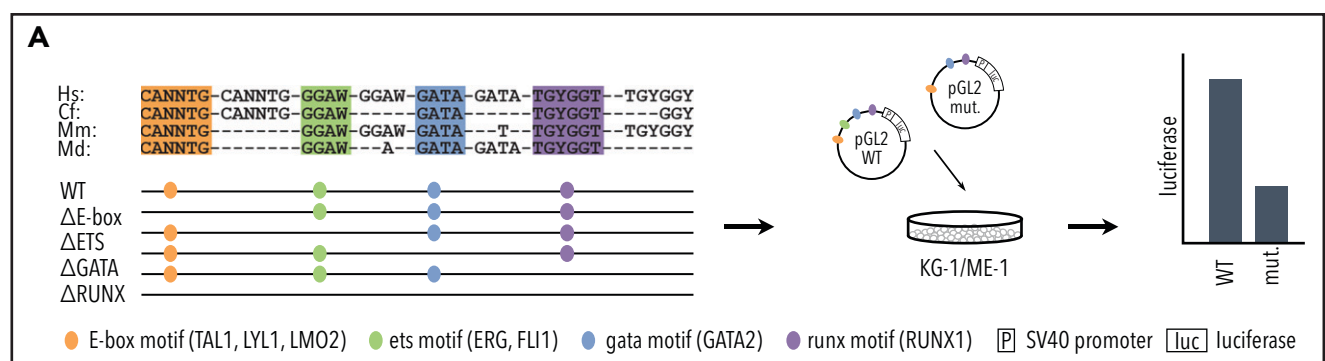
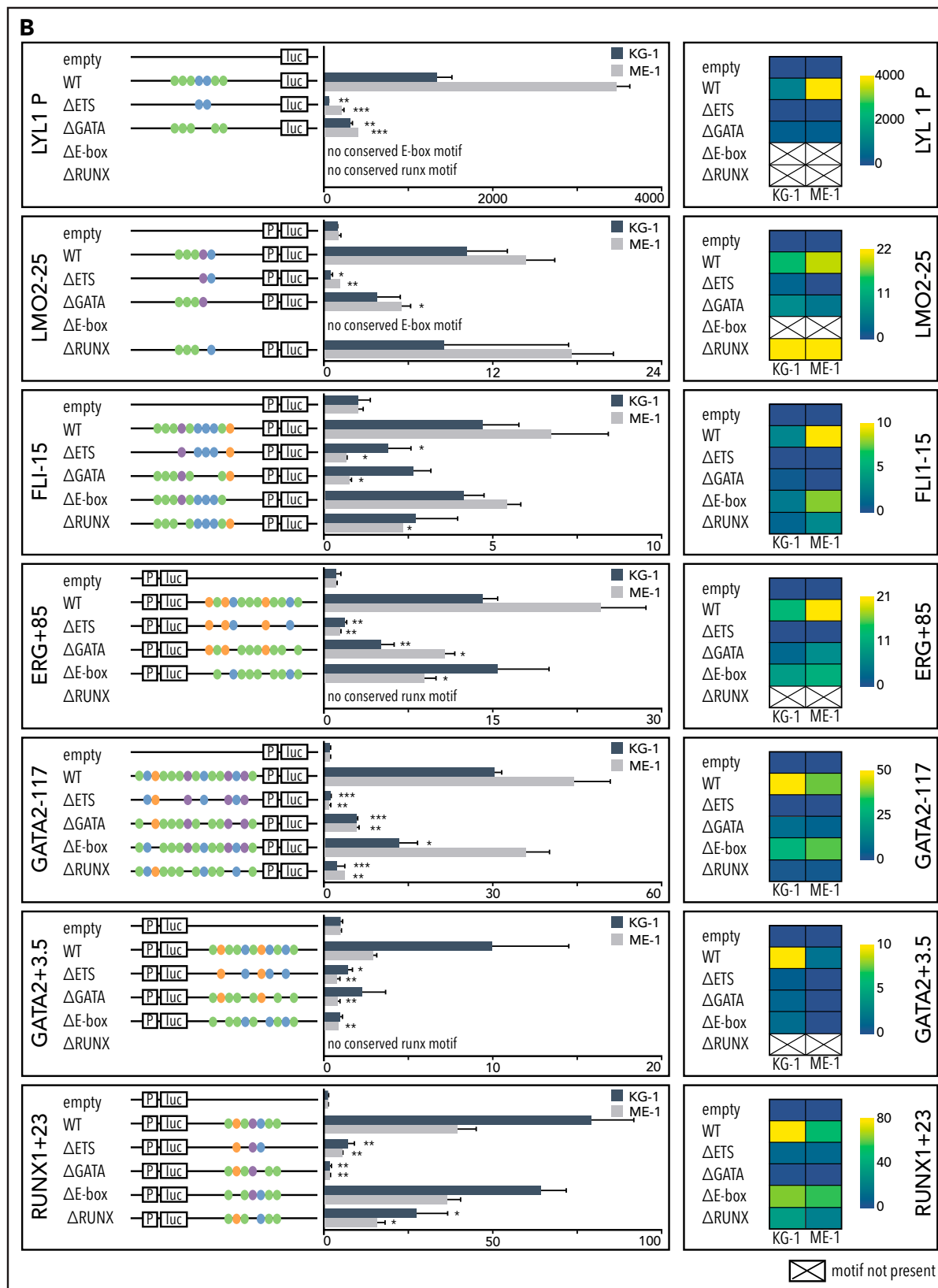
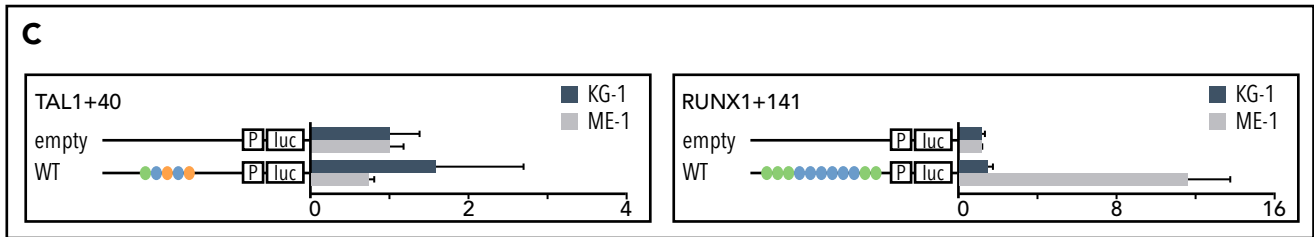


Figure 3.



