Analysis of chromatin-state plasticity identifies cell-type–specific regulators of H3K27me3 patterns

Luca Pinello[a,1], Jian Xu[b,1], Stuart H. Orkin[b,c,2], and Guo-Cheng Yuan[a,2]

[a]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Heath, Boston, MA 02215; [b]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115; and [c]Howard Hughes Medical Institute, Boston, MA 02115

Contributed by Stuart H. Orkin, December 6, 2013 (sent for review August 13, 2013)

Chromatin states are highly cell-type–specific, but the underlying mechanisms for the establishment and maintenance of their genome-wide patterns remain poorly understood. Here we present a computational approach for investigation of chromatin-state plasticity. We applied this approach to investigate an ENCODE ChIP-seq dataset profiling the genome-wide distributions of the H3K27me3 mark in 19 human cell lines. We found that the high plasticity regions (HPRs) can be divided into two functionally and mechanistically distinct subsets, which correspond to CpG island (CGI) proximal or distal regions, respectively. Although the CGI proximal HPRs are typically associated with continuous variation across different cell-types, the distal HPRs are associated with binary-like variations. We developed a computational approach to predict putative cell-type–specific modulators of H3K27me3 patterns and validated the predictions by comparing with public ChIP-seq data. Furthermore, we applied this approach to investigate mechanisms for poised enhancer establishment in primary human erythroid precursors. Importantly, we predicted and experimentally validated that the principal hematopoietic regulator T-cell acute lymphocytic leukemia-1 (TAL1) is involved in regulating H3K27me3 variations in collaboration with the transcription factor growth factor independent 1B (GFI1B), providing fresh insights into the context-specific role of TAL1 in erythropoiesis. Our approach is generally applicable to investigate the regulatory mechanisms of epigenetic pathways in establishing cellular identity.

polycomb | hematopoiesis | histone modifications motifs

In eukaryotic cells the genome is organized into chromatin. The structure of the chromatin is highly cell-type–specific, providing an important additional layer of gene regulation. Recent genome-wide studies have identified the configuration of chromatin states with high resolution in diverse cell-types, and shown that genome-wide transcriptional levels are highly correlated with chromatin-state switches (1–5). Even within the same cell-type, chromatin-state switches are closely involved in fine-tuning gene-expression patterns in a developmental stage-specific manner (6, 7). These studies have provided strong evidence that the chromatin state plays an important role in establishing cell identity during development.

Despite the extensive epigenomic data generated during the past decade, mechanistic understanding of the determinants for chromatin states remains lacking. Whereas numerous regulatory pathways have been suggested (8), few studies have evaluated the contribution of each pathway to genome-wide patterns. Computational methods have been developed, but it remains difficult to predict genome-wide chromatin states ab initio (9).

One of the most intensely studied chromatin marks is H3K27me3, which is highly associated with gene repression and catalyzed by the EZH2 (enhancer of zeste homolog 2) subunit of the Polycomb repressive complex 2 (PRC2). Previous studies of H3K27me3 patterns have mainly been focused on promoter regions, where H3K27me3 target promoters are highly enriched with GC content (10, 11). Additional specificity is associated with a number of transcription factor (TF) motifs (12–14). Although less studied, a significant fraction of H3K27me3 is located in distal regions. It has been proposed that distal H3K27me3 marks poised enhancers, which can be activated through replacing H3K27me3 by H3K27ac (15, 16). An important question is whether there is a general principle controlling the plastic changes of H3K27me3 patterns across different cell types.

To systematically investigate the mechanisms modulating chromatin-state plasticity, we developed and validated a computational approach to identify distinct lineage-restricted regulators by focusing on high plasticity regions (HPRs). These regions were selected based on analyzing ChIP-seq data in numerous cell lines. We applied this approach to analysis of H3K27me3 ChIP-seq datasets obtained from the ENCODE consortium (17). We showed that the locations of the HPRs can be predicted by the underlying DNA sequences. These HPRs can be divided into two groups, corresponding to CpG island (CGI)-proximal and CGI-distal regions, with distinct properties. We found that the CGI-distal regions are more cell-type–specific and associated with distinct lineage-restricted transcriptional regulators. We applied this approach to investigate the regulation of the H3K27me3 pattern in primary human erythroid progenitor (ProE) cells, and identified a previously unrecognized context-specific function of the master regulator T-cell acute leukemia-1 (TAL1) in gene silencing through modulating poised enhancer activities.

