HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2023 January 26.

Published in final edited form as:

Cell Rep. 2022 December 20; 41(12): 111830. doi:10.1016/j.celrep.2022.111830.

EKLF/Klf1 regulates erythroid transcription by its pioneering activity and selective control of RNA Pol II pause-release

Kaustav Mukherjee[1,2,4], **James J. Bieker**[1,2,3,4,5,*]

[1]Icahn School of Medicine at Mount Sinai, Department of Cell, Developmental, and Regenerative Biology, Biology Box 1020, One Gustave L. Levy Place, New York, NY 10029, USA

[2]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[3]Tisch Cancer Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[4]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[5]Lead contact

SUMMARY

EKLF/Klf1 is a zinc-finger transcription activator essential for erythroid lineage commitment and terminal differentiation. Using ChIP-seq, we investigate EKLF DNA binding and transcription activation mechanisms during mouse embryonic erythropoiesis. We utilize the Nan/+ mouse that expresses the EKLF-E339D (Nan) variant mutated in its conserved zinc-finger region and address the mechanism of hypomorphic and neomorphic changes in downstream gene expression. First, we show that Nan-EKLF limits normal EKLF binding to a subset of its sites. Second, we find that ectopic binding of Nan-EKLF occurs largely at enhancers and activates transcription through pioneering activity. Third, we find that for a subset of ectopic targets, gene activation is achieved in Nan/+ only by Nan-EKLF binding to distal enhancers, leading to RNA polymerase II pause-release. These results have general applicability to understanding how a DNA binding variant factor confers dominant disruptive effects on downstream gene expression even in the presence of its normal counterpart.

Graphical abstract

*Correspondence: james.bieker@mssm.edu.

AUTHOR CONTRIBUTIONS

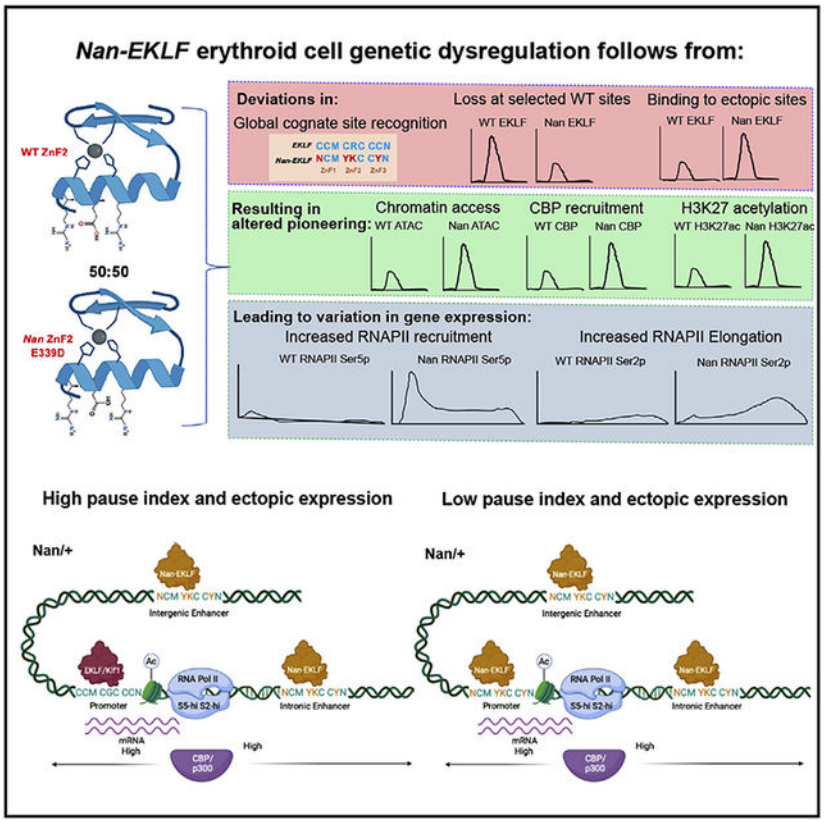
K.M. performed experiments and analyzed the data; K.M. and J.J.B. designed the experiments, interpreted the results, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111830.



In brief

Mukherjee and Bieker identify transcriptional mechanisms leading to dysregulation of gene expression in the Nan/+ erythroid cells, which express a mutant variant (E339D) of the master EKLF/KLF1 transcriptional regulator. They illuminate the ways by which co-expression of a variant factor interferes with the proper function of its wild-type counterpart.

INTRODUCTION

The Krüppel-like factor (KLF) family of transcription factors regulates gene expression to facilitate diverse biological processes.[1-3] A notable member, KLF4, is involved in maintenance of pluripotency and is one of four factors used for cellular reprogramming to generate induced pluripotent stem cells.[4] Many KLF proteins play crucial roles during hematopoiesis, especially during embryonic development.[5-8]

EKLF/Klf1 is an erythroid transcription factor that activates expression of genes required for lineage commitment, terminal erythropoiesis, enucleation, and erythroblast island macrophage function.[9-13] EKLF/Klf1 binds a tripartite DNA sequence, CCM-CRC-CCN, using three tandem C2H2 zinc fingers at its C terminus.[14,15] EKLF binds to the regulatory elements of most erythroid genes, such as globins, erythrocyte membrane proteins, cell-cycle factors, and iron/heme regulators.[16]

EKLF mutations have profound effects on mammalian erythropoiesis.[17-24] While embryonic lethality at E14.5 in homozygous EKLF$^{-/-}$ mice[25,26] can be attributed to the global downregulation of EKLF targets,[27,28] monoallelic point mutations in the EKLF zinc finger, such as the mouse EKLF-E339D (neonatal anemia; Nan) and human EKLF-E325K (congenital dyserythropoietic anemia; CDA type IV), cause dominant phenotypes resulting from aberrant transcriptional changes.[17,29-34] In each case the same conserved glutamate in the second zinc finger of EKLF is mutated (reviewed in Kulczynska-Figurny et al.[35]). The Nan E339D mutation is embryonic lethal at E10–11 in Nan/Nan homozygotes,[36,37] and Nan/− heterozygotes are dysmorphic prior to E12.5[18], thus causing a more severe phenotype than EKLF$^{-/-}$. Nan/+ heterozygotes survive until adulthood with severe anemia,[17,31] and global gene expression and EKLF DNA binding analysis in the context of Nan reveal altered EKLF binding preference and ectopic expression of non-erythroid genes.[30-32,35] The human EKLF/KLF1-E325K mutation causes CDA IV (OMIM: 613673), with which patients present with severe hemolytic anemia and a range of other symptoms[38-41] resulting from aberrant transcriptomic effects during terminal erythroid differentiation.[33,34,42,43]

EKLF/Klf1 utilizes the cellular transcription machinery comprising general transcription factors and co-factors via interactions with its N-terminal transactivation domain (reviewed in Yien and Bieker[11]). Among the prominent interactors are protein acetyltransferases CBP and P300, each of which acetylates EKLF at two specific sites[44-47] and also acetylates H3K27,[48] leading to transcriptionally poised promoters[49,50] and active enhancers.[51-53] Acetylation of EKLF facilitates its interaction with the BRG1 subunit of the SWI/SNF chromatin remodeling complex,[6,54-56] thus providing a mechanistic link to EKLF's role in chromatin remodeling.[57] The EKLF-CBP interaction may thus be crucial for transcription activation, since it can facilitate chromatin opening and pre-initiation complex (PIC) formation at transcription start sites (TSSs).[58] This leads to transcription initiation by RNA polymerase II (Pol II) and pausing near the TSS and subsequent release of the paused RNA Pol II into elongation upon enhancer activation.[59-62] RNA Pol II pause-release is the rate-limiting step of transcription in many metazoan genes governing cell-fate decisions and development.[63-65] Paused RNA Pol II is phosphorylated at serine 5 of its C-terminal domain (CTD) repeat (YSPTSPS), while elongating RNA Pol II is phosphorylated at Ser2 of its CTD[66] by the action of positive elongation factors recruited in response to acetylated histones at promoters and enhancers.[67]

Our study focuses on global EKLF binding dynamics during embryonic erythropoiesis in primary wild-type (WT) and Nan/+ mouse fetal liver and its correlation with chromatin accessibility, CBP occupancy, and histone acetylation and finally its effect on RNA Pol II pausing and elongation. Our goal is to elucidate the mechanisms of transcription activation by EKLF/Klf1 during embryonic erythropoiesis *in vivo* and in the context of RNA Pol II pause-release control. In addition, we aim to understand the unusually severe effects of conservative E-to-D change in Nan-EKLF and the molecular mechanisms leading to dominant anemia through global gene dysregulation. The Nan/+ mouse as a model to study transcription activation by EKLF provides a significant advantage, since ectopic DNA binding and gene activation by Nan-EKLF are exclusive to this system. These details are also globally relevant to understanding how cell-restricted transcriptional modulators

play a coordinating role in establishing selective gene regulation during development and differentiation.

RESULTS

Terminal erythroid maturation is reduced during embryonic erythropoiesis in Nan/+

E13.5 Nan/+ fetal livers (FLs) are pale with reduced total cell numbers, indicative of impaired erythropoiesis.[17,18] To quantify the progress of erythroid maturation in the FL, we examined the expression of the terminal erythroid marker Ter119 (Figure S1A). Mature Ter119+ erythroid cells in Nan/+ FL are significantly reduced compared with WT, indicating blocks in erythroid progression (Figure S1B). Moreover, the Ter119+ Nan/+ FL cells have lower Ter119 expression compared with WT (Figures S1C and S1D). Further, we assessed the expression of CD44, whose expression reduces progressively with decreasing cell size during erythroid maturation[68] in the Ter119+ population (Figure S1E). WT FL shows three distinct populations with progressive maturation stages from P1 to P3 (Figure S1F). The mature P3 population is significantly reduced in Nan/+, whereas the less mature P1 and P2 populations are higher in Nan/+ (Figure S1G), indicating that most of the erythroblasts in Nan/+ FL are early stage. These observations depict a delay or partial block in terminal erythroid maturation in Nan/+ during embryonic erythropoiesis.

The Nan mutation alters EKLF DNA binding preference as well as the genomic distribution of EKLF binding sites in Nan/+ FL

To determine the effect of the heterozygous Nan mutation on EKLF DNA binding during embryonic erythropoiesis in Nan/+, we used chromatin immunoprecipitation sequencing (ChIP-seq) on E13.5 FLs with a custom-designed anti-EKLF antibody called 7B2a. Peak analysis (Table S1) showed that about one-third of EKLF peak locations in WT are unoccupied in Nan/+ (Figure 1A) and vice versa (Figure 1B). There are more peaks in Nan/+ compared with WT (Figure 1C). Most WT EKLF peaks are located at promoters or TSSs, but significant binding is also seen at annotated intronic and intergenic regions (Figure 1D). This is significantly altered in Nan/+, with reduced binding at promoters and increased binding to introns and intergenic regions, indicating ectopic binding (Figure 1D). On examining select overexpressed or ectopically expressed genes in Nan/+,[32] we found that for *Ass1* and *Gng2* there is ectopic Nan-EKLF binding at promoters and introns (Figure S2A), while *Serpinel* has Nan-EKLF bound exclusively at its promoter (Figure S2B). Others, such as *IL4ra*, *Pacs2*, and *Psap*, have ectopic Nan-EKLF binding at a specific intron (Figure S2C).

To determine the specificity of 7B2a, we performed motif analysis of peaks in WT and Nan/+. We found that in WT, the canonical EKLF/Klf1 motif is the top enriched motif, thus attesting to the high specificity of 7B2a (Figure 1E). The motif enriched in Nan/+ is a variant of the EKLF motif, with a strong preference for G (on the C-rich strand) at the second nucleotide of the middle zinc-finger binding sequence (Figure 1E, asterisk). This is consistent with earlier findings that Nan-EKLF does not prefer binding to the cognate EKLF binding sequence with CAC (on the C-rich strand) due to the nature of the Nan mutation.[17,30] For peaks called exclusively in Nan/+ and not in WT (likely bound by

Nan-EKLF), we found that, along with the altered preference in the middle nucleotide (Figure S3A, denoted by star), there is also degeneracy at the adjacent nucleotide (Figure S3A, denoted by circle) which is normally strictly G (on the G-rich strand) in WT.[30]

Other transcription factor motifs, such as GATA, are co-enriched with WT EKLF motifs (Figure 1F). This is consistent with the known cooperativity of EKLF and Gata1 in transcription regulation during definitive erythropoiesis.[27] Interestingly, GATA motifs are not co-enriched with EKLF peaks in Nan/+ (Figure 1G), suggesting that Nan-EKLF binds to regions of the genome independent of Gata1 binding. On examining the occupancy of Gata1 using ChIP-seq data from E12.5 FL,[69] we found that most WT EKLF peaks in E13.5 FL have co-occupancy with Gata1 (Figure S3B), consistent with our motif analysis (Figure 1F). In contrast, very few of the EKLF peaks in Nan/+ show co-occupancy with Gata1 (Figure S3C), explaining the lack of co-enriched GATA and EKLF motifs in Nan/+.

Pioneering by EKLF is robust at promoter regions in the presence of Gata1 and CBP and is altered in Nan/+ in specific contexts

EKLF's known interaction with CBP[45-47] led us to measure the global effect of EKLF binding on CBP occupancy. We also measured the effect of CBP binding by examining H3K27ac using ChIP-seq, since H3K27 is a prominent substrate of CBP[70] and allows us to locate enhancer regions. We separated the EKLF peaks at promoters and enhancers in WT and Nan/+ and determined co-occupancy with CBP, H3K27ac, and Gata1 (Figure 2). Pioneering activity also correlates with high chromatin accessibility, and thus, we corroborated CBP and H3K27ac levels with ATAC-seq.