**Figure 3. Specific TF consensus binding motifs, particularly ETS and eiGATA motifs, are critical for function of heptad regulatory elements.** (A) Schematic showing process for selecting TF binding motifs for mutation and luciferase reporter workflow. (B) Conserved TF binding motifs in heptad regulatory elements that were highly bound by heptad TFs in AML cell lines, and activity of wild-type (WT) and mutated luciferase constructs in KG-1 and ME-1 cells (left). Activity is scaled relative to the empty vector, and graphs show representative data from a single transfection experiment. \* $P < .05$ ; \*\* $P < .01$ ; \*\*\* $P < .001$ , t test. Heat maps showing aggregate data from all luciferase experiments (right). Data from biological replicates were normalized to WT activity for each experiment, then aggregate data scaled relative to empty vector. Heat maps are scaled from 0 to maximum luciferase activity for each regulatory element.





**Figure 3 (continued)** (C) Schematics showing conserved TF binding motifs in heptad regulatory elements that were highly bound by heptad TFs in AML cell lines, and activity of WT luciferase constructs in KG-1 and ME-1 cells. Activity is scaled relative to the empty vector, and graphs show representative data from a single transfection experiment. \* $P < .05$ ; \*\* $P < .01$ ; \*\*\* $P < .001$  (Student t test).

tended to decrease (Figure 5D, right). To evaluate functional consequences of *ERG* knockdown in HSPCs (Figure 5E) we measured colony forming capacity and found that cells with reduced *ERG* expression were skewed toward erythroid colony formation (Figure 5F). Together, the perturbation data supports the notion that heptad genes, and in particular the triplet *GATA2*, *TAL1*, and *ERG*, form a functionally relevant interconnected network and play a key role in regulating cell state transitions in healthy blood cells and in leukemic cells.

## Discussion

Gene regulatory networks control cell fate decisions in development and disease. We focused on heptad TFs and identified parallel phenotypes between healthy hematopoiesis and leukemic cells spanning single-cell gene expression, chromatin state, and enhancer use (Figure 6A). Our data suggest that *GATA2*, *TAL1*, and *ERG* constitute a heptad subcircuit that regulates stem cell-to-erythroid transition in healthy blood cells and leukemia cells (Figure 6B).

### Insights into enhancer biology

The genome-wide chromatin state can be used to classify cell types.<sup>4</sup> We showed that chromatin accessibility at only 9 heptad enhancers could be used to classify all early stages of hematopoiesis and subpopulations of AML cells. Although the transcriptional network determining hematopoietic cell fate undoubtedly contains additional enhancers, the heptad enhancers in this study give significant insight into the transcriptional control of blood cell identity. Most heptad enhancers were accessible in HSPCs and became selectively inaccessible at terminal differentiation, though exceptions were observed. We found the *GATA2-117* (mice: *Gata2-77*) enhancer was open only in common myeloid progenitors (CMPs) and MEPs, suggesting a central role for this enhancer in erythroid transition and confirming previous murine models, where its deletion blocked erythroid and megakaryocytic differentiation.<sup>66</sup>

This enhancer has been studied in *inv(3)* AML, where it is translocated close to oncogene *MECOM/EVI1*, leading to increased *EVI1* and decreased *GATA2* expression.<sup>58,59</sup> We found that the enhancer was accessible in a subset of leukemic cells and was strongly heptad-bound in both AML cell lines compared with HSPCs. In our reporter assays *GATA2-117* also drove more luciferase activity than *GATA2+3.5*, the other *GATA2* regulatory element. Thus, even in its normal genomic context *GATA2-117* may play a role in driving *GATA2*

expression in AML. Unlike *GATA2-117*, the *ERG+85* enhancer was open in all HSPC subsets and across AML subtypes (supplemental Figure 1A). This enhancer has been linked to AML prognosis<sup>43</sup> and used to identify LSCs within bulk AML populations.<sup>67,68</sup> Enhancers are replete with sequence motifs enabling binding of distinct TF families, either directly to DNA or indirectly via protein scaffolding, as observed for *LMO2*<sup>69,70</sup> and *RUNX1*.<sup>42,44</sup> In this study, evolutionarily conserved heptad enhancers relied heavily on ETS and GATA motifs, in agreement with previous reports that ETS-ETS-GATA motifs were enriched at blood enhancers.<sup>71</sup>

### Regulation of cell fate transitions by *GATA2*, *TAL1*, and *ERG*

Combinatorial binding of TFs is a key component of cell fate transitions.<sup>38</sup> We identify a triad of TFs-*GATA2*, *TAL1*, and *ERG*, whereby high *GATA2* and *TAL1*, and low *ERG* expression biased fate decisions toward the erythroid lineage in both HSPCs and ME-1 leukemic cells. A similar circuit, comprising *GATA2*, *TAL1*, and *FLI1* (an ETS TF closely related to *ERG*) has been reported during embryonic HSC specification,<sup>72</sup> whereas *GATA1*, *TAL1* and *KLF1* form a subcircuit in erythroid cells.<sup>73</sup> Indeed, recycling of regulatory modules is a key feature of developmental networks,<sup>38</sup> emphasizing the utility of cell classification strategies such as Northstar.<sup>55</sup>