Significance

We developed a computational approach to characterize chromatin-state plasticity across cell types, using the repressive mark H3K27me3 as an example. The high plasticity regions (HPRs) can be divided into two functionally and mechanistically distinct groups, corresponding to CpG island proximal and distal regions, respectively. We identified cell-type–specific regulators correlating with H3K27me3 patterns at distal HPRs in ENCODE cell lines as well as in primary human erythroid precursors. We predicted and validated a previously unrecognized role of T-cell acute lymphocytic leukemia-1 (TAL1) in modulating H3K27me3 patterns through interaction with additional cofactors, such as growth factor independent 1B (GFI1B). Our integrative approach provides mechanistic insights into chromatin-state plasticity and is broadly applicable to other epigenetic marks.

Author contributions: L.P., J.X., S.H.O., and G.-C.Y. designed research; L.P. and J.X. performed research; L.P. and J.X. contributed new reagents/analytic tools; L.P., J.X., S.H.O., and G.-C.Y. analyzed data; and L.P., J.X., S.H.O., and G.-C.Y. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no GSE52924).

[1]L.P. and J.X. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: stuart_orkin@dfci.harvard.edu or gcyuan@jimmy.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1322570111/-/DCSupplemental.

Results

Genome-Wide Characterization of H3K27me3 Plasticity. We obtained the ChIP-seq datasets of H3K27me3 in 19 human cell lines from the ENCODE consortium (17) (*SI Appendix*, Table S1). The raw sequence reads were normalized and mapped to nonoverlapping bins of 200 bp in size. We initially observed that the H3K27me3 patterns were highly variable across different cell types, with certain regions associated with high-degree of plasticity, such as the homeobox (HOX) gene clusters (Fig. 1*A*). We quantified the plasticity of H3K27me3 for each bin based on a metric known as the index of dispersion (IOD), and ranked the bins according to their associated plasticity scores (*Materials and Methods*). We selected 133,980 bins whose plasticity scores were significantly higher than the genome background ($P < 0.01$) (Fig. 1*B*). After merging neighboring selected bins, we obtained 46,755 contiguous regions that we refer to as the HPRs in the rest of the article. The average length of the identified HPRs is 507 bp (*SI Appendix*, Fig. S1). For comparison, we also defined low plasticity regions (LPRs) by using a similar criterion on the opposite end of distribution (*Materials and Methods*).

The HPRs are highly conserved compared with the genome-background (Fig. 1*C*), suggesting that they are enriched with functional elements. To rule out the possibility that the elevated conservation can be simply explained by an enrichment of exons, we divided the HPRs into three groups corresponding to exon, intron, and intergenic regions, respectively, and repeated the analysis within each group. For each group, the conservation scores are significantly higher at the HPRs compared with the genomic background and LPRs (*SI Appendix*, Fig. S2), suggesting the association is not an artifact. We further investigated the overlap between HPRs and various annotated functional elements, and found significant enrichment ($P < 1E\text{-}4$, Fisher's exact test) in CGIs, CGI shores, promoters, enhancers, and DNase I hypersensitivity sites, suggesting that variation of H3K27me3 is strongly associated with transcriptional regulatory activities (Fig. 1*D*). A closer examination suggested that the HPRs were associated with elevated variation of DNase-seq signal across cell types