First, we found that EKLF at promoters in both WT and Nan/+ has high co-occupancy with CBP and H3K27ac and the promoters have accessible chromatin (Figures 2A and 2B). Promoter-bound EKLF in both WT and Nan/+ also correlates highly with Gata1 occupancy at E12.5 (Figures 2A and 2B). An example of this is the EKLF target gene *Epb42*, with EKLF binding at the promoter in both WT and Nan/+ correlating with a Gata1 peak at E12.5 and accessible chromatin at the promoter (Figure S4A). Second, we found that mostly WT EKLF intronic and intergenic peaks are cooccupied with CBP, have high H3K27ac levels indicating active enhancers, and have accessible chromatin (Figure 2C). This is visible at the *E2f2* (Figure S4B) and *Dmtn* (Figure S4C) loci, where there are multiple EKLF peaks at intronic enhancers, some of which also correlate with Gata1 peaks. Interestingly, at the *Dmtn* locus, the third and fourth intronic enhancer sites, which show both WT EKLF and Gata1 peaks, lose EKLF binding in Nan/+ and also lose accessible chromatin and CBP in Nan/+ (Figure S4C). This indicates that at these sites pioneering requires the presence of EKLF, without which chromatin accessibility is lost. In addition, in some cases, intronic enhancers may be bound by both EKLF and Gata1, but pioneering is visible only at the promoters, such as in the *Alad* locus (Figure S4D).

In Nan/+, CBP, H3K27ac, and chromatin accessibility do not correlate as highly with non-promoter-bound EKLF (Figure 2D) as they do with promoter-bound EKLF (Figure 2B). Cooperativity in pioneering activity between EKLF and Gata1 is likely mediated at least in part by their interactions with the CBP/p300 complex.[71] Overall, the intronic and intergenic EKLF peaks in Nan/+ have low Gata1, CBP, H3K27ac, and chromatin accessibility (Figure

2D) compared with promoters in Nan/+ (Figure 2B). These observations suggest that the pioneering ability of EKLF is robust in the presence of Gata1 and more effective at promoter regions than at intronic/intergenic regions in Nan/+. This is clearly visible at the *Fam159b* (*Shisal2b*) locus, where a strong Nan-EKLF peak at the TSS correlates with high amounts of CBP, H3K27ac, and accessible chromatin specifically in Nan/+ (Figure S4E). The *Il4ra* locus has ectopic Nan-EKLF binding at a Gata1-occupied site in an intronic enhancer, which causes increased chromatin accessibility at that site along with higher CBP, but no difference is observed at the TSS between WT and Nan/+ (Figure S4F). The strongest evidence that EKLF alone can act as a pioneer transcription factor comes from genes such as *Sema4b* and *Larp4b* (Figures S4G and S4H), along with *Fam159b*, shown earlier (Figure S4E). The *Sema4b* locus shows a couple of different patterns (Figure S4G). At intron 1 there is strong ectopic Nan-EKLF binding leading to a moderate increase in CBP and accessible chromatin, but no Gata1 (Figure S4G, first box from left). At intron 2 there is a WT EKLF peak coinciding with Gata1 binding, which loses EKLF binding in Nan/+ (Figure S4G, second box from left). Interestingly, despite the loss of EKLF binding here, chromatin is still accessible in Nan/+, indicating that at this site the presence of Gata1 alone is sufficient for pioneering, whereas at the intron 1 site the presence of Nan-EKLF alone can perform pioneering. *Larp4b* also has an ectopic Nan-EKLF site in intron 1, which has minimal Gata1 binding, but has increased chromatin accessibility in Nan/+ due to ectopic Nan-EKLF binding (Figure S4H).

Loss of EKLF and CBP binding downregulates genes in Nan/+ by decreased enhancer activation and reduced RNA polymerase II pausing and elongation

Since EKLF is a transcription activator, we examined the effect of EKLF occupancy and pioneering in WT and Nan/+ on gene expression. We correlated EKLF peaks in WT and Nan/+ with RNA-sequencing (RNA-seq) data from our earlier studies on Nan/+[30,32] to determine direct gene targets of EKLF that are differentially expressed in Nan/+. We reanalyzed the RNA-seq data using Salmon[72] by mapping it to the latest mouse genome annotations and found 204 downregulated and 260 ectopic or overexpressed genes in Nan/+ (Table S2). Of these, 91 downregulated genes also had EKLF peaks in WT at either promoters or enhancers and were thus directly activated by EKLF (Figure 3A and Table S3). We found that EKLF binding is lost or reduced at both promoters and intronic regions of these targets in Nan/+ (Figures 3A and S5A), and motif analysis of these peaks revealed that the EKLF motif with CAC loses EKLF binding in Nan/+ (Figure S5B). This is consistent with our earlier findings[17] (Figure 1E), but reveals that even in the presence of WT EKLF in Nan/+, EKLF binding to specific CAC motifs is limited, and these genes lose expression. Downregulated EKLF targets such as *Dmtn*, *Tfrc* (*CD71*), and *Steap3* have EKLF peaks at their promoters and introns, and these are absent in Nan/+ (Figure S5C). Others, such as *Tspo2*, *Rgcc*, and *Cdkn2c* (*p18*), have promoter-bound EKLF, whereas *E2f2* has EKLF binding at intronic enhancers, which is absent in Nan/+ (Figure S5C). EKLF binding is dramatically reduced in Nan/+ at the β-like globin locus Locus Control Regions and also (but less pronounced) at the β-globin promoters (Figure S5C).

Next, we determined the mechanism of gene downregulation in Nan/+ by assessing the contribution of CBP/p300-mediated H3K27 acetylation and its effect on RNA Pol II

pausing and elongation. Reduced EKLF binding in Nan/+ leads to reduced CBP occupancy predominantly at introns (Figure 3B). Enhancer activation, however, is drastically reduced at both TSSs and introns as evidenced by a significant reduction in H3K27ac in Nan/+ (Figure 3C). At the highly downregulated *Dmtn* locus, there is reduced CBP, H3K27ac, and accessible chromatin in Nan/+ at the third and fourth EKLF peak *Dmtn* introns (Figure S5D). In contrast, at the *Rgcc* locus, there is reduced EKLF and chromatin accessibility at the TSS in Nan/+ and reduced CBP and H3K27ac at the TSS and introns (Figure S5E). Finally, at the *E2f2* locus, reduced EKLF, CBP, H3K27ac, and accessible chromatin in Nan/+ are seen only at intronic enhancers (Figure S5F).

To assess the effect of reduced enhancer activation on RNA Pol II, we examined global occupancy of both paused (phospho-Ser5) and elongating (phospho-Ser2) forms of RNA Pol II. We found overall reduced paused (Figure 3D) and elongating (Figure 3E) RNA Pol II at the downregulated EKLF targets in Nan/+. This indicates that gene downregulation due to a loss of EKLF binding is mainly regulated by RNA Pol II recruitment and pausing at the TSS rather than by regulating RNA Pol II pause-release. This is apparent when we examine the correlation between paused Pol II at the TSS and elongating Pol II at the gene bodies in WT and Nan/+ (Figure S5G). In Nan/+, there is reduced Ser5p at the TSS and reduced Ser5p and Ser2p at the gene bodies compared with WT (Figure S5G).

To determine the consequence of gene downregulation on the phenotypic effects seen in Nan/+, we performed gene ontology (GO) analysis on the downregulated EKLF targets (Table S4). We found that they are involved in erythrocyte homeostasis, transport, and localization and mostly comprise erythrocyte membrane proteins critical for membrane integrity and signaling (Table S4). Of these, *Dmtn* and *Tspo2* are important erythrocyte membrane proteins strongly downregulated in Nan/+ (Figure 3F) due to the loss of both paused and elongating RNA Pol II in Nan/+ resulting from reduced enhancer activation by CBP at the TSS and intronic regions (Figure 3G). In addition, *Tfrc* (*CD71*) shows a moderate 2-fold downregulation in Nan/+ (Figure 3F) with roughly equivalent pausing in WT and Nan/+ (Figure 3G, compare Pol2 Ser5), but lower elongation in Nan/+ (Figure 3G – compare Pol2 Ser2). A few downregulated EKLF targets important for cell-cycle regulation, such as *E2f2*, *Cdkn2c*, and *Rgcc* (Figure 3H), show reduced enhancer activation, lower paused RNA Pol II at their TSS, and lower elongation in Nan/+ (Figure 3I). Thus, important erythroid EKLF targets lose expression in Nan/+ due to reduced enhancer activation and RNA Pol II recruitment.

Ectopic binding of Nan-EKLF and CBP to intronic and intergenic distal enhancers influences pioneering and enhancer activation at the TSS of certain genes, leading to their ectopic or overexpression in Nan/+

Since a major cause of deleterious effects in Nan/+ can be attributed to the ectopic or overexpression of genes not normally expressed during terminal erythroid maturation,[32] we assessed which mechanisms of transcription activation are employed by Nan-EKLF at these loci. First, we examined EKLF occupancy at the ectopic or overexpressed genes in Nan/+ that also have EKLF ChIP-seq peaks and are direct targets of Nan-EKLF (Table S3) and found higher occupancy at intronic and intergenic regions in Nan/+ (Figure 4A). Further,

we examined the enrichment of EKLF peaks at these loci (promoter or intronic/intergenic enhancer) and found higher enrichment at introns/intergenic locations in Nan/+ (Figure 4B). This suggests that Nan-EKLF (and by extension, EKLF) can activate transcription effectively by binding to intronic or intergenic enhancers and influence events at the promoter/TSS regions.

Next, we correlated ectopic Nan-EKLF binding and transcription activation with enhancer activation mediated by its interaction with CBP. We found that CBP levels are significantly higher at both the promoter/TSS and the intronic/intergenic regions (Figure 4C), and this also correlates with moderately higher H3K27ac at the TSS of these genes (Figure 4D). This suggests that Nan-EKLF can independently activate enhancers near the TSS through CBP even by binding elsewhere in the gene body. GO analysis of ectopic Nan-EKLF targets revealed genes with functions such as apoptosis and immune functions that could likely impede terminal erythroid maturation (Table S4). Some of these genes are also involved in stress-responsive transcription and proinflammatory responses mediated by IL-6 that could worsen the effects on terminal erythropoiesis in Nan/+ (Table S4).

In Nan/+ there is a sharp intronic enhancer peak at the first intron that correlates with an ATAC peak at that location, higher CBP in the vicinity, and higher H3K27ac along the entire body of the *Psap* and *Il4rα* genes (Figures 4E and 4F) and *Pacs2* (Figure S6A), all of which have functions related to immune regulation and cell death (see supplemental information). At the *Psap* locus, the ectopic Nan-EKLF peak at the distal enhancer correlates with higher CBP and ATAC peaks at the TSS and the enhancer in Nan/+ (Figure 4E). Thus, pioneering and enhancer activation may depend on the context of EKLF and CBP binding, with higher likelihood at promoters marked by higher H3K27ac levels, but enhanced by distal EKLF/CBP at intronic or intergenic regions. A minor variation is seen at the *Il4rα* locus, where EKLF, CBP, and accessible chromatin at the TSS are comparable between WT and Nan/+, but ectopic Nan-EKLF binding to an intronic enhancer drives higher CBP occupancy and H3K27 acetylation in Nan/+ (Figure 4F). *Hmox1* is another important ectopic gene that leads to potentially deleterious effects in Nan/+[32] and has two ATAC peaks upstream of its promoter (Figure 4G). Peak 1 has no discernible differences in any of the factors, while at peak 2 there is ectopic Nan-EKLF binding and a slightly higher chromatin accessibility (Figure 4G). Thus, *Hmox1* overexpression in Nan/+ is likely driven by Nan-EKLF binding to the upstream distal intergenic enhancer site 2. The gene with highest ectopic expression in Nan/+ is *Ass1* (Table S2), for which ectopic expression is mostly driven by Nan-EKLF binding at the *Ass1* promoter, and this correlates with higher CBP, H3K27ac, and chromatin accessibility (Figure 4H, dotted box). Nan-EKLF also binds to intronic regions of *Ass1*, leading to higher CBP deposition, but pronounced effects are seen only at the TSS (Figure 4H).

In contrast, in *Egln3*, ectopic Nan-EKLF binding to the promoter and distal intronic enhancer correlates with Nan-specific CBP, ATAC, and H3K27ac deposition (Figure S6B). Thus, EKLF binding to distal enhancers can lead to active transcription from a canonical TSS and points to the involvement of long-range chromatin interactions that likely mediate this process. Other ectopic genes such as *Serpine1* and *Fam159b* have ectopic Nan-EKLF binding at the promoter, correlating with ectopic ATAC and CBP peaks at the TSS and

correspondingly higher H3K27ac occupancy along the gene in Nan/+ (Figures S6C and S6D), indicating that for *Fam159b* and *Serpine1*, ectopic expression in Nan/+ is driven by the promoter.