Each member of this triad is known to play complex roles in healthy blood and leukemia development. *GATA2* controls blood cell emergence in the embryonic aorta<sup>74</sup> and is necessary for HSC maintenance.<sup>75</sup> Germline loss-of-function mutations in *GATA2* predisposes to myelodysplastic syndrome and AML,<sup>76</sup> and high *GATA2* expression is associated with poor prognosis in patients with AML.<sup>77</sup> *TAL1* is also necessary for embryonic blood formation<sup>48,78</sup> and drives erythroid and megakaryocytic differentiation programs<sup>79</sup> but is dispensable for HSC maintenance.<sup>48,80,81</sup> However, dysregulation of *TAL-1* is associated with T-ALL.<sup>48</sup> *ERG* is not necessary for HSC specification or differentiation, but it promotes HSC maintenance by restricting differentiation.<sup>82,83</sup> High *ERG* expression is a poor prognostic marker for AML<sup>49,84-86</sup> and is leukemogenic in mouse models,<sup>87-90</sup> although its role in human leukemia is more subtle.<sup>64</sup>

### Clinical implications

Therapeutic approaches to AML that force LSCs to differentiate have been sought.<sup>91</sup> Although TFs are relatively difficult drug targets, small molecules upregulating *CEBPA*<sup>92,93</sup> or downregulating *PU.1*<sup>94</sup> and *RUNX1*<sup>95</sup> have been developed. Regulatory circuits,

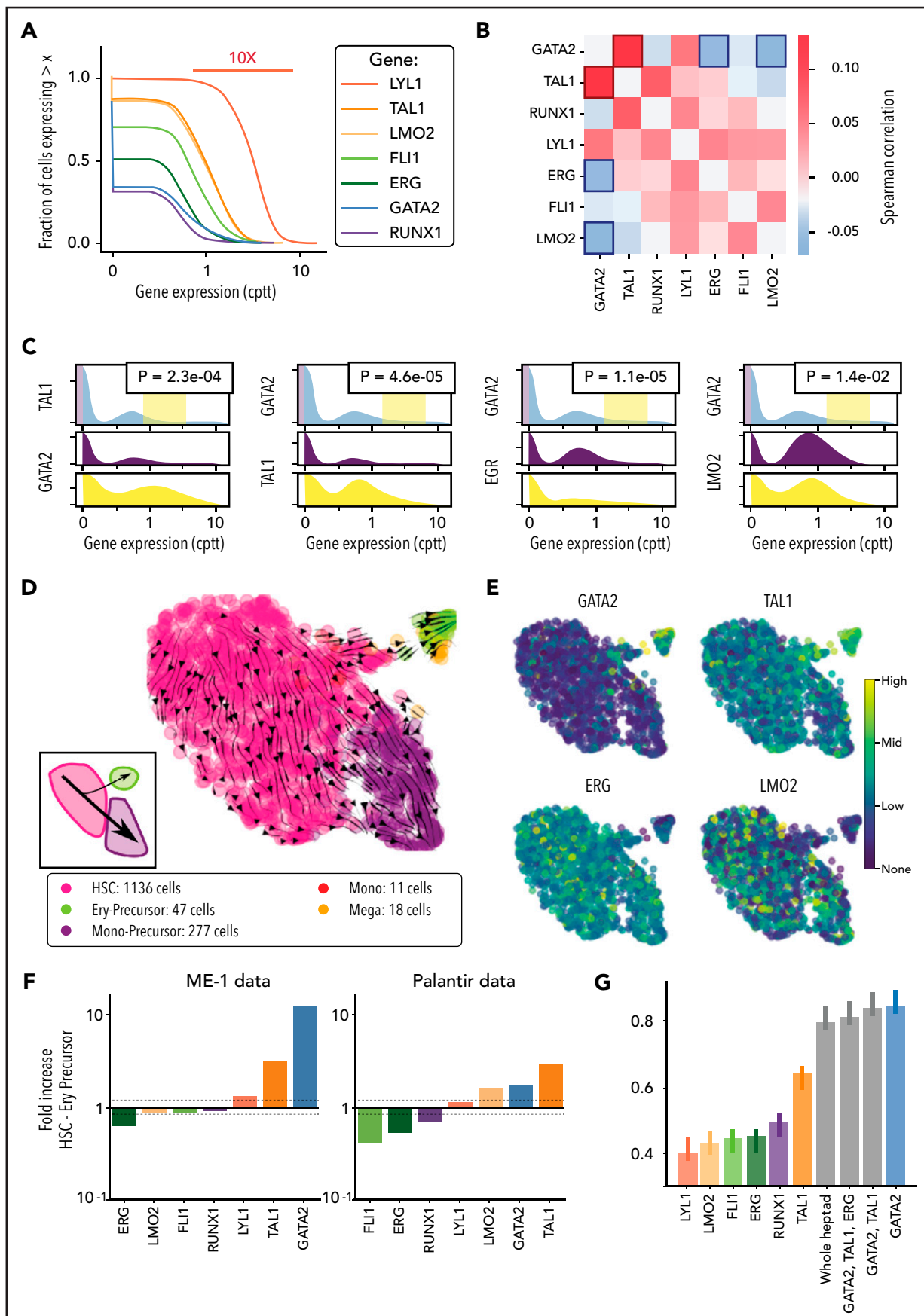
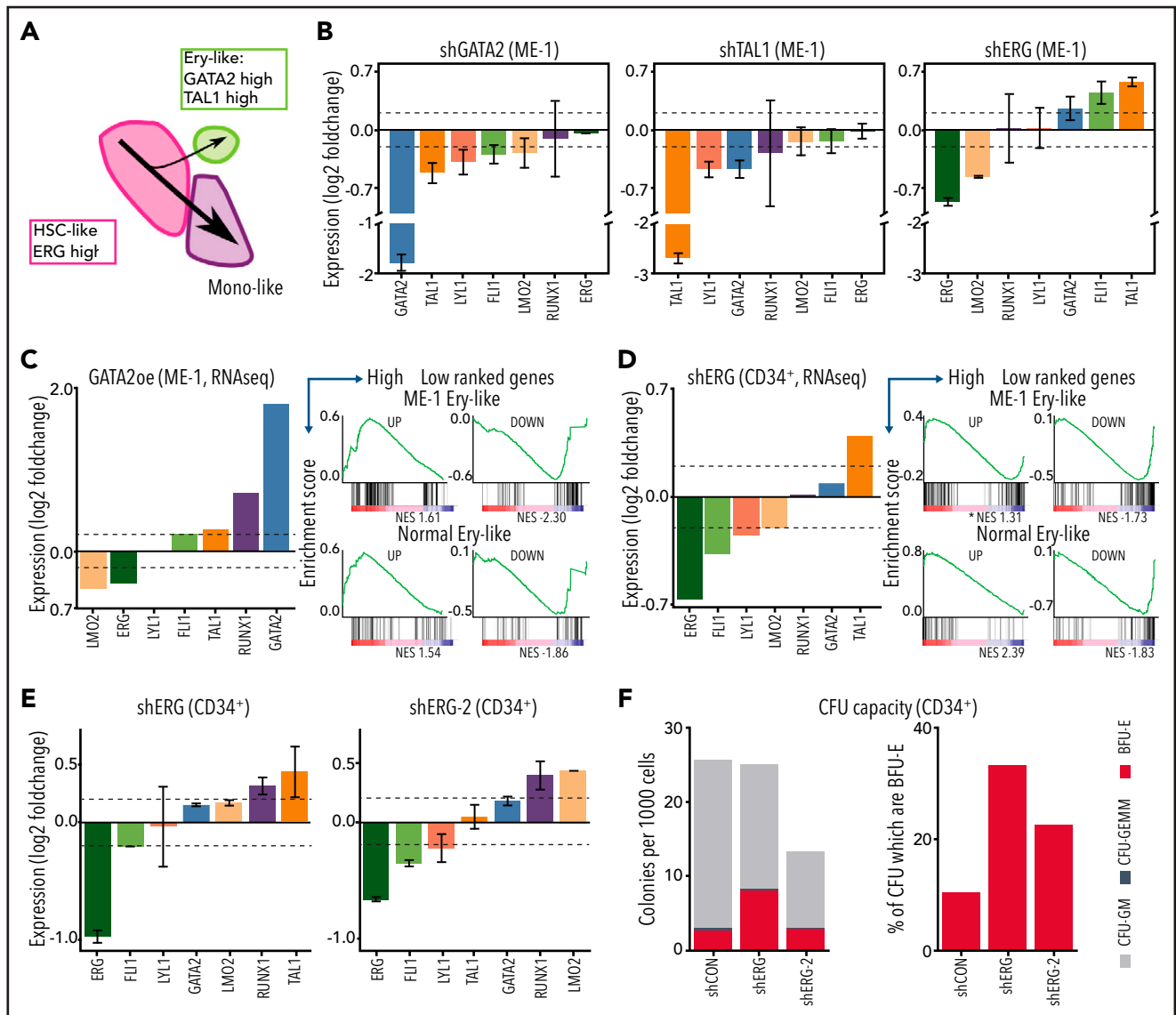