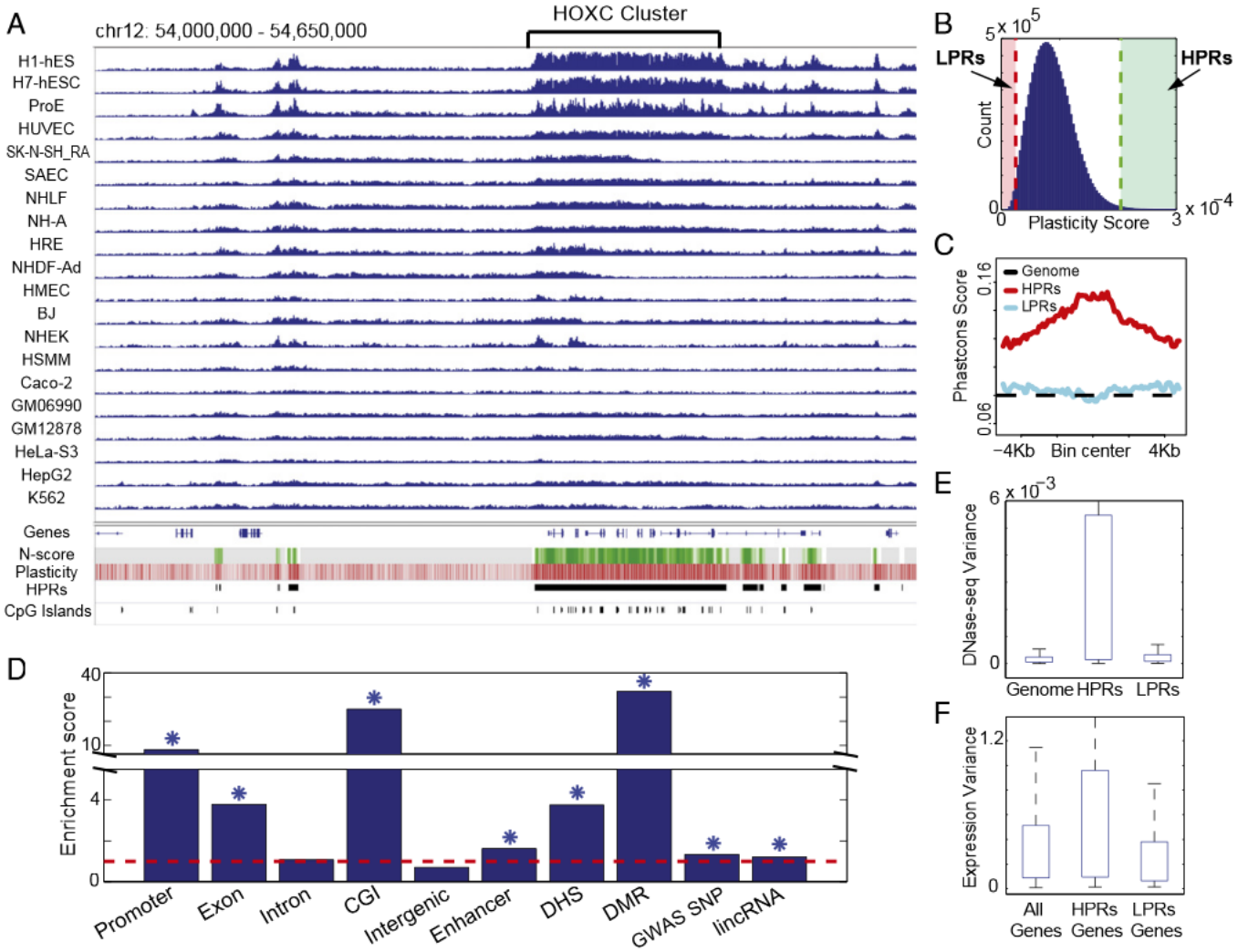

Fig. 1. Overview of the properties of HPRs. (*A*) ChIP-seq signals of H3K27me3 in 19 human cell lines from ENCODE (blue tracks) and corresponding plasticity scores (red track). Locations of HPRs are marked by the black segments. Predicted plasticity values (based on the N-score model) are shown as the green track. (*B*) Genome-wide distribution of the plasticity scores. Shaded areas show the location of HPRs and LPRs. (*C*) PhastCons scores surround HPRs (red) and LPRs (light blue). Baseline genome-wide average is also shown for reference (dashed black line). (*D*) Enrichment score of various annotated functional elements in HPRs. The stars indicate statistically significant enrichment ($P < 1E\text{-}4$, Fisher's exact test). (*E*) Boxplot showing the distribution of the variance of DNase-Seq signal in HPRs, LPRs, and genome-wide regions across the 19 cell lines. (*F*) Boxplot showing the distribution of the variance of the expression levels of the genes harboring either HPRs or LPRs in their promoters across the 19 cell lines. Genome-wide distribution is also shown for reference.