Ectopic Nan-EKLF binding and CBP-mediated pioneering and enhancer activation regulates RNA polymerase II pausing

When we examined paused RNA Pol II at the TSS of ectopic or overexpressed targets, we were surprised to notice that a subset of genes had high levels of paused RNA Pol II even in WT, despite higher elongating RNA Pol II only in Nan/+ (i.e., S5p-hi S2p-lo vs. S5p-hi S2p-hi) (Figure 5A, red line). When we compared the TSS and gene body occupancies of Pol II Ser5p and Ser2p in WT and Nan/+, we also found that the distribution of Ser5p at the TSS does not change in Nan/+, while Ser2p at the gene bodies increases significantly (Figure 5B). We calculated the difference between WT and Nan/+ in paused Pol II near the TSS and found that, indeed, about half of the ectopic or overexpressed genes had similar levels of paused Pol II at their TSS in Nan/+ compared with WT (Table S5). To explore this mechanism further, we analyzed these ectopic or overexpressed gene sets separately and denoted them by their RNA Pol II pausing pattern in WT as high pausing index (S5-hi S2-lo, “hi-PI”) and low pausing index (S5-lo S2-lo, “lo-PI”) (Table S5). We calculated the pausing index of these categories in WT based on the ratio of Ser5p at the TSS and Ser5p at the gene bodies and found that the hi-PI genes indeed have a higher pausing index in WT than lo-PI genes (Figure 5C). This also becomes apparent when we plot the difference in paused and elongating Pol II in Nan/+ over WT by means of a quadrant plot showing that the correlation is lower for the hi-PI genes (Figure 5D, left) and much higher for the lo-PI genes (Figure 5D, right). This suggests that this subset of genes already has paused RNA Pol II in WT and yet is not highly expressed. Their ectopic expression or overexpression in Nan/+ is probably driven by ectopic binding of Nan-EKLF at promoters or distal enhancers.

The RNA expression levels for genes in both categories are significantly higher in Nan/+, as expected (Figure 5E). However, when we quantify and plot the distribution of RNA Pol II Ser5p occupancy around the TSS (−30, +100) in hi-PI genes, we find no significant difference (Figure 5F). Similarly for the same TSS regions, the EKLF, CBP, and H3K27ac levels are also similar in WT and Nan/+ (Figure 5F). This is also apparent when we analyze the ChIP-seq enrichment profile ±3 kb around the TSS (Figure 5F). This strongly indicates that there is EKLF-driven pioneering mediated by CBP at the TSS of hi-PI genes in WT leading to RNA Pol II recruitment and pausing, but no or low expression (Figure 5E). In contrast, lo-PI genes have higher paused Pol II in Nan/+ compared with WT, and this correlates with ectopic Nan-EKLF binding leading to CBP recruitment and H3K27 acetylation at the TSS in Nan/+ (Figure 5G). This suggests that the lo-PI genes are regulated by a canonical mechanism of RNA Pol II recruitment, pausing, and subsequent pause-release driven by Nan-EKLF-mediated pioneering.

Earlier studies have shown that certain genes involved in stress-response or developmental pathways may be regulated by modulating the release of TSS-proximal paused RNA Pol II.[63-65,73,74] These genes usually have a high pausing index or traveling ratio under normal conditions, and RNA Pol II is released when the cell experiences stress, or

differentiation signals, in the latter case. Thus, we asked whether the hi-PI and lo-PI genes were functionally different, which would explain their different modes of regulation. GO analysis of each gene set showed that the hi-PI genes were involved primarily in development and differentiation as well as in stress-responsive transcription (Figure 6A), with some overlapping genes in both functional categories, such as *Nck2* and *Mt3* (Figure 6A). Representative loci from the hi-PI category (Figures 6B and 6C), all having at least one common function of cytoskeleton reorganization,[75,76] demonstrate the pattern of transcription factor and co-factor occupancy shown in Figure 5F. At the TSS there are comparable levels of EKLF, CBP, H3K27ac, and RNA Pol II Ser5p (paused) in both WT and Nan/+ (Figures 6B and 6C). However, certain genes show a variation of this pattern with an ectopic binding site for EKLF in an intron that is not annotated as an alternate TSS, and interestingly, there can be paused RNA Pol II at this site in both WT and Nan/+ (Figures 6D and S7A), or there is higher paused RNA Pol II specifically in Nan/+ (Figures S7B and S7C). In both cases, there is higher CBP at the intronic regions, although H3K27ac levels are mostly similar in WT and Nan/+ (Figures 6B-6D and S7A-S7C). This suggests that ectopic Nan-EKLF binding can mediate ectopic RNA Pol II pausing through CBP-mediated pioneering. Finally, *Hmox1* shows a distinct pattern of regulation with respect to RNA Pol II pausing, in which there are similar amounts of EKLF at the TSS and site 1 distal intergenic enhancer (Figure S7D). In Nan/+ the ectopic Nan-EKLF peak at the site 2 distal intergenic enhancer correlates with higher pausing at sites 1 and 2, but pausing at the TSS remains the same (Figure S7D). This suggests that ectopic Nan-EKLF binding at distal enhancers can influence RNA Pol II pausing upstream and downstream of its binding site, likely mediated by long-range interactions.

Lo-PI genes upregulated in Nan/+ are mostly involved in membrane-associated functions such as transport, membrane biogenesis, and membrane raft functions (Figure 6E). They show a canonical mechanism of Nan-EKLF-driven ectopic expression where there is Nan-specific recruitment of EKLF and CBP at the TSS, leading to higher H3K27ac levels and higher paused and elongating RNA Pol II (Figures 6F, 6G, and S8A-S8C). However, there a few striking exceptions to this pattern, such as those seen at *IL4rα* and *Psap* (Figures 6H and S8D). Here, there are comparable levels of EKLF at the promoter in both WT and Nan/+, but ectopic Nan-EKLF binding at a distal intron leads to higher pausing specifically in Nan/+ at that intronic site as well as the TSS (Figures 6H and S8D), and this pattern is similar to *Crtc1*, *Pacs2*, and *Mob2* of the hi-PI gene set (Figures S7A-S7C). CBP levels at the TSS or intronic regions are also higher in Nan/+ (Figures 6H and S8D), indicating cross talk between the intron-bound Nan-EKLF and the canonical TSS of *IL4rα* and *Psap*. Together, these data strongly indicate that Nan-EKLF (and by extension EKLF/Klf1) can mediate RNA Pol II pausing at its binding sites by CBP-mediated pioneering.

Ectopic Nan-EKLF binding to distal enhancers of high-pausing-index genes drives ectopic expression or overexpression by releasing paused RNA polymerase II at the TSS

The hi-PI gene category is intriguing due to the presence of paused RNA Pol II in WT but low or no expression (Figures 5E and 5F). At these genes, paused RNA Pol II is released specifically in Nan/+, as seen from the higher levels of RNA Pol II Ser2p along the gene and accumulating closer to the transcription end site (TES) (Figures 7A and 7B).

We find correspondingly similar levels of EKLF, CBP, and H3K27ac at the TSS in WT and Nan/+, but higher EKLF and CBP at the gene body specifically in Nan/+ (Figure 7A). EKLF at gene bodies of hi-PI genes is likely Nan-EKLF bound to intronic enhancers (Figure 1D) as evidenced by higher EKLF and CBP at intronic enhancers of hi-PI genes only in Nan/+ (Figure 7B). Surprisingly, H3K27ac levels at the intronic regions of hi-PI genes are similar (Figure 7B). This could suggest that EKLF and CBP drive pause-release in Nan/+ independent of H3K27 acetylation and points to additional mechanisms of CBP action. Thus, we propose one model for Nan-EKLF-driven gene activation in Nan/+ where in WT, EKLF and CBP bound to the promoters/TSS drive promoter-proximal pausing of RNA Pol II but no or low elongation (Figure 7C). In Nan/+, ectopic Nan-EKLF binding to distal enhancers recruits CBP, and paused RNA Pol II is released independent of H3K27 acetylation (Figure 7C).

In contrast, lo-PI genes are regulated by a canonical mechanism where there is ectopic Nan-EKLF binding to promoters, correlating with higher CBP, H3K27ac, and RNA Pol II pausing at the TSS (Figures 7D and 5G). Further, significantly higher elongating RNA Pol II Ser2p at the gene body in Nan/+ correlates with higher EKLF, CBP, and H3K27ac (Figure 7E). There is also higher chromatin accessibility at both the TSS and the specific intronic enhancers of the lo-PI genes in Nan/+ (Figure 7F). This points to a model of gene activation where ectopic Nan-EKLF binding at promoters can initiate pioneering and CBP-mediated H3K27 acetylation leading to recruitment and pausing of RNA Pol II (Figure 7G). Further, ectopic Nan-EKLF at intronic or intergenic distal enhancers can activate RNA Pol II pause-release into productive elongation (Figure 7F). This mechanism of Nan-EKLF-driven transcription activation involves both CBP recruitment and its canonical function of H3K27 acetylation at promoters and enhancers.

DISCUSSION

Our studies have examined the transcriptional control of erythropoiesis from multiple facets. First, we take advantage of the Nan/+ mouse to examine normal and aberrant red cell transcriptional activation *in vivo* during early development. Second, a global analysis and comparison of EKLF binding in WT and Nan/+ cells illuminates the particular effects of the dominant E339D mutant co-existing with WT in the same cell. Third, binding has been correlated with open/closed chromosomal structure, the presence/absence of the CBP co-activator, and acetylation at H3K27 residues. Finally, the surprising effects of Nan-EKLF on RNA Pol II have revealed a previously undescribed non-homogeneous partitioning of pause-release effects at intrinsic and, most excitingly, at ectopic sites of EKLF-E339D binding. Our study has thus delineated the genome-wide impact of EKLF binding on downstream events leading to transcription activation. Together, these studies provide a detailed picture of how a transcription factor acts as a pioneer to establish transcription during embryonic development and how this function becomes subverted in the presence of a seemingly minor (i.e., conservative) mutation.

The presence of Nan-EKLF in the midst of WT EKLF leads to a number of abnormal effects on erythroid gene expression. It remains perplexing that, even in heterozygosity, the E339D mutation can limit binding of WT EKLF to a specific subset of cognate sites, resulting in

a selective and drastic reduction in the expression of erythroid genes. Based on absolute quantification of EKLF protein,[77] there are ~8,000 molecules per erythroid cell, which is similar to the number of occupied binding sites in the genome. The amount of WT protein is then decreased by 50% in the Nan/+ cell, rendering a limiting amount of protein available to bind normal target sites. A further explanation for the negative effect is that binding is dictated by EKLF's interactions with transcription co-factors through posttranslational modifications,[11] such that the WT and mutant EKLF in Nan/+ must compete with limiting amounts of these co-factors.[78,79]

From a structural point of view, it is interesting that the conservative E-to-D change not only limits the recognition of the central nucleotide in the target DNA sequence (as noted previously[17,30]), but also reduces the specificity of the overall 9-bp target recognition sequence from CCM CRC CCN to NCM YKC CYN. Nevertheless, we find that Nan-EKLF DNA binding preferences from our data are consistent with earlier biochemical studies and previous determinations of Nan-EKLF target site preference (CCM NKC CYN).[17,30] The central glutamate is conserved at that site within zinc-finger 2 in all KLF proteins across all species,[80-82] and X-ray studies of the highly similar KLF4 zinc fingers demonstrate that substitution of D alters the cognate recognition.[83,84] Further, the co-occurrence of Gata1 motifs in WT but not Nan-EKLF sites suggests that binding sites of lineage transcription factors have "clustered" during evolution to activate genes involved in a particular developmental process synchronously.[85-87] In addition, our studies demonstrate that the presence of a KLF protein with a change that significantly alters binding is detrimental and not biologically well tolerated within the cell, leading to a significantly altered genetic output even in the company of the WT.

The pioneering activity of EKLF/Klf1 has been inferred from its similarity to KLF4, which has zinc fingers that are >90% identical to those of EKLF[82,88] and is a well-established pioneer factor.[89] KLF family proteins are also implicated in facilitating accessibility of co-localized factors and chromatin assembly at GC-rich regulatory target sites.[87] KLF4 belongs to the "group 1 pioneer factors" that exhibit strong nucleosome binding via their α-helical domains,[90] and due to their strong identity in the C2H2 region, EKLF likely uses the same means for this effect. Although binding to nucleosomes has not been directly demonstrated, EKLF has been shown to bind histone H3 via its zinc-finger region.[44,91] Interactions between pioneer factors and core histones is proposed to be critical for opening chromatin and forming/establishing genetic networks during development.[92] The importance of the P300/CBP co-activator for enhancer/promoter interactions has been well established[52]; EKLF binds both these proteins and by this property alters the surrounding chromatin and activates adjacent gene expression.[44,45]

Quite unexpectedly, we find that WT EKLF is normally bound to genes not expressed in the WT cell (hi-PI genes), which also retain H3K27ac, low levels of CBP, and paused RNA Pol II. In the presence of Nan-EKLF, the availability of a Nan-EKLF consensus at nearby intergenic/intronic regions enables these genes to increase their levels of CBP, resulting in "ectopic" expression by selective release of RNA Pol II. Their observed functions in differentiation, development, and stress responses may well correspond to observations in *Drosophila* wherein genes not likely to be expressed nonetheless have poised RNA Pol II

at the TSS that is ready to respond to external signals.[93,94] This may enable synchronous activation of genes, such as in response to cytokine signals,[95] e.g., during later development in the present case.

A more direct ectopic expression scenario follows from lo-PI genes without poised Pol II/CBP/H3K27ac or expression in WT but active only in Nan/+ and bound by all components. These targets provide the strongest evidence for the pioneering activity of EKLF (Nan), and selective activation in Nan/+ is a true example of mis-expression. These two examples of ectopic expression fall into the general paradigm of lineage transcription factors being pioneers but then requiring co-activator recruitment for ultimate and optimal target gene activation.[96] We suggest that this is because, at a basal level, there is an excess of transcription repressors over activators in the erythroid cell,[79] and this must be overcome by directed and muscular recruitment of co-activators to ensure robust gene activation and differentiation.