Figure 4.

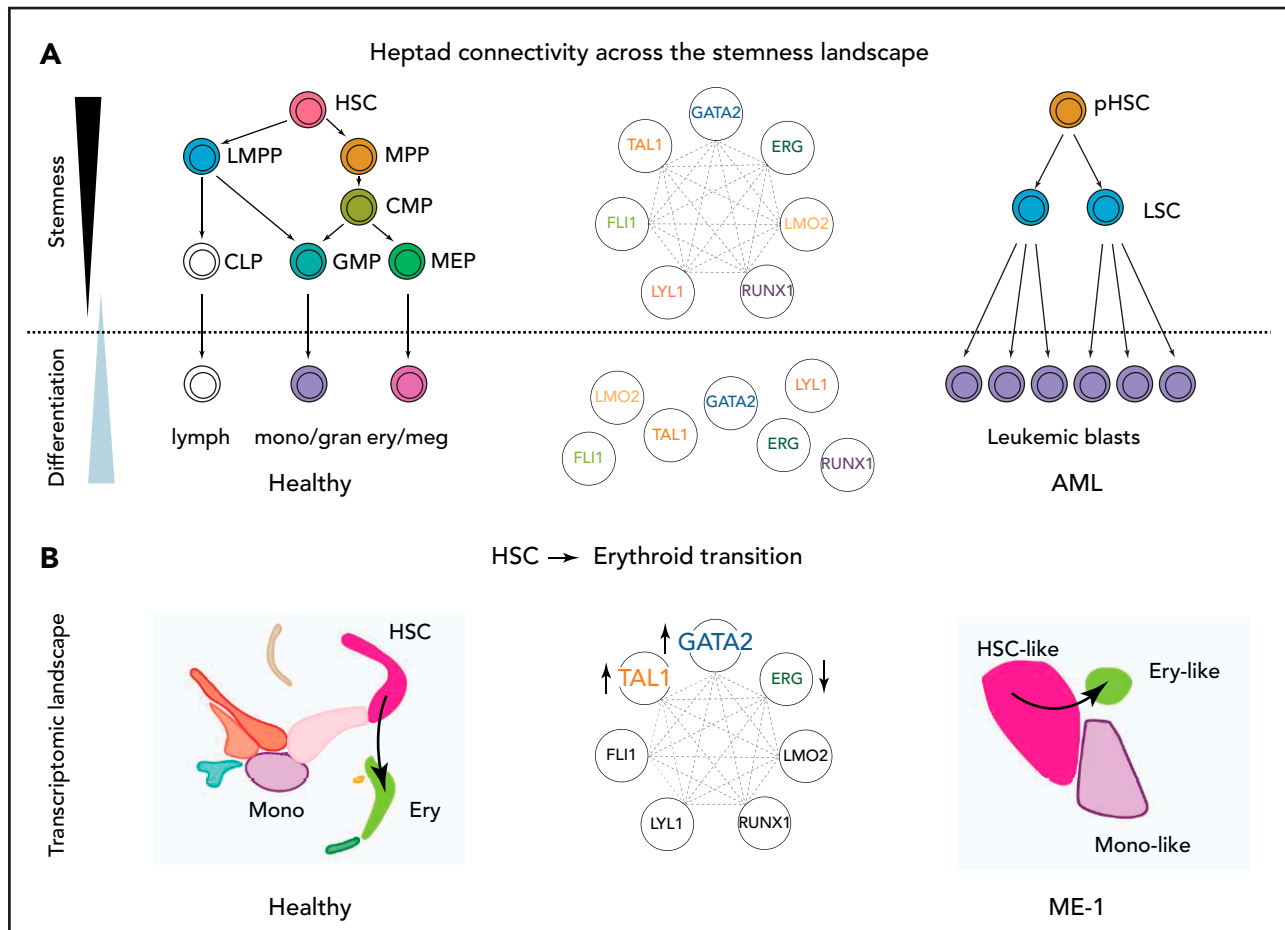


**Figure 5. Manipulating GATA2 and ERG in bulk ME-1 cells and normal CD34<sup>+</sup> HSPCs leads to altered heptad expression and can push cells toward the ery-like state.** (A) The branching phenotype within ME-1 cells indicating relative expression of key heptad genes highlighted in Figure 4. (B) Effect of knocking down GATA2, TAL1, or ERG on heptad genes in ME-1 cells (error bars show 95% confidence interval [CI]). (C) Effect of overexpressing GATA2 on heptad genes in ME-1 cells (RNAseq) (left). GSEA plots showing enrichment of genes associated with the ery-precursor/ery-precursor-like state in response to overexpressing GATA2 in ME-1 cells (right). (D) Effect of knocking down ERG on heptad genes in CD34<sup>+</sup> HSPCs (RNAseq) (left). GSEA plots showing enrichment of genes associated with the ery-precursor/ery-precursor-like state in response to knocking down ERG in CD34<sup>+</sup> HSPCs (right). False discovery rate  $q$  value for GSEA plots = 0, except where indicated by \* $q$  value = 0.02. (E) Effect of knocking down ERG on heptad genes in CD34<sup>+</sup> HSPCs using 2 different shRNAs (error bars, 95% CI). (F) Colony forming capacity of CD34<sup>+</sup> cells transduced with control (shCON) or ERG (shERG, shERG-2) shRNAs (left). CD34<sup>+</sup> cells produce colonies derived from granulocyte and/or macrophage progenitor cells (CFU-GM; gray), multipotential progenitor cells (CFU-GEMM; dark blue), and erythroid progenitor cells (blast forming unit-erythroid [BFU-E]; red). Proportion of total colonies that are erythroid (BFU-E) (right). NES, normalized enrichment score.

such as the GATA2-TAL1-ERG triad described herein may provide a conceptual framework within which to develop such therapies. A first approach would be to alter TF expression directly, as upregulating GATA2 or downregulating ERG promotes erythroid

differentiation. However, population structure of malignant cells within primary AML varies between patients and different leukemias may be primed toward specific differentiation pathways.<sup>61</sup> As such, ERG perturbation is especially promising, as this TF

**Figure 4. Single-cell transcriptomics in ME-1 cells reveals branching heterogeneity consistent with GATA2 regulation.** (A) Cumulative expression distributions for heptad genes in single ME-1 cells. cppt: counts per 10 000 reads. (B) Pairwise Spearman correlations between heptad genes in single cells. (C) Censored distributions of gene expression for the gene pairs highlighted in panel B. The 2 bottom panels show the expression of the second gene in the lowest 10% and highest 5% of expressing cells for the first gene.  $P$  values refer to a Kolmogorov-Smirnov 2-sample test between the purple and green distributions. (D) Uniform manifold approximation and projection (UMAP) embedding of ME-1 cells and cell state assignment based on Northstar<sup>55</sup> and the Palantir data, used as an atlas (Figure 1). Streamlines show RNA velocity as computed by scVelo,<sup>57</sup> projected onto the same embedding. Inset: the branching phenotype within ME-1 cells, indicating that the cell flux into the ery-precursor-like state is a rare event. (E) Expression of the 4 heptad genes highlighted in panel B on the embedded cells. (F) Fold increase in heptad gene expression across the HSC-to-ery-precursor-like state in ME-1 cells (left). Fold increase in heptad gene expression across the HSC-to-ery-precursor-like state in normal CD34<sup>+</sup> HSPCs cells (right). (G) Performance of random forest classifiers between HSC-like and ery-precursor-like states in ME-1 cells, trained solely on Palantir data with a spectrum of selected features. The presence of GATA2 expression in the model is essential for its accuracy. Error bars indicate standard deviation over 10 runs of the predictor with data resampling in each run.



**Figure 6. Proposed model of heptad activity across hematopoietic differentiation.** (A) Heptad TFs form a densely interconnected network, with key regulatory elements accessible and heptad-bound in normal and leukemic stem cells. Accessibility of regulatory elements and, consequently, heptad connectivity are reduced as cells become more differentiated. (B) scRNAseq populations in normal and ME-1 cell populations. GATA2, TAL1, and ERG promote cell state changes along the HSC-ery-precursor axis in both normal CD34<sup>+</sup> HSPCs and ME-1 cells.

appears to preserve the progenitor state rather than bias the cell toward a particular fate, and knockdown may favor exit from the stem cell state across a range of primary AMLs. A second approach would be to focus on transcriptional regulators of these TFs. USP9X, a deubiquitinase that regulates ERG stability<sup>96</sup> and is positively regulated by ERG in a feed-forward loop is one such candidate.<sup>67</sup> A third approach would be to focus on specific enhancers such as GATA2-117, which is inaccessible in normal HSCs but open in the transitional progenitor states characteristic of AML, enabling preferential cytotoxicity in leukemic cells. Overall, a deeper understanding of heptad regulatory circuits and their roles in maintaining and exiting normal and leukemic stem cell states can help shape novel, data-based approaches to innovative cancer therapies.

## Acknowledgments

The authors thank the staff and donors of the Sydney Cord Blood Bank for providing cord bloods for research.