Limitations of the study

The reengineered anti-EKLF antibody worked well for the global ChIP analysis, as verified by its specificity for the EKLF cognate recognition motif. However, its limitation in the present study is that it does not distinguish between extant WT and Nan-EKLF proteins in the Nan/+ cell. Tagging each EKLF allele with a singular in-frame antigen will be required for this discriminatory ability to be attained. In this context, the global presence of positive and negative transcriptional elongation effectors (such as P-TEFb, BRD4, DSIF, and NELF) will be of interest to assess and compare at lost/gained EKLF binding sites in WT vs. Nan/+ erythroid cells.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, James J Bieker (james.bieker@mssm.edu).

Materials availability—There are restrictions to the availability of the mouse monoclonal EKLF 7B2a antibody due to the lack of an external centralized repository for its distribution and our need to maintain the stock. We are negotiating with suppliers for commercial availability. In the interim we are glad to share this antibody with compensation by requestor for its shipping, and after completion of a Materials Transfer Agreement.

Data and code availability

- Statement about the data: All NGS data generated from ChIP-Seq and ATAC-Seq is deposited in Gene Expression Omnibus (GEO) and are publicly available as of the date of publication, accession number GSE210779. Gata1 ChIP-Seq data was obtained from GEO accession number GSE32110.
- Statement about the code: The original code is deposited on Github at https://github.com/mkaustav84/biekerlab/blob/

19bbdcf064436e7dd7502c60d8f2db8285829dba/scripts.md and Zenodo with https://doi.org/10.5281/zenodo.7315756.

- General statement: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice—The *Nan* mouse strain ($Klf1^{Nan}/Klf1^{+}$) established previously[17,37,97] was used for all experiments in this study. As hematologic values do not differ between sexes,[17] we did not genotype the sex of the embryos. The adult heterozygous *Nan* mouse is designated *Nan/+* and the strain is fully inbred (C3H/101 background). Embryos are generated from intercrosses between Nan/+ males and +/+ females. Mice were sacrificed at embryonic day E13.5. All animal experiments are approved by the IACUC committee at Mount Sinai.

METHOD DETAILS

Cell isolation—Fetal livers were dissected from WT and Nan/+ E13.5 littermate embryos and mechanically dispersed into single cells by gentle pipetting and straining through a 70μM filter into ice-cold PBS containing 2% FBS.

Flow cytometry—Primary cells from littermate WT and Nan/+ E13.5 fetal livers were stained for FACS with 1:100 dilution of the following antibodies: anti-Ter119-APC and anti-CD44-FITC. Flow Cytometry data was analyzed by FCS Express, and gating was performed based on unstained and single-color compensation controls from the same samples, using the same dyes and within the same experiment. Cell populations were gated for erythroid differentiation stage based on CD44 expression and cell size.[68]

Generation of 7B2a anti-EKLF antibody—One of our early monoclonals (7B2) has proven useful for some studies,[98-100] and we had optimized the ChIP protocol such that we are able to use 7B2 for directed target analysis.[32] However, it still failed when used for ChIP-seq (unpublished). As IgG3 antibodies (the 7B2 isotype) can have issues with solubility and aggregation, we amplified the 7B2 heavy and light chains via PCR of clone 7B2 hybridoma DNA, cloned them into vectors containing mouse IgG2a constant and IgK regions, and isolated the antibody after cotranfection of these two constructs into Epsi293 or HD293F cells, which secreted the antibody into the supernatant (performed in conjunction with GenScript USA). This was purified via protein A, yielding the 7B2a antibody. ELISA testing indicated that affinity/avidity were equivalent to our parental antibody, IP test results were better than with 7B2, and importantly, directed ChIP (tests on binding to ßmaj and E2F2 promoters) gave excellent signal/noise. As a result we used 7B2a for all experiments.

Chromatin immunoprecipitation and sequencing (ChIP-Seq)—Antibodies used for ChIP are shown below:

ChIP experiment	Antibody
EKLF/Klf1	EKLF 7B2a mouse monoclonal antibody

ChIP experiment	Antibody
H3K27ac	Anti-Histone H3 (acetyl K27) antibody - ChIP Grade
CBP	CBP (D6C5) Rabbit mAb
RNA Polymerase II Ser5p	Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody
RNA Polymerase II Ser2p	Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody
Rabbit IgG control	Normal rabbit IgG
Mouse IgG control	Normal mouse IgG

For each ChIP experiment, cells from one WT or one Nan/+ E13.5 fetal liver were used as a biological replicate, and each experiment constituted 3 biological replicates except for RNA Pol II Ser5p and Ser2p ChIPs which had two biological replicates each. Where possible, littermates were used for replicates within each ChIP experiment. 3 million cells from each fetal liver were used for RNA Pol II and H3K27ac ChIPs, and for transcription factor ChIPs we used 7.5 million cells. Primary fetal liver cells were cross-linked in 1% formaldehyde and cell pellets were stored at −80°C. Frozen cells were thawed and resuspended in MNase lysis buffer (0.2% NP-40, 10mM NaCl, 10mM Tris pH 8.0) and kept on ice for 15 mins. Cells were pelleted and resuspended in MNase lysis buffer with 1mM $CaCl_2$, 40 units Micrococcal nuclease (MNase) and incubated at 37°C for 15 min. The MNase reaction was stopped by adding EDTA to a final concentration of 2mM and nuclei were pelleted by centrifugation. The pellet was resuspended in 1X RIPA buffer with 1X Protease Inhibitor Cocktail and sonicated using a cup-horn sonicator (Ultrasonic VC505) at 80 Amplitude for 50 cycles with pulses of 40s on and 50s off. The lysed and fragmented chromatin was cleared by centrifugation and 5%v/v was stored at −80°C as input. 5–10μg of antibody (depending on the titer) was added to the lysate along with a control IgG (for enrichment estimation) and antibody-lysate mixture was incubated at 4°C for 16h (overnight). The lysate-antibody mixture was then incubated with Protein-A or Protein-G Dynabeads, depending on the antibody serotype, for 2h at 4°C. Immunoprecipitated complexes on Dynabeads were then washed three times in 1X RIPA with protease inhibitors, and three times in wash buffer containing 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris pH 8.0, 500mM NaCl, and protease inhibitors. This was followed by three washes in wash buffer containing 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris pH 8.0, and 250 mM LiCl. Finally, beads were washed three times in 1X TE buffer containing 10mM Tris pH 8.0 and 1mM EDTA. Elution was performed by adding elution buffer containing 10mM Tris pH 8.0, 1mM EDTA, 200mM NaCl and 1% SDS and incubating at 65°C in a Thermomixer (Eppendorf Cat# 5382000023) with gentle mixing for 20 mins. Eluate containing ChIP DNA and input chromatin was removed and incubated at 65°C with gentle mixing overnight for reverse crosslinking. DNA was treated with 0.1μg/μL Proteinase K for 1h at 55°C followed by purification using Phenol:chloroform extraction and ethanol precipitation. ChIP efficiency was verified by qPCR by comparing enrichment over input for various ChIP experiments with enrichments obtained with IgG control. ChIP-Seq libraries were made using a Neb Next DNA Ultra II DNA library preparation kit following manufacturer's instructions. Libraries were quantified using Qubit fluorometer (Invitrogen) and library quality was assessed using a high-sensitivity DNA kit on Agilent Bioanalyzer, followed by Illumina next generation sequencing.

Assay for transposase accessible chromatin (ATAC-Seq)—ATAC-Seq was performed using 100,000 cells from WT and Nan/+ E13.5 fetal livers in two biological replicates each from independent Nan/+ mice. The protocol was only slightly modified from.[101,102] Briefly, cells were lysed in 10mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAL CA-630. Transposition mix included Illumina 2x Tagment DNA Buffer and Tn5 Transposase. AmpureXP beads were used for DNA purification and SPRIselect beads for size selection. PCR amplification utilized NEB Next High Fidelity 2x PCR master mix and Nextera primers. In order to reduce GC and size bias, and to avoid saturation, the extent of PCR after five cycles was determined using 10% of the initial PCR in a side reaction that included SYBR Green 1. The suitable number of cycles required for the second round of PCR for the final library was then determined by adding 3 additional cycles to the number of cycles corresponding to 25% maximum intensity. Libraries were quantified with a Qubit fluorometer (Invitrogen) and library quality was assessed using a high-sensitivity DNA kit on an Agilent Bioanalyzer. The ATAC libraries were sequenced on a HiSeq platform set for 100nt/single reads. ATAC-seq oligos were as described.[101,102]

QUANTIFICATION AND STATISTICAL ANALYSIS

The analysis, software, and quantification methodology that are specific to NGS experiments are included under the relevant subsections below. Information regarding replicate numbers and error bars is provided in figure legends. If degrees of significance as p *value* is depicted in a figure, details regarding the statistical test used are provided in the figure legend.

Bioinformatics and computational methods

RNA-seq: Reads were mapped using Salmon to the mouse transcriptome version Ensembl GRCm38. Raw counts were imported using tximport package and count normalization and differential gene expression analysis was performed using DESeq2.[103] Subsequent analyses and plotting were done using Python Pandas, Matplotlib, and Seaborn libraries. Genes were designated as differentially expressed based on a *p-adjusted* threshold of 0.01. This *p-adjusted* value is based on the inherent Benjamini-Hochberg correction for FDRs in the DESeq2 package.

ChIP-seq and ATAC-Seq—Reads were aligned to the mouse genome mm10 version using Bowtie2[104] and sorted bam files were generated using samtools.[105] For IGV visualization, input-normalized bigwig files were generated using Deeptools using the RPKM method[106] and then sorted and combined using bedtools (Encode). ChIP enrichment profiles and heatmaps were also generated using Deeptools computeMatrix, plotHeatmap and plotProfile functions using the resulting input-normalized sorted, and combined bigwig files. Peak calling, motif analysis and enrichment calculation was performed using Homer[107] using the findPeaks.pl, getDifferentialPeaksReplicates.pl, find-Motifs.pl, and annotatePeaks.pl programs. All subsequent analysis was performed using Python Pandas, Numpy, MS Excel and plots were generated using the Python Seaborn library. All the code and options used for each of the above program is deposited at Github: https://github.com/mkaustav84/biekerlab/blob/19bbdcf064436e7dd7502c60d8f2db8285829dba/scripts.md and Zenodo: https://doi.org/10.5281/zenodo.7315756.

Gata1 E12.5 Fetal Liver ChIP-Seq data was obtained from[69] and analyzed using the same methods as above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by a Black Family Stem Cell Institute postdoctoral award to K.M. and by NIH grant R01 DK046865 to J.J.B. We thank Mohan Dangeti for initial RNA Pol II experiments, Chrysoula Deligianni for ATAC-seq, Zelda Salfati at the MSSM Genomics Core for advice on NGS library optimization, Tom Moran and J. Andrew Duty of the MSSM Hybridoma Core Facility for establishing the 7B2a/IgG2a clones, and the MSSM Flow Cytometry Core facility. We thank Cheryl A. Keller and Ross Hardison for advice on the ATAC-seq protocol. We acknowledge Nithya Gnanapragasam, Sanjana Pillay, and Antanas Planutis for ongoing discussions; Sree Chinta for RNA analysis; and Li Xue for mouse husbandry.

REFERENCES

1. Pollak NM, Hoffman M, Goldberg IJ, and Drosatos K (2018). Kruppel-like factors: crippling and un-crippling metabolic pathways. JACC. Basic Transl. Sci 3, 132–156. 10.1016/j.jacbts.2017.09.001. [PubMed: 29876529]
2. McConnell BB, and Yang VW (2010). Mammalian Kruppel-like factors in health and diseases. Physiol. Rev 90, 1337–1381. 10.1152/physrev.00058.2009. [PubMed: 20959618]
3. Chang E, Nayak L, and Jain MK (2017). Kruppel-like factors in endothelial cell biology. Curr. Opin. Hematol 24, 224–229. 10.1097/MOH.0000000000000337. [PubMed: 28306664]
4. Takahashi K, and Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676. [PubMed: 16904174]
5. Southwood CM, Downs KM, and Bieker JJ (1996). Erythroid Kruppel-like Factor (EKLF) exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. Dev. Dyn 206, 248–259. [PubMed: 8896981]
6. Funnell APW, Mak KS, Twine NA, Pelka GJ, Norton LJ, Radziewic T, Power M, Wilkins MR, Bell-Anderson KS, Fraser ST, et al. (2013). Generation of mice deficient in both KLF3/BKLF and KLF8 reveals a genetic interaction and a role for these factors in embryonic globin gene silencing. Mol. Cell Biol 33, 2976–2987. [PubMed: 23716600]
7. Vinjamur DS, Wade KJ, Mohamad SF, Haar JL, Sawyer ST, and Lloyd JA (2014). Kruppel-like transcription factors KLF1 and KLF2 have unique and coordinate roles in regulating embryonic erythroid precursor maturation. Haematologica 99, 1565–1573. 10.3324/haematol.2014.104943. [PubMed: 25150253]
8. Basu P, Morris PE, Haar JL, Wani MA, Lingrel JB, Gaensler KML, and Lloyd JA (2005). KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. Blood 106, 2566–2571. 10.1182/blood-2005-02-0674. [PubMed: 15947087]
9. Gnanapragasam MN, and Bieker JJ (2017). Orchestration of late events in erythropoiesis by KLF1/EKLF. Curr. Opin. Hematol 24, 183–190. 10.1097/MOH.0000000000000327. [PubMed: 28157724]
10. Siatecka M, and Bieker JJ (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood 118, 2044–2054. 10.1182/blood-2011-03-331371. [PubMed: 21613252]
11. Yien YY, and Bieker JJ (2013). EKLF/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination. Mol. Cell Biol 33, 4–13. [PubMed: 23090966]
12. Mukherjee K, and Bieker JJ (2021). Transcriptional control of gene expression and the heterogeneous cellular identity of erythroblastic island macrophages. Front. Genet 12, 756028. 10.3389/fgene.2021.756028. [PubMed: 34880902]
13. Tallack MR, and Perkins AC (2010). KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. IUBMB Life 62, 886–890. 10.1002/iub.404. [PubMed: 21190291]