Some of the data presented in this work were acquired by personnel and/or instruments at the Mark Wainwright Analytical Centre (MWAC) of UNSW Sydney, which is funded in part by the Research Infrastructure Programme of UNSW. This work was supported by the Anthony Rothe Memorial Trust (J.T.); a University International Postgraduate Award from UNSW Sydney and Translational Cancer Research Network-a

Translational Cancer Research Centre funded by the Cancer Institute NSW (P.T.); an International Postgraduate Student scholarship from UNSW Sydney and the Prince of Wales Clinical School and Translational Cancer Research Network-a Translational Cancer Research Centre funded by the Cancer Institute NSW (S.S.); a Peter Doherty Fellowship from the National Health and Medical Research Council of Australia (APP1073768), a Cancer Institute NSW Early Career Fellowship, the Anthony Rothe Memorial Trust, and Gilead Sciences (D.B.); a Wellcome Investigator award (206328/Z/17/Z) (B.G.); project grants from the National Health and Medical Research Council of Australia (APP1042934, APP1102589, and APP1008515), a translational program grant from the Leukemia Lymphoma Society (LLS)-Snowdome Foundation-Leukaemia Foundation, project funds from the Translational Cancer Research Network-a Translational Cancer Research Centre funded by the Cancer Institute NSW, the Anthony Rothe Memorial Trust, and philanthropic funding from Christina's Light (J.E.P.). This research was funded in whole, or in part, by the Wellcome Trust [203151/Z/16/Z] and the UKRI Medical Research Council [MC\_PC\_17230].

## Authorship

Contribution: J.A.I.T., P.T., S.S., K.K., G.H., Y.H., J.A.S., D.R.C., S.J., and J.S. performed the research and analyzed the data; D.C., A.S., D.B., and J.W.H.W. analyzed the data; I.d.J. and J.L. provided key reagents; B.G. and J.W.H.W. discussed and interpreted the data; and J.A.I.T., F.Z., and J.E.P. conceived the study and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: J.A.I.T., 0000-0002-4876-7230; Y.H., 0000-0002-7003-3110; J.A.S., 0000-0003-0413-5829; F.Z., 0000-0001-7097-8539; J.E.P., 0000-0002-0509-8962.

Correspondence: Julie I. Thoms, Level 2, Lowy Cancer Research Centre, UNSW Sydney, NSW, Australia; e-mail: j.thoms@unsw.edu.au; Fabio Zanini, Level 2, Lowy Cancer Research Centre, UNSW Sydney, NSW, Australia; e-mail: fabio.zanini@unsw.edu.au; and John E. Pimanda, Level 2, Lowy Cancer Research Centre, UNSW Sydney, NSW, Australia; e-mail: jpimanda@unsw.edu.au.

## Footnotes

Submitted 27 October 2020; accepted 3 May 2021; prepublished online on *Blood* First Edition 1 June 2021. DOI 10.1182/blood.2020009707.

\*F.Z. and J.E.P. contributed equally to this study.

Supplemental Table 3 shows the public data sets. New data are deposited under accession GSE158797. Code is available from [https://github.com/ionosofabio/heptad\\_paper](https://github.com/ionosofabio/heptad_paper).

Original data are available by e-mail request to any corresponding author.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

## REFERENCES

- Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell Stem Cell*. 2012;10(2):120-136.
- Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281.
- Buenrostro JD, Corces MR, Lareau CA, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell*. 2018;173(6):1535-1548.e1516.
- Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48(10):1193-1203.
- Karamitros D, Stoilova B, Aboukhalil Z, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol*. 2018;19(1):85-97.
- Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir [published correction appears in *Nat Biotechnol*. 2019;37(10):1237]. *Nat Biotechnol*. 2019;37(4):451-460.
- Watcham S, Kucinski I, Gottgens B. New Insights into Haematopoietic Differentiation Landscapes from scRNA-seq. *Blood*. 2019;133(13):1415-1426.
- Pimanda JE, Göttgens B. Gene regulatory networks governing haematopoietic stem cell development and identity. *Int J Dev Biol*. 2010;54(6-7):1201-1211.
- Sive JI, Göttgens B. Transcriptional network control of normal and leukaemic haematopoiesis. *Exp Cell Res*. 2014;329(2):255-264.
- Enver T, Pera M, Peterson C, Andrews PW. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell*. 2009;4(5):387-397.
- Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet*. 2016;17(11):693-703.
- Wilkinson AC, Nakauchi H, Göttgens B. Mammalian transcription factor networks: recent advances in interrogating biological complexity. *Cell Syst*. 2017;5(4):319-331.
- Thoms JAI, Beck D, Pimanda JE. Transcriptional networks in acute myeloid leukemia. *Genes Chromosomes Cancer*. 2019;58(12):859-874.
- Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med*. 2015;373(12):1136-1152.
- Horton SJ, Huntly BJ. Recent advances in acute myeloid leukemia stem cell biology. *Haematologica*. 2012;97(7):966-974.
- Jan M, Snyder TM, Corces-Zimmerman MR, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med*. 2012;4(149):149ra118.
- Shlush LI, Zandi S, Mitchell A, et al; HALT Pan-Leukemia Gene Panel Consortium. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia [published correction appears in *Nature*. 2014;508(7496):420]. *Nature*. 2014;506(7488):328-333.
- Basilico S, Göttgens B. Dysregulation of haematopoietic stem cell regulatory programs in acute myeloid leukaemia. *J Mol Med (Berl)*. 2017;95(7):719-727.
- Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci USA*. 2014;111(7):2548-2553.
- Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 1994;367(6464):645-648.
- Goardon N, Marchi E, Atzberger A, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*. 2011;19(1):138-152.
- Sarry JE, Murphy K, Perry R, et al. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2R $\gamma$ -deficient mice. *J Clin Invest*. 2011;121(1):384-395.
- Shlush LI, Mitchell A, Heisler L, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature*. 2017;547(7661):104-108.
- Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011;17(9):1086-1093.
- Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA*. 2010;304(24):2706-2715.
- Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. 1997;3(7):730-737.
- Sanz MA, Grimwade D, Tallman MS, et al. Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood*. 2009;113(9):1875-1891.
- DiNardo CD, Stein EM, de Botton S, et al. Durable remissions with ivosidenib in IDH1-mutated relapsed or refractory AML. *N Engl J Med*. 2018;378(25):2386-2398.
- Stein EM, DiNardo CD, Pollyea DA, et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood*. 2017;130(6):722-731.
- Hansen E, Quivoron C, Straley K, et al. AG-120, an oral, selective, first-in-class, potent inhibitor of mutant IDH1, reduces intracellular 2HG and induces cellular differentiation in TF-1 R132H cells and primary human IDH1 mutant AML patient samples treated ex vivo [abstract]. *Blood*. 2014;124(21). Abstract 3734.
- Popovici-Muller J, Lemieux RM, Artin E, et al. Discovery of AG-120 (ivosidenib): a first-in-class mutant IDH1 inhibitor for the treatment of IDH1 mutant cancers. *ACS Med Chem Lett*. 2018;9(4):300-305.
- Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374(23):2209-2221.
- Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
- Ley TJ, Miller C, Ding L, et al; Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo

- acute myeloid leukemia. *N Engl J Med*. 2013; 368(22):2059-2074.
35. Assi SA, Imperato MR, Coleman DJL, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*. 2019; 51(1):151-162.
  36. Yi G, Wierenga ATJ, Petraglia F, et al. Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes. *Cell Rep*. 2019;26(4):1059-1069.e6.
  37. McKeown MR, Corces MR, Eaton ML, et al. Superenhancer analysis defines novel epigenomic subtypes of non-APL AML, including an RAR $\alpha$  dependency targetable by SY-1425, a potent and selective RAR $\alpha$  agonist. *Cancer Discov*. 2017;7(10):1136-1153.
  38. Davidson EH. Emerging properties of animal gene regulatory networks. *Nature*. 2010; 468(7326):911-920.
  39. Bell CC, Fennell KA, Chan YC, et al. Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. *Nat Commun*. 2019;10(1):2723.
  40. Fennell KA, Bell CC, Dawson MA. Epigenetic therapies in acute myeloid leukemia: where to from here? *Blood*. 2019;134(22):1891-1901.
  41. Guo L, Li J, Zeng H, et al. A combination strategy targeting enhancer plasticity exerts synergistic lethality against BET1-resistant leukemia cells. *Nat Commun*. 2020;11(1):740.
  42. Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013;122(14):e12-e22.
  43. Diffner E, Beck D, Gudgin E, et al. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia [published correction appears in *Blood*. 2014;123(18):2901]. *Blood*. 2013;121(12): 2289-2300.
  44. Wilson NK, Foster SD, Wang X, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010;7(4):532-544.
  45. Guibentif C, Rönn RE, Böiers C, et al. Single-cell analysis identifies distinct stages of human endothelial-to-hematopoietic transition. *Cell Rep*. 2017;19(1):10-19.
  46. Bergiers I, Andrews T, Vargel Bölükbaşı Ö, et al. Single-cell transcriptomics reveals a new dynamical function of transcription factors during embryonic hematopoiesis. *eLife*. 2018; 7:e29312.
  47. Oram SH, Thoms JA, Pridans C, et al. A previously unrecognized promoter of LMO2 forms part of a transcriptional regulatory circuit mediating LMO2 expression in a subset of T-acute lymphoblastic leukaemia patients. *Oncogene*. 2010;29(43):5796-5808.
  48. Curtis DJ, Salmon JM, Pimanda JE. Concise review: Blood relatives: formation and regulation of hematopoietic stem cells by the basic helix-loop-helix transcription factors stem cell leukemia and lymphoblastic leukemia-derived sequence 1. *Stem Cells*. 2012;30(6):1053-1058.
  49. Marcucci G, Baldus CD, Ruppert AS, et al. Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2005;23(36):9234-9242.
  50. Li Y, Luo H, Liu T, Zacksenhaus E, Ben-David Y. The ets transcription factor Fli-1 in development, cancer and disease. *Oncogene*. 2015;34(16):2022-2031.
  51. Mandoli A, Singh AA, Jansen PW, et al. CBF $\beta$ -MYH11/RUNX1 together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. *Leukemia*. 2014;28(4):770-778.
  52. Mandoli A, Singh AA, Prange KHM, et al. The hematopoietic transcription factors RUNX1 and ERG prevent AML1-ETO oncogene overexpression and onset of the apoptosis program in t(8;21) AMLs. *Cell Rep*. 2016;17(8): 2087-2100.
  53. Sotoca AM, Prange KH, Reijnders B, et al. The oncofusion protein FUS-ERG targets key hematopoietic regulators and modulates the all-trans retinoic acid signaling pathway in t(16;21) acute myeloid leukemia. *Oncogene*. 2016;35(15):1965-1976.
  54. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-W191.
  55. Zanini F, Berghuis BA, Jones RC, et al. Northstar enables automatic classification of known and novel cell types from tumor samples. *Sci Rep*. 2020;10(1):15251.
  56. La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. *Nature*. 2018; 560(7719):494-498.
  57. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;38(12):1408-1414.
  58. Yamazaki H, Suzuki M, Otsuki A, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell*. 2014; 25(4):415-427.
  59. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014; 157(2):369-381.
  60. Rucker FG, Sander S, Döhner K, Döhner H, Pollack JR, Bullinger L. Molecular profiling reveals myeloid leukemia cell lines to be faithful model systems characterized by distinct genomic aberrations. *Leukemia*. 2006;20(6):994-1001.
  61. van Galen P, Hovestadt V, Wadsworth Ii MH, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 2019;176(6):1265-1281.e24.
  62. Yanagisawa K, Horiuchi T, Fujita S. Establishment and characterization of a new human leukemia cell line derived from M4E0. *Blood*. 1991;78(2):451-457.
  63. Yi G, Mandoli A, Jussen L, et al. CBF $\beta$ -MYH11 interferes with megakaryocyte differentiation via modulating a gene program that includes GATA2 and KLF1. *Blood Cancer J*. 2019;9(3): 33.
  64. Tursky ML, Beck D, Thoms JA, et al. Overexpression of ERG in cord blood progenitors promotes expansion and recapitulates molecular signatures of high ERG leukemias. *Leukemia*. 2015;29(4):819-827.
  65. Kucinski I, Wilson NK, Hannah R, et al. Interactions between lineage-associated transcription factors govern haematopoietic progenitor states. *EMBO J*. 2020;39(24): e104983.
  66. Johnson KD, Conn DJ, Shishkova E, et al. Constructing and deconstructing GATA2-regulated cell fate programs to establish developmental trajectories. *J Exp Med*. 2020; 217(11):e20191526.
  67. Aqaq N, Yassin M, Yassin AA, et al. An ERG enhancer-based reporter identifies leukemia cells with elevated leukemogenic potential driven by ERG-USP9X feed-forward regulation. *Cancer Res*. 2019;79(15):3862-3876.
  68. Yassin M, Aqaq N, Yassin AA, et al. A novel method for detecting the cellular stemness state in normal and leukemic human hematopoietic cells can predict disease outcome and drug sensitivity. *Leukemia*. 2019;33(8):2061-2077.
  69. Osada H, Grutz G, Axelson H, Forster A, Rabbitts TH. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci USA*. 1995; 92(21):9585-9589.
  70. Wadman I, Li J, Bash RO, et al. Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. *EMBO J*. 1994;13(20):4831-4839.
  71. Donaldson IJ, Chapman M, Kinston S, et al. Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. *Hum Mol Genet*. 2005;14(5): 595-601.
  72. Pimanda JE, Ottersbach K, Knezevic K, et al. Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci USA*. 2007;104(45):17692-17697.
  73. Wontakal SN, Guo X, Smith C, et al. A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation. *Proc Natl Acad Sci USA*. 2012; 109(10):3832-3837.
  74. Eich C, Arlt J, Vink CS, et al. In vivo single cell analysis reveals Gata2 dynamics in cells transitioning to hematopoietic fate. *J Exp Med*. 2018;215(1):233-248.
  75. Menendez-Gonzalez JB, Vukovic M, Abdelfattah A, et al. Gata2 as a crucial regulator of stem cells in adult hematopoiesis

- and acute myeloid leukemia. *Stem Cell Reports*. 2019;13(2):291-306.
76. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011;43(10):1012-1017.
  77. Vicente C, Vazquez I, Conchillo A, et al. Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities. *Leukemia*. 2012;26(3):550-554.
  78. Lancrin C, Sroczynska P, Stephenson C, Allen T, Kouskoff V, Lacaud G. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature*. 2009;457(7231):892-895.
  79. Elwood NJ, Zogos H, Pereira DS, Dick JE, Begley CG. Enhanced megakaryocyte and erythroid development from normal human CD34(+) cells: consequence of enforced expression of SCL. *Blood*. 1998;91(10):3756-3765.
  80. Mikkola HK, Klintman J, Yang H, et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature*. 2003;421(6922):547-551.
  81. Robertson SM, Kennedy M, Shannon JM, Keller G. A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. *Development*. 2000;127(11):2447-2459.
  82. Taoudi S, Bee T, Hilton A, et al. ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. *Genes Dev*. 2011;25(3):251-262.
  83. Knudsen KJ, Rehn M, Hasemann MS, et al. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev*. 2015;29(18):1915-1929.
  84. Marcucci G, Maharry K, Whitman SP, et al; Cancer and Leukemia Group B Study. High expression levels of the ETS-related gene, ERG, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2007;25(22):3337-3343.
  85. Schwind S, Marcucci G, Maharry K, et al. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood*. 2010;116(25):5660-5669.
  86. Metzeler KH, Dufour A, Benthaus T, et al. ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J Clin Oncol*. 2009;27(30):5031-5038.
  87. Thoms JA, Birger Y, Foster S, et al. ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*. 2011;117(26):7079-7089.
  88. Goldberg L, Tijssen MR, Birger Y, et al. Genome-scale expression and transcription factor binding profiles reveal therapeutic targets in transgenic ERG myeloid leukemia. *Blood*. 2013;122(15):2694-2703.
  89. Carmichael CL, Metcalf D, Henley KJ, et al. Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc Natl Acad Sci USA*. 2012;109(38):15437-15442.
  90. Salek-Ardakani S, Smooha G, de Boer J, et al. ERG is a megakaryocytic oncogene. *Cancer Res*. 2009;69(11):4665-4673.
  91. Nowak D, Stewart D, Koeffler HP. Differentiation therapy of leukemia: 3 decades of development. *Blood*. 2009;113(16):3655-3665.
  92. Namasu CY, Katzerke C, Bräuer-Hartmann D, et al. ABR, a novel inducer of transcription factor C/EBP $\alpha$ , contributes to myeloid differentiation and is a favorable prognostic factor in acute myeloid leukemia. *Oncotarget*. 2017;8(61):103626-103639.
  93. Radomska HS, Jernigan F, Nakayama S, et al. A cell-based high-throughput screening for inducers of myeloid differentiation. *J Biomol Screen*. 2015;20(9):1150-1159.
  94. Antony-Debré I, Paul A, Leite J, et al. Pharmacological inhibition of the transcription factor PU.1 in leukemia. *J Clin Invest*. 2017;127(12):4297-4313.
  95. Morita K, Suzuki K, Maeda S, et al. Genetic regulation of the RUNX transcription factor family has antitumor effects. *J Clin Invest*. 2017;127(7):2815-2828.
  96. Wang S, Kollipara RK, Srivastava N, et al. Ablation of the oncogenic transcription factor ERG by deubiquitinase inhibition in prostate cancer. *Proc Natl Acad Sci USA*. 2014;111(11):4251-4256.