14. Feng WC, Southwood CM, and Bieker JJ (1994). Analyses of β-thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. J. Biol. Chem 269, 1493–1500. [PubMed: 8288615]
15. Miller IJ, and Bieker JJ (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Mol. Cell Biol 13, 2776–2786. [PubMed: 7682653]
16. Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, and Perkins AC (2010).A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res. 20, 1052–1063. 10.1101/gr.106575. [PubMed: 20508144]
17. Siatecka M, Sahr KE, Andersen SG, Mezei M, Bieker JJ, and Peters LL (2010). Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor. Proc. Natl. Acad. Sci. USA 107, 15151–15156. [PubMed: 20696915]
18. Heruth DP, Hawkins T, Logsdon DP, Gibson MI, Sokolovsky IV, Nsumu NN, Major SL, Fegley B, Woods GM, Lewing KB, et al. (2010). Mutation in erythroid specific transcription factor KLF1 causes Hereditary Spherocytosis in the Nan hemolytic anemia mouse model. Genomics 96, 303–307. 10.1016/j.ygeno.2010.07.009. [PubMed: 20691777]
19. Sorolla A, Tallack MR, Oey H, Harten SK, Daxinger LC, Magor GW, Combes AN, Ilsley M, Whitelaw E, and Perkins AC (2015). Identification of novel hypomorphic and null mutations in Klf1 derived from a genetic screen for modifiers of alpha-globin transgene variegation. Genomics 105, 116–122. 10.1016/j.ygeno.2014.09.013. [PubMed: 25451176]
20. Korporaal A, Gillemans N, Heshusius S, Cantu I, van den Akker E, van Dijk TB, von Lindern M, and Philipsen S (2021). Hemoglobin switching in mice carrying the Klf1(Nan) variant. Haematologica 106, 464–473. 10.3324/haematol.2019.239830. [PubMed: 32467144]
21. Perkins A, Xu X, Higgs DR, Patrinos GP, Arnaud L, Bieker JJ, and Philipsen S; KLF1 Consensus Workgroup (2016). Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. Blood 127, 1856–1862. 10.1182/blood-2016-01-694331. [PubMed: 26903544]
22. Waye JS, and Eng B (2015). Kruppel-like factor 1: hematologic phenotypes associated with KLF1 gene mutations. Int. J. Lab. Hematol 37 (Suppl 7), 78–84. 10.1111/ijlh.12356. [PubMed: 25976964]
23. Viprakasit V, Ekwattanakit S, Riolueang S, Chalaow N, Fisher C, Lower K, Kanno H, Tachavanich K, Bejrachandra S, Saipin J, et al. (2014). Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood 123, 1586–1595. 10.1182/blood-2013-09-526087. [PubMed: 24443441]
24. Liu D, Zhang X, Yu L, Cai R, Ma X, Zheng C, Zhou Y, Liu Q, Wei X, Lin L, et al. (2014). KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. Blood 124, 803–811. [PubMed: 24829204]
25. Nuez B, Michalovich D, Bygrave A, Ploemacher R, and Grosveld F (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature (London) 375, 316–318. [PubMed: 7753194]
26. Perkins AC, Sharpe AH, and Orkin SH (1995). Lethal ß-thalassemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature (London) 375, 318–322. [PubMed: 7753195]
27. Tallack MR, Magor GW, Dartigues B, Sun L, Huang S, Fittock JM, Fry SV, Glazov EA, Bailey TL, and Perkins AC (2012). Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. Genome Res. 22, 2385–2398. [PubMed: 22835905]
28. Hodge D, Coghill E, Keys J, Maguire T, Hartmann B, McDowall A, Weiss M, Grimmond S, and Perkins A (2006).A global role for EKLF in definitive and primitive erythropoiesis. Blood 107, 3359–3370. [PubMed: 16380451]
29. Cantú I, van deWerken HJG, Gillemans N, Stadhouders R, Heshusius S, Maas A, Esteghamat F, Ozgur Z, van IJcken WFJ, Grosveld F, et al. (2019). The mouse KLF1 Nan variant impairs nuclear condensation and erythroid maturation. PLoS One 14, e0208659. 10.1371/journal.pone.0208659. [PubMed: 30921348]

30. Gillinder KR, Ilsley MD, Nébor D, Sachidanandam R, Lajoie M, Magor GW, Tallack MR, Bailey T, Landsberg MJ, Mackay JP, et al. (2017). Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability. Nucleic Acids Res. 45, 1130–1143. [PubMed: 28180284]

31. Nébor D, Graber JH, Ciciotte SL, Robledo RF, Papoin J, Hartman E, Gillinder KR, Perkins AC, Bieker JJ, Blanc L, and Peters LL (2018). Mutant KLF1 in adult anemic nan mice leads to profound transcriptome changes and disordered erythropoiesis. Sci. Rep 8, 12793. 10.1038/s41598-018-30839-2. [PubMed: 30143664]

32. Planutis A, Xue L, Trainor CD, Dangeti M, Gillinder K, Siatecka M, Nebor D, Peters LL, Perkins AC, and Bieker JJ (2017). Neomorphic effects of the neonatal anemia (Nan-Eklf) mutation contribute to deficits throughout development. Development 144, 430–440. 10.1242/dev.145656. [PubMed: 28143845]

33. Varricchio L, Planutis A, Manwani D, Jaffray J, Mitchell WB, Migliaccio AR, and Bieker JJ (2019). Genetic disarray follows mutant KLF1-E325K expression in a congenital dyserythropoietic anemia patient. Haematologica 104, 2372–2380. 10.3324/haematol.2018.209858. [PubMed: 30872368]

34. Kohara H, Utsugisawa T, Sakamoto C, Hirose L, Ogawa Y, Ogura H, Sugawara A, Liao J, Aoki T, Iwasaki T, et al. (2019). KLF1 mutation E325K induces cell cycle arrest in erythroid cells differentiated from congenital dyserythropoietic anemia patient-specific induced pluripotent stem cells. Exp. Hematol 73, 25–37.e8. 10.1016/j.exphem.2019.03.001. [PubMed: 30876823]

35. Kulczynska-Figurny K, Bieker JJ, and Siatecka M (2020). Severe anemia caused by dominant mutations in Kruppel-like factor 1 (KLF1). Mutat. Res 786, 108336. 10.1016/j.mrrev.2020.108336.

36. White RA, Sokolovsky IV, Britt MI, Nsumu NN, Logsdon DP, McNulty SG, Wilmes LA, Brewer BP, Wirtz E, Joyce HR, et al. (2009). Hematologic characterization and chromosomal localization of the novel dominantly inherited mouse hemolytic anemia, neonatal anemia (Nan). Blood Cells Mol. Dis 43, 141–148. [PubMed: 19409822]

37. Lyon MF (1983). Dominant haemolytic anaemia. Mouse News Letter 68, 68.

38. Arnaud L, Saison C, Helias V, Lucien N, Steschenko D, Giarratana MC, Prehu C, Foliguet B, Montout L, de Brevern AG, et al. (2010). A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. Am. J. Hum. Genet 87, 721–727. 10.1016/j.ajhg.2010.10.010. [PubMed: 21055716]

39. Jaffray JA, Mitchell WB, Gnanapragasam MN, Seshan SV, Guo X, Westhoff CM, Bieker JJ, and Manwani D (2013). Erythroid transcription factor EKLF/KLF1 mutation causing congenital dyserythropoietic anemia type IV in a patient of Taiwanese origin: review of all reported cases and development of a clinical diagnostic paradigm. Blood Cells Mol. Dis 51, 71–75. 10.1016/j.bcmd.2013.02.006. [PubMed: 23522491]

40. Singleton BK, Lau W, Fairweather VSS, Burton NM, Wilson MC, Parsons SF, Richardson BM, Trakarnsanga K, Brady RL, Anstee DJ, and Frayne J (2011). Mutations in the second zinc finger of human EKLF reduce promoter affinity but give rise to benign and disease phenotypes. Blood 118, 3137–3145. 10.1182/blood-2011-04-349985. [PubMed: 21778342]

41. Ravindranath Y, Johnson RM, Goyette G, Buck S, Gadgeel M, and Gallagher PG (2018). KLF1 E325K-associated congenital dyserythropoietic anemia type IV: insights into the variable clinical severity. J. Pediatr. Hematol. Oncol 40, e405–e409. 10.1097/MPH.0000000000001056. [PubMed: 29300242]

42. Kulczynska K, Bieker JJ, and Siatecka M (2020). A kruppel-like factor 1 (KLF1) mutation associated with severe congenital dyserythropoietic anemia alters its DNA-binding specificity. Mol. Cell Biol 40, 004444–e519. 10.1128/MCB.00444-19.

43. Ilsley MD, Huang S, Magor GW, Landsberg MJ, Gillinder KR, and Perkins AC (2019). Corrupted DNA-binding specificity and ectopic transcription underpin dominant neomorphic mutations in KLF/SP transcription factors. BMC Genom. 20, 417. 10.1186/s12864-019-5805-z.

44. Sengupta T, Chen K, Milot E, and Bieker JJ (2008). Acetylation of EKLF is essential for epigenetic modification and transcriptional activation of the beta-globin locus. Mol. Cell Biol 28, 6160–6170. [PubMed: 18710946]

45. Zhang W, and Bieker JJ (1998). Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. Proc. Natl. Acad. Sci. USA 95, 9855–9860. [PubMed: 9707565]
46. Zhang W, Kadam S, Emerson BM, and Bieker JJ (2001). Site-specific acetylation by p300 or CREB binding protein regulates erythroid Kruppel-like factor transcriptional activity via its interaction with the SWI-SNF complex. Mol. Cell Biol 21, 2413–2422. [PubMed: 11259590]
47. Mas C, Lussier-Price M, Soni S, Morse T, Arseneault G, Di Lello P, Lafrance-Vanasse J, Bieker JJ, and Omichinski JG (2011). Structural and functional characterization of an atypical activation domain in erythroid Kruppel-like factor (EKLF). Proc. Natl. Acad. Sci. USA 108, 10484–10489. [PubMed: 21670263]
48. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, and Harte PJ (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development 136, 3131–3141. 10.1242/dev.037127. [PubMed: 19700617]
49. Shiama N (1997). The p300/CBP family: integrating signals with transcription factors and chromatin. Trends Cell Biol. 7, 230–236. 10.1016/S0962-8924(97)01048-9. [PubMed: 17708951]
50. Hsu E, Zemke NR, and Berk AJ (2021). Promoter-specific changes in initiation, elongation, and homeostasis of histone H3 acetylation during CBP/p300 inhibition. Elife 10, e63512. 10.7554/eLife.63512. [PubMed: 33704060]
51. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA 107, 21931–21936. 10.1073/pnas.1016071107. [PubMed: 21106759]
52. Narita T, Ito S, Higashijima Y, Chu WK, Neumann K, Walter J, Satpathy S, Liebner T, Hamilton WB, Maskey E, et al. (2021). Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Mol. Cell 81, 2166–2182.e6. 10.1016/j.molcel.2021.03.008. [PubMed: 33765415]
53. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, and Wysocka J (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature (London) 470, 279–283. 10.1038/nature09692. [PubMed: 21160473]
54. Armstrong JA, Bieker JJ, and Emerson BM (1998). A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. Cell 95, 93–104. [PubMed: 9778250]
55. Kadam S, McAlpine GS, Phelan ML, Kingston RE, Jones KA, and Emerson BM (2000). Functional selectivity of recombinant mammalian SWI/SNF subunits. Genes Dev. 14, 2441–2451. [PubMed: 11018012]
56. Bottardi S, Ross J, Pierre-Charles N, Blank V, and Milot E (2006). Lineage-specific activators affect beta-globin locus chromatin in multi-potent hematopoietic progenitors. EMBO J. 25, 3586–3595. [PubMed: 16858401]
57. Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, and de Laat W (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev. 18, 2485–2490. [PubMed: 15489291]
58. Haberle V, and Stark A (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol 19, 621–637. 10.1038/s41580-018-0028-8. [PubMed: 29946135]
59. Kwak H, and Lis JT (2013). Control of transcriptional elongation.Annu. Rev. Genet 47, 483–508. 10.1146/annurev-genet-110711-155440. [PubMed: 24050178]
60. Jonkers I, and Lis JT (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol 16, 167–177. 10.1038/nrm3953. [PubMed: 25693130]
61. Lee K, and Blobel GA (2016). Chromatin architecture underpinning transcription elongation. Nucleus 7, 1–8. 10.1080/19491034.2016.1200770.
62. Chen FX, Smith ER, and Shilatifard A (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol 19, 464–478. 10.1038/s41580-018-0010-5. [PubMed: 29740129]

63. Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, Gaudenz K, De Kumar B, Krumlauf R, and Zeitlinger J (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. 2, 1670–1683. 10.1016/j.celrep.2012.11.024. [PubMed: 23260668]
64. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, and Young RA (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet 39, 1512–1516. 10.1038/ng.2007.26. [PubMed: 17994019]
65. Murphy ZC, Murphy K, Myers J, Getman M, Couch T, Schulz VP, Lezon-Geyda K, Palumbo C, Yan H, Mohandas N, et al. (2021). Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. Blood 138, 1740–1756. 10.1182/blood.2020009903. [PubMed: 34075391]
66. Komarnitsky P, Cho EJ, and Buratowski S (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 14, 2452–2460. 10.1101/gad.824700. [PubMed: 11018013]
67. Saunders A, Core LJ, and Lis JT (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol 7, 557–567. 10.1038/nrm1981. [PubMed: 16936696]
68. Chen K, Liu J, Heck S, Chasis JA, An X, and Mohandas N (2009). Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. Proc. Natl. Acad. Sci. USA 106, 17413–17418. [PubMed: 19805084]
69. Papadopoulos GL, Karkoulia E, Tsamardinos I, Porcher C, Ragoussis J, Bungert J, and Strouboulis J (2013). GATA-1 genome-wide occupancy associates with distinct epigenetic profiles in mouse fetal liver erythropoiesis. Nucleic Acids Res. 41, 4938–4948. 10.1093/nar/gkt167. [PubMed: 23519611]
70. Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SYR, and Ge K (2011). Distinct roles of GCN5/ PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. 30, 249–262. 10.1038/emboj.2010.318. [PubMed: 21131905]
71. Blobel GA, Nakajima T, Eckner R, Montminy M, and Orkin SH (1998). CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. Proc. Natl. Acad. Sci. USA 95, 2061–2066. [PubMed: 9482838]
72. Patro R, Duggal G, Love MI, Irizarry RA, and Kingsford C (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. 10.1038/nmeth.4197. [PubMed: 28263959]
73. Rougvie AE, and Lis JT (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell 54, 795–804. [PubMed: 3136931]
74. Gilmour DS, and Lis JT (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol. Cell Biol 6, 3984–3989. 10.1128/mcb.6.11.3984-3989.1986. [PubMed: 3099167]
75. Alfaidi M, Scott ML, and Orr AW (2021). Sinner or saint?: Nck adaptor proteins in vascular biology. Front. Cell Dev. Biol 9, 688388. 10.3389/fcell.2021.688388. [PubMed: 34124074]
76. Sepulveda JL, Gkretsi V, and Wu C (2005). Assembly and signaling of adhesion complexes. Curr. Top. Dev. Biol 68, 183–225. 10.1016/S0070-2153(05)68007-6. [PubMed: 16125000]
77. Gautier EF, Leduc M, Ladli M, Schulz VP, Lefèvre C, Boussaid I, Fontenay M, Lacombe C, Verdier F, Guillonneau F, et al. (2020). Comprehensive proteomic analysis of murine terminal erythroid differentiation. Blood Adv. 4, 1464–1477. 10.1182/bloodadvances.2020001652. [PubMed: 32282884]
78. Blobel GA (2002). CBP and p300: versatile coregulators with important roles in hematopoietic gene expression. J. Leukoc. Biol 71, 545–556. [PubMed: 11927640]
79. Gillespie MA, Palii CG, Sanchez-Taltavull D, Shannon P, Longabaugh WJR, Downes DJ, Sivaraman K, Espinoza HM, Hughes JR, Price ND, et al. (2020). Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis. Mol. Cell 78, 960–974.e11. 10.1016/j.molcel.2020.03.031. [PubMed: 32330456]

80. Presnell JS, Schnitzler CE, and Browne WE (2015). KLF/SP transcription factor family evolution: expansion, diversification, and innovation in eukaryotes. Genome Biol. Evol 7, 2289–2309. 10.1093/gbe/evv141. [PubMed: 26232396]

81. Shao M, Ge GZ, Liu WJ, Xiao J, Xia HJ, Fan Y, Zhao F, He BL, and Chen C (2017). Characterization and phylogenetic analysis of Kruppel-like transcription factor (KLF) gene family in tree shrews (Tupaia belangeri chinensis). Oncotarget 8, 16325–16339. 10.18632/oncotarget.13883. [PubMed: 28032601]

82. Chen L, Huang R, Li Y, Li Y, Li Y, Liao L, He L, Zhu Z, and Wang Y (2021). Genome-wide identification, evolution of Kruppel-like factors (klfs) and their expressions during GCRV challenge in grass carp (Ctenopharyngodonidella). Dev. Comp. Immunol 120, 104062. 10.1016/j.dci.2021.104062. [PubMed: 33667530]

83. Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, and Cheng X (2014). Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res. 42, 4859–4867. 10.1093/nar/gku134. [PubMed: 24520114]

84. Hashimoto H, Wang D, Steves AN, Jin P, Blumenthal RM, Zhang X, and Cheng X (2016). Distinctive Klf4 mutants determine preference for DNA methylation status. Nucleic Acids Res. 44, 10177–10185. 10.1093/nar/gkw774. [PubMed: 27596594]

85. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, Sarro R, Rakic P, and Noonan JP (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science 347, 1155–1159. 10.1126/science.1260943. [PubMed: 25745175]

86. Sarropoulos I, Sepp M, Fromel R, Leiss K, Trost N, Leushkin E, Okonechnikov K, Joshi P, Giere P, Kutscher LM, et al. (2021). Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. Science 373, eabg4696. 10.1126/science.abg4696. [PubMed: 34446581]

87. Zhao Y, Vartak SV, Conte A, Wang X, Garcia DA, Stevens E, Kyoung Jung S, Kieffer-Kwon KR, Vian L, Stodola T, et al. (2022). Stripe" transcription factors provide accessibility to co-binding partners in mammalian genomes. Mol. Cell 82, 3398–3411.e11. 10.1016/j.molcel.2022.06.029. [PubMed: 35863348]

88. Bieker JJ (2001). Kruppel-like factors: three fingers in many pies. J. Biol. Chem 276, 34355–34358. [PubMed: 11443140]

89. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, and Zaret KS (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161, 555–568. 10.1016/j.cell.2015.03.017. [PubMed: 25892221]

90. Fernandez Garcia M, Moore CD, Schulz KN, Alberto O, Donague G, Harrison MM, Zhu H, and Zaret KS (2019). Structural features of transcription factors associating with nucleosome binding. Mol. Cell 75, 921–932.e6. 10.1016/j.molcel.2019.06.009. [PubMed: 31303471]

91. Soni S, Pchelintsev N, Adams PD, and Bieker JJ (2014). Transcription factor EKLF (KLF1) recruitment of the histone chaperone HIRA is essential for beta-globin gene expression. Proc. Natl. Acad. Sci. USA 111, 13337–13342. 10.1073/pnas.1405422111. [PubMed: 25197097]

92. Iwafuchi M, Cuesta I, Donahue G, Takenaka N, Osipovich AB, Magnuson MA, Roder H, Seeholzer SH, Santisteban P, and Zaret KS (2020). Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. Nat. Genet 52, 418–427. 10.1038/s41588-020-0591-8. [PubMed: 32203463]

93. Gaertner B, and Zeitlinger J (2014). RNA polymerase II pausing during development. Development 141, 1179–1183. 10.1242/dev.088492. [PubMed: 24595285]

94. Core L, and Adelman K (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. 33, 960–982. 10.1101/gad.325142.119. [PubMed: 31123063]

95. Henriques T, Gilchrist DA, Nechaev S, Bern M, Muse GW, Burkholder A, Fargo DC, and Adelman K (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell 52, 517–528. 10.1016/j.molcel.2013.10.001. [PubMed: 24184211]

96. Field A, and Adelman K (2020). Evaluating enhancer function and transcription. Annu. Rev. Biochem 89, 213–234. 10.1146/annurev-biochem-011420-095916. [PubMed: 32197056]

97. Lyon MF (1986). Position of neonatal anaemia (Nan) on chromosome 8. Mouse News Letter 74, 95.

98. Mansoor A, Mansoor MO, Patel JL, Zhao S, Natkunam Y, and Bieker JJ (2020). KLF1/EKLF expression in acute leukemia is correlated with chromosomal abnormalities. Blood Cells Mol. Dis 83, 102434. 10.1016/j.bcmd.2020.102434. [PubMed: 32311573]

99. Im H, Grass JA, Johnson KD, Kim SI, Boyer ME, Imbalzano AN, Bieker JJ, and Bresnick EH (2005). Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. Proc. Natl. Acad. Sci. USA 102, 17065–17070. [PubMed: 16286657]

100. Siatecka M, Lohmann F, Bao S, and Bieker JJ (2010). EKLF directly activates the p21WAF1/CIP1 gene by proximal promoter and novel intronic regulatory regions during erythroid differentiation. Mol. Cell Biol 30, 2811–2822. [PubMed: 20368355]

101. Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol 109, 21.29.1–21.29.9. 10.1002/0471142727.mb2129s109.

102. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. 10.1038/nmeth.2688. [PubMed: 24097267]

103. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. 10.1186/s13059-014-0550-8. [PubMed: 25516281]

104. Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. 10.1038/nmeth.1923. [PubMed: 22388286]

105. Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. 10.1093/bioinformatics/btp324. [PubMed: 19451168]

106. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165. 10.1093/nar/gkw257. [PubMed: 27079975]

107. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. 10.1016/j.molcel.2010.05.004. [PubMed: 20513432]

Highlights

- EKLF binding *in vivo* in WT compared with Nan/+ erythroid cells has been established
- Heterozygous expression of Nan-EKLF interferes with WT binding at select sites
- Nan-EKLF is a pioneer at ectopic sites of gene activation, altering CBP and H3K27ac
- Nan-EKLF alters RNA Pol II pause-release status at ectopic activation sites

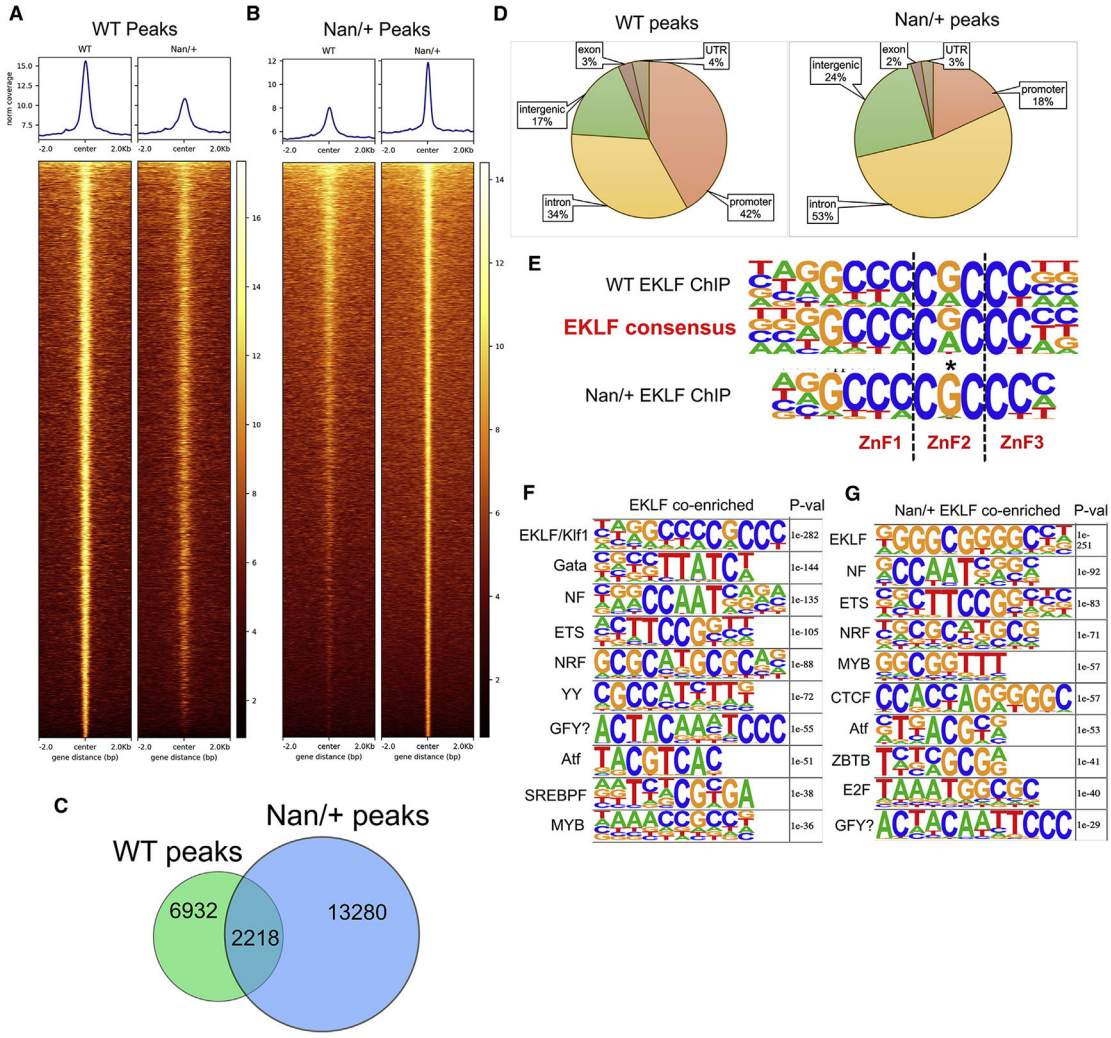


Figure 1. The EKLF/Klf1 Nan mutation alters EKLF DNA binding in E13.5 mouse FL

(A and B) Input-normalized coverage and enrichment heatmaps ±2 kb from the peak center of 7B2a EKLF ChIP-seq peaks in (A) WT and (B) Nan/+.

(C) Venn diagram comparing EKLF peak numbers in WT and Nan/+ depicting overlapping and non-overlapping peaks.

(D) Distribution of genomic locations of EKLF peaks in WT (left) and Nan/+ (right).

(E) Alignment of the motif sequences determined by Homer in WT and Nan/+ EKLF ChIP-seq to the consensus EKLF motif. The asterisk is explained in the text.

(F and G) Top 10 motifs determined by Homer from sequences of EKLF ChIP-seq peaks in (F) WT and (G) Nan/+. See also Figures S2 and S3 and Table S1.

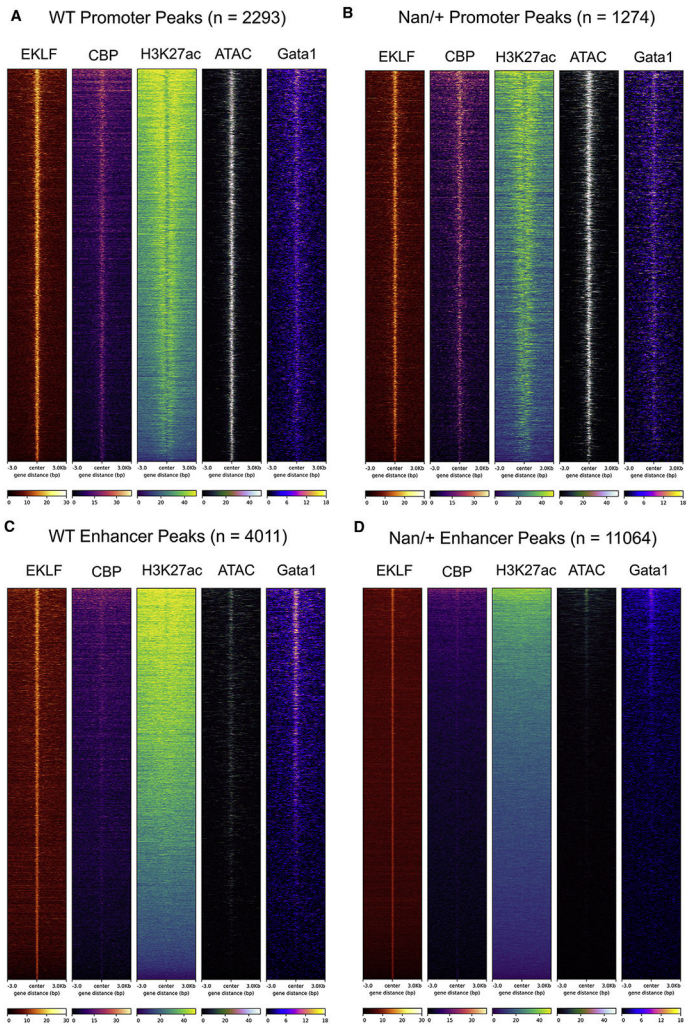


Figure 2. Pioneering by EKLF and Nan-EKLF in the context of EKLF peak location and presence or absence of additional transcription factors and co-factors

(A and B) ChIP-seq heatmaps of EKLF, CBP, H3K27ac, Gata1, and ATAC-seq showing input-normalized enrichment ±3 kb from the center of WT EKLF promoter peaks in (A) WT and (B) Nan/+ FL.

(C and D) Same data as in (A) and (B) from introns and intergenic EKLF peaks in (C) WT and (D) Nan/+. E12.5 Gata1 ChIP-seq data were obtained from Papadopoulos et al.[69] Number of peaks is mentioned above the heatmaps. See also Figure S4.

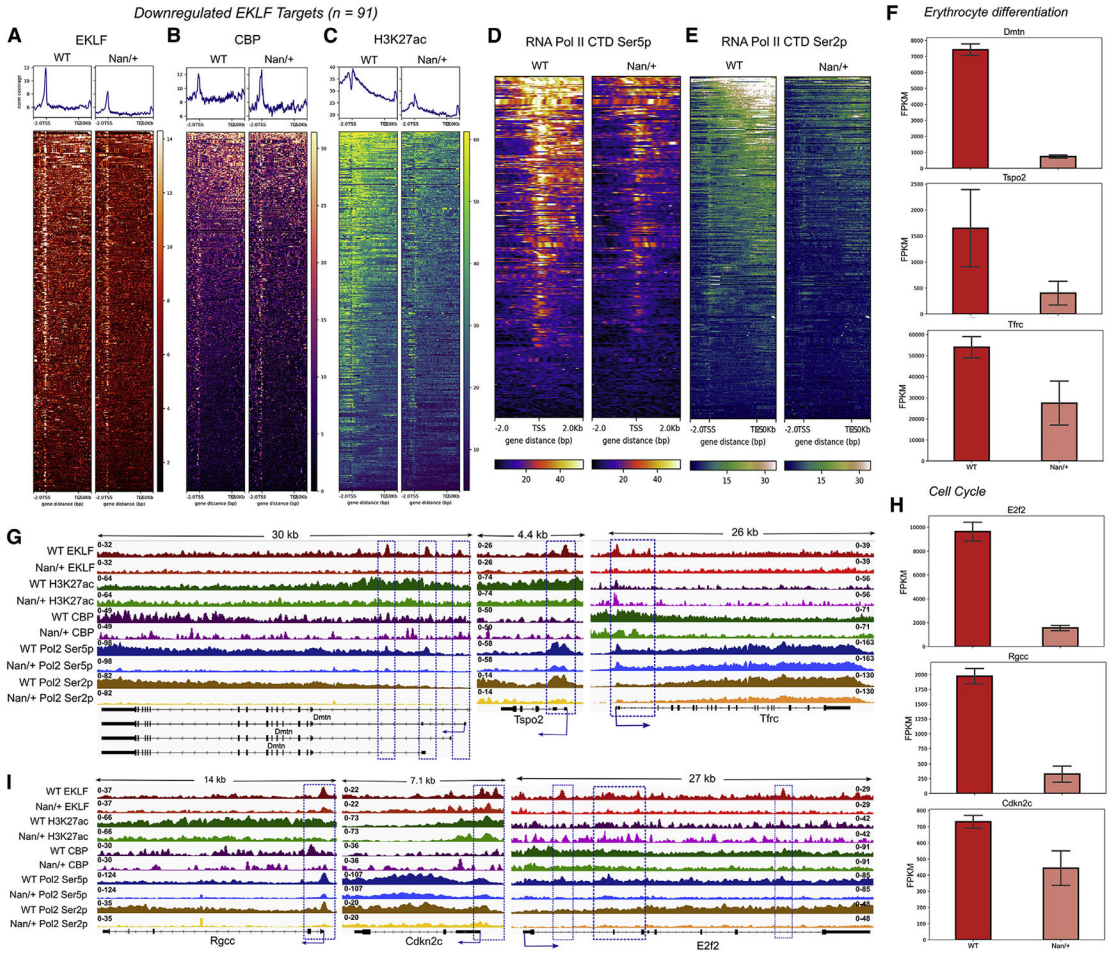


Figure 3. Loss of Nan-EKLF binding at specific sites downregulates genes by reducing pioneering and RNA polymerase II pausing and elongation

(A) Input-normalized ChIP-seq coverage and enrichment heatmap of EKLF from −2 kb upstream of the TSS to +2 kb downstream of the TTS at direct EKLF targets downregulated in Nan/+.

(B and C) Normalized ChIP-seq enrichment of (B) CBP and (C) H3K27ac at same locations as in (A).

(D and E) Input-normalized ChIP-seq heatmap of (D) RNA Pol II Ser5p ±2 kb of the TSS and (E) RNA Pol II Ser2p from −2 kb upstream of the TSS to +2 kb downstream of the TTS of the genes in (A).

(F and G) (F) Select EKLF targets downregulated in Nan/+ that have functions in erythroid maturation and (G) Integrative Genomics Viewer (IGV) tracks showing normalized ChIP-seq coverage of EKLF, CBP, H3K27ac, and RNA Pol II Ser5p and Ser2p at these loci. For *Dmtn* alternate transcripts are shown.

(H and I) (H) Cell-cycle targets downregulated in Nan/+ and (I) EKLF, CBP, H3K27ac, and RNA Pol II Ser5p and Ser2p ChIP-seq data for these loci.

Dotted blue boxes indicate locations of EKLF peaks and blue arrows indicate direction of transcription. The y-axis scale is normalized within each experiment and the number range at the left or right end indicates the scale.See also Figure S5 and Table S3. Error bars

represent the standard deviation from mean of the fragments per kilobase of exon per million mapped fragments (FPKM) values from three biological RNA-seq experiments.

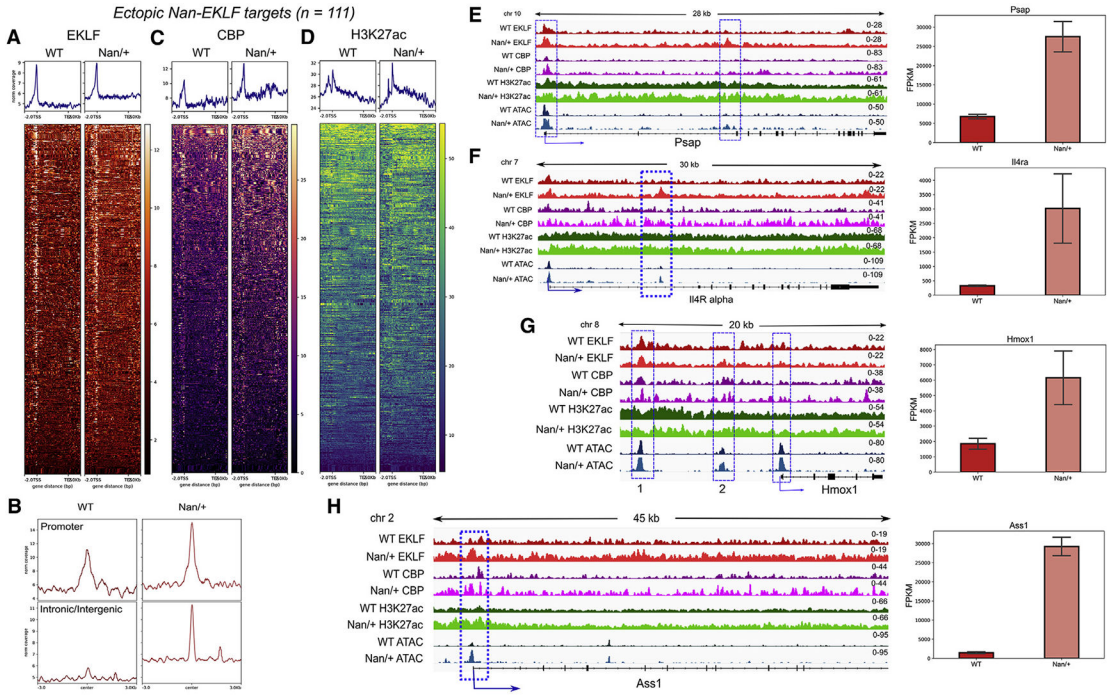


Figure 4. Ectopic or overexpression of genes caused by ectopic Nan-EKLF binding and CBP-mediated pioneering

(A) Input-normalized EKLF ChIP-seq coverage and enrichment heatmap −2 kb upstream of the TSS to +2 kb downstream of the TTS at Nan-EKLF targets upregulated in Nan/+.

(B) Input-normalized coverage profile ±3 kb from the peak center of Nan-EKLF peaks at promoters (top) and intronic/intergenic enhancers (bottom) at Nan-EKLF targets upregulated in Nan/+.

(C and D) Input-normalized ChIP-seq coverage and enrichment heatmap −2 kb upstream of the TSS to +2 kb downstream of the TTS at Nan-EKLF targets upregulated in Nan/+ for (C) CBP and (D) H3K27ac.

(E–H) Left: IGV tracks showing normalized ChIP-seq coverage of EKLF, CBP, H3K27ac, and ATAC-seq. Dotted blue boxes indicate locations of EKLF peaks and blue arrows indicate direction of transcription. The y-axis scale is normalized within each experiment and the number range at the right end indicates the scale. Right: RNA-seq expression level of ectopically expressed genes *Psap*, *Il4rμ*, *Hmox1*, and *Ass1*. See also Figure S6 and Table S4. Error bars represent the standard deviation from the mean of the FPKM values from three biological RNA-seq experiments.

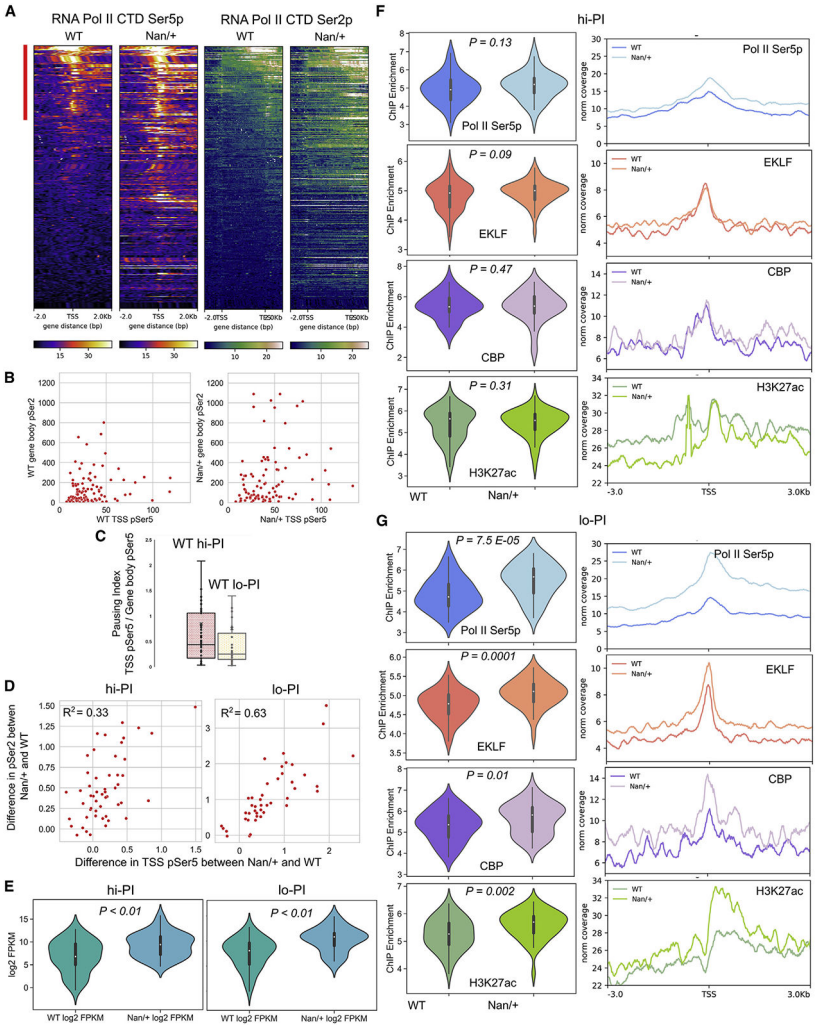


Figure 5. Ectopic expression or overexpression in Nan/+ is driven by modulating both RNA Pol II pausing and RNA Pol II pause-release

(A) Heatmap of input-normalized ChIP-seq enrichment of RNA Pol II Ser5p ±2 kb from the TSS and RNA Pol II Ser2p −2 kb upstream of the TSS to +2 kb downstream of the TTS of ectopic or overexpressed Nan-EKLF targets. Thick red line indicates high-pausing-index genes.

(B) Scatterplot of the Ser5p RNA Pol II at the TSS and Ser2p Pol II at gene bodies of ectopic genes in WT and Nan/+.

(C) Distribution of pausing index of genes in high- and low-pausing-index categories based on Pol II Ser5p at the TSS and gene bodies. Whiskers indicate outlier variability outside the upper and lower quartiles.

(D) Quadrant plots showing the difference between WT and Nan/+ of paused RNA Pol II Ser5p at the TSS (x axis) and elongating RNA Pol II Ser2p at the gene body (y axis) of genes in (A) with high and low pausing index.

(E) Violin plot showing distribution of RNA-seq expression as log2 FPKM for genes designated as hi-PI (high pause index) and lo-PI (low pause index) in Table S5. The p value

is indicated as the false discovery rate (FDR) threshold of DESeq2 analysis from RNA-seq data (adjusted p using the Benjamini-Hochberg method).
(F) Left: violin plots showing the distribution of RNA Pol II Ser5p, EKLF, CBP, and H3K27ac enrichment around the TSS (−30 to +100) of hi-PI genes. Right: corresponding input-normalized ChIP-seq coverage profiles ±3 kb of the TSS of hi-PI genes.
(G) Left: violin plots showing the distribution of RNA Pol II Ser5p, EKLF, CBP, and H3K27ac enrichment (−30 to +100) around the TSS of lo-PI genes quantified by Homer. Right: corresponding input-normalized ChIP-seq coverage profile ±3 kb of the TSS of lo-PI genes. The p values of the violin plots were calculated based on a pairwise Student t test assuming a homoscedastic one-tailed distribution.

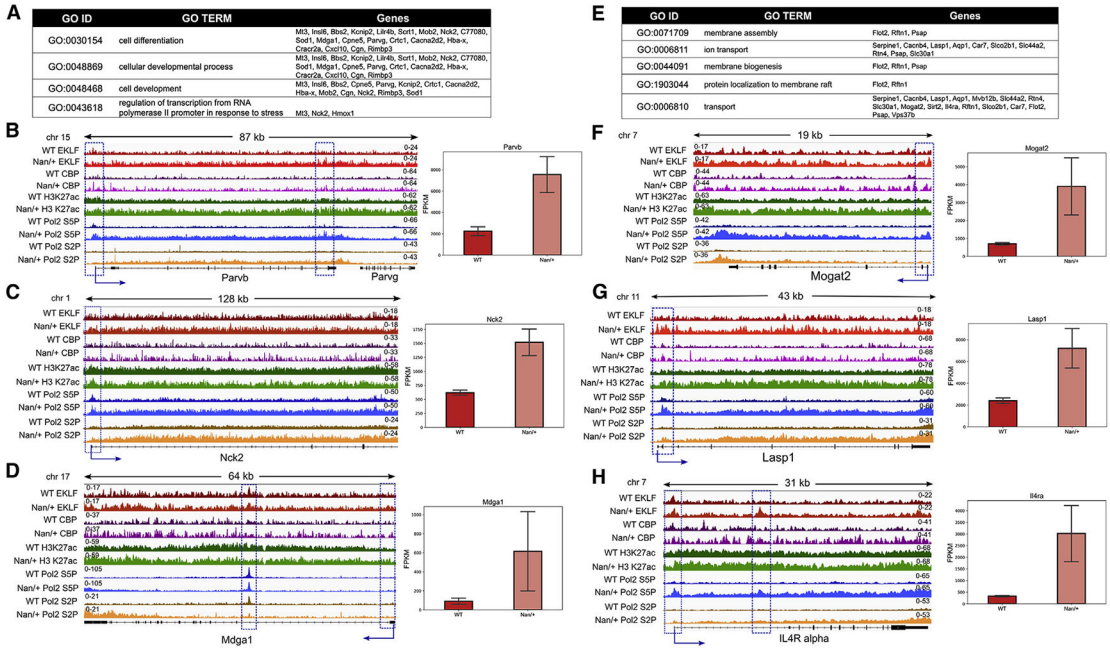


Figure 6. Functional differences in genes with high or low pausing index

(A) Gene ontology (GO) analysis of hi-PI genes.

(B–D) Left: IGV tracks of individual hi-PI genes from various GO categories in (A). Right: RNA-seq expression data.

(E) GO analysis of lo-PI genes.

(F–H) Left: IGV tracks of individual lo-PI genes from various GO categories in (E). Dotted blue boxes indicate locations of EKLF peaks and blue arrows indicate direction of transcription. The y-axis scale is normalized within each experiment and the number range at the left or right end indicates the scale. Right: RNA-seq expression data. Error bars represent the standard deviation from mean of the FPKM values from three biological RNA-seq experiments, adjusted $p < 0.01$. See also Figures S7 and S8.

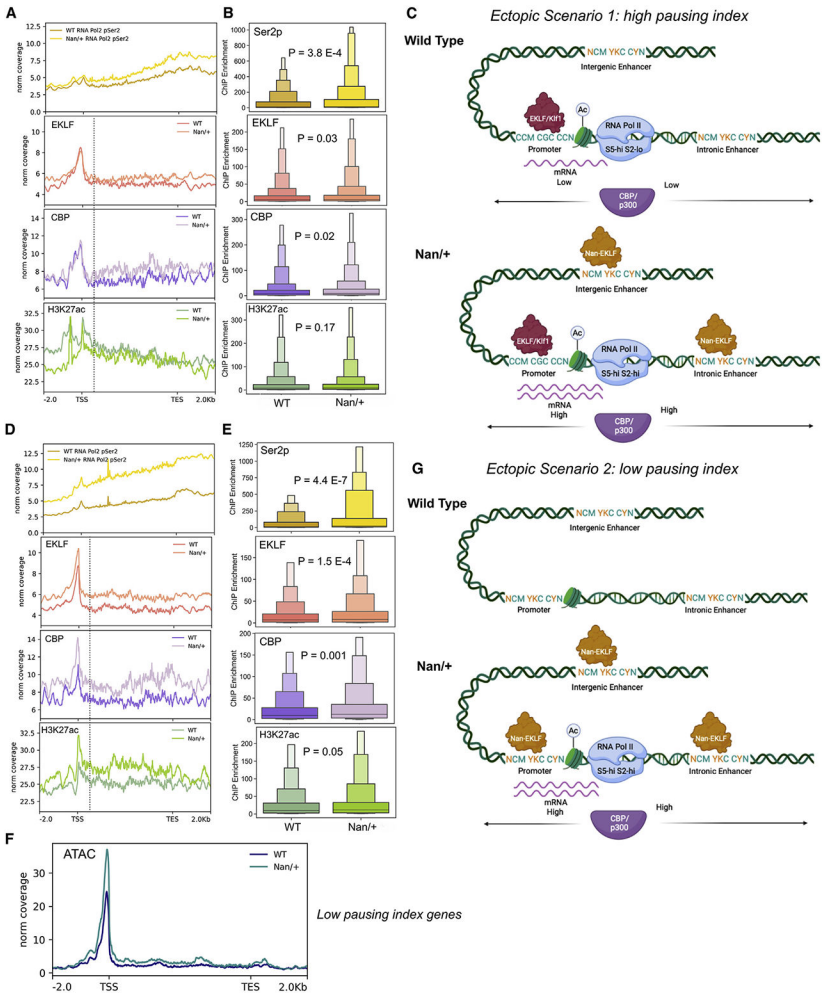


Figure 7. Ectopic or overexpression in Nan/+ by enhancer-driven RNA Pol II pause-release

(A) Input-normalized RNA Pol II Ser2p, EKLF, CBP, and H3K27ac ChIP-seq coverage profiles from −2 kb upstream of the TSS to +2 kb downstream of the TES of hi-PI genes.

(B) Enhanced boxplots showing quantile distributions for cumulative ChIP-seq enrichment of RNA Pol II Ser2p at the gene body and up to +200 bp downstream of the TES. For EKLF, CBP, and H3K27ac, cumulative ChIP-seq enrichment distribution is shown at intronic enhancers of hi-PI genes. The p values were calculated based on a pairwise Student t test assuming a heteroscedastic one-tailed distribution.

(C) Model for the EKLF-dependent RNA Pol II pausing and subsequent Nan-EKLF-mediated enhancer-driven release of paused RNA Pol II leading to overexpression or ectopic expression of hi-PI genes.

(D) Input-normalized RNA Pol II Ser2p, EKLF, CBP, and H3K27ac ChIP-seq coverage profiles from −2 kb upstream of the TSS to +2 kb downstream of the TES of lo-PI genes.

(E) Same analysis and plots as in (B) for lo-PI genes.

(F) ATAC-seq coverage profile −2 kb upstream of the TSS to +2 kb downstream of the TES of lo-PI genes.

(G) Model for the Nan-EKLF-driven RNA Pol II pausing and subsequent pause-release leading to ectopic expression of lo-PI genes.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
EKLF 7B2a mouse monoclonal antibody	this study	this study
Anti-Histone H3 (acetyl K27) antibody - ChIP Grade	Abcam	Cat# ab4729; RRID:AB_2118291
CBP (D6C5) Rabbit mAb	Cell Signaling technologies	Cat# 7389; RRID:AB_2616020
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody	Abcam	Cat# ab5131; RRID:AB_449369
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody	Abcam	Cat# ab5095; RRID:AB_304749
Normal rabbit IgG	Millipore Sigma	Cat# 12–370; RRID:AB_145841
Normal mouse IgG	Santa Cruz Biotechnologies	Cat# sc-2025; RRID:AB_737182
Anti-Ter119-APC	eBioscience	Cat# 17-5921-81; RRID:AB_469472
anti-CD44-FITC	Biorad	Cat# MCA89FT; RRID:AB_2076574
Chemicals, peptides, and recombinant proteins		
Micrococcal Nuclease	Millipore Sigma	N5386
10X RIPA buffer	Cell Signaling Technologies	9806
Protease Inhibitor Cocktail tablets EDTA free	Roche	11873580001
Dynabeads Protein-A	Life Technologies	100–02D
Dynabeads Protein-G	Life Technologies	100–04D
Proteinase K	Millipore Sigma	P2308
Critical commercial assays		
Neb Next DNA Ultra II DNA library preparation kit	New England Biolabs	E7645
DNA High Sensitivity Assay kit	Agilent	5067–4627
2x Tagment Buffer	Illumina	FC-121-1030
Tn5 Transposase	Illumina	FC-121-1030
QuantiTect SYBR Green PCR Kit	Qiagen	204143
Deposited data		
ChIP-seq and ATAC-seq	this study	NCBI-GEO: GSE210779
Code	this study	https://github.com/mkaustav84/biekerlab/blob/19bbdcf064436e7dd7502c60d8f2db8285829dba/scripts.md
Code	this study	https://doi.org/10.5281/zenodo.7315756
Experimental models: Organisms/strains		
Klf1^{Nan}/*Klf1*$^{+}$	Luanne Peters	MGI:4819280
Software and algorithms		
FCS Express 7	denovosoftware.com	DeNovo software™
Salmon	salmon.readthedocs.io	Version 1.9.0
Tximport	Bioconductor	https://doi.org/10.18129/B9.bioc.tximport
DESeq2	Bioconductor	https://doi.org/10.18129/B9.bioc.DESeq2

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Python Pandas	pandas.pydata.org	N/A
Python Seaborn	matplotlib.org	N/A
Python Matplotlib	seaborn.pydata.org	N/A
Bowtie2	bowtie-bio.sourceforge.net/bowtie2/index.shtml	Version 2.4
Samtools	www.htslib.org	Version 1.1
Deeptools	deeptools.readthedocs.io	Version 3.5
Bedtools	bedtools.readthedocs.io	Version 2.30
Homer	homer.ucsd.edu	Version 4.1
Biorender	biorender.com	N/A
Other		
AmpureXP beads	Beckman Coulter	A63880
SPRIselect beads	Beckman Coulter	B23317