(Fig. 1*E*). We also observed a moderate enrichment in exons (Fig. 1*D*). Although the function of H3K27me3 in gene bodies remains unclear, recent studies have suggested that it may be related to alternative promoter use (18), alternative splicing (19, 20), or monoallelic gene expression (21). Long noncoding RNAs have recently been implicated in gene regulation and Polycomb recruitment (22, 23). Consistent with these studies, our analysis shows a moderate but statistically significant enrichment of long noncoding RNAs in the HPRs. In addition, we investigated the enrichment of HPRs in all of the repetitive classes of the RepeatMasker annotations. We found that the HPRs were significantly depleted in SINE, LINE, and LTR elements but enriched in Satellite elements. Similar biases were previously observed in mouse embryonic stem (ES) cells (24) (*SI Appendix*, Fig. S3).

We next investigated the correlation between H3K27me3 and DNA methylation variability. Previous studies have suggested that the transition from H3K27me3 to DNA methylation is a precursor event for cancer (25). High-throughput bisulphite sequencing of multiple cancer samples identified hundreds of differentially methylated regions (DMRs) in cancer among different cancer samples (26). These DMRs are highly colocalized with the HPRs defined herein based on the H3K27me3 patterns ($P < 1E-40$, Fisher's exact test; ES = 32.4) (Fig. 1*D*), confirming the close correlation between these two epigenetic marks. To test whether this colocalization can be simply explained by their common association with CGIs, we repeated our analysis by further constraining the distance with respect to CGIs. By analyzing CGI proximal (<2 kb) and distal (>2 kb) regions separately, we found that the HPRs were strongly associated with DMRs in both groups, with a much higher enrichment score in CGI distal DMRs (*SI Appendix*, Fig. S4).

If variation of H3K27me3 levels at HPRs plays a significant role in transcriptional regulation, we should expect that the expression levels of their target genes are also highly variable. To avoid the ambiguity associated with target gene identification, we considered only the subset of HPRs located within promoter regions, and compared the variance of the expression levels of their downstream genes relative to that of the genomic background. We found an increased variance in expression levels associated with HPR-associated genes (Fig. 1*F*). We applied GREAT (27) to search for enriched functional categories associated with HPRs and identified "transcription regulatory region sequence-specific DNA binding" ($P < 1E-40$) and "sequence-specific DNA binding" ($P < 1E-40$) as among the most enriched categories, further supporting a role of HPRs in establishing cell-type–specific gene expression programs.

Recently, a number of groups (3, 28) have identified domain-like histone modification patterns by analyzing ChIP-seq data. We compared the locations of HPRs identified by our model with the domains detected in these studies. We observed that 58% of the H3K27me3 domains previously identified by ref. 28 (total 2,498, 100-kb size) overlap with at least one HPR ($P < 1E-4$, permutation test) (*SI Appendix*, Fig. S5*A*). Similarly, of the 9,397 H3K27me3 domains identified by ref. 3 in ES cells, 57% overlap with at least one HPR ($P < 1E-4$, permutation test) (*SI Appendix*, Fig. S5*B*). A smaller percentage (26%) of the IMR90-associated H3K27me3 domains overlap with HPRs (*SI Appendix*, Fig. S5*C*), probably because IMR90 was not included in the 19 cell lines based on which the HPRs were identified.

Taken together, the above results strongly suggest that the HPRs mark critical regulatory regions for establishment of cell-type–specific gene regulatory programs. Further characterization of variation pattern within the HPRs to identify the underlying sequences and associated factors will likely provide important mechanistic insights into the establishment and maintenance of the cell-type specific epigenetic patterns.

Prediction of H3K27me3 Plasticity from DNA Sequences. To understand the molecular determinants of cell-type–specific HPRs, we analyzed DNA sequence signatures that were predictive of HPRs. Previous studies have identified a number of DNA sequence features associated with H3K27me3, including CGIs (10), TF sequence motifs (12, 13), and short RNA hairpins (29). However, these aforementioned studies were focused on one cell type (primarily ES cells) at a time; thus, to what extent the DNA sequence impacts the overall plasticity across cell-types remains poorly understood.

As an initial evaluation, we applied motif independent metric, which is a Kullback–Leibler distance-based method, to quantify DNA sequence specificity (30). We found that the HPRs were associated with a significantly higher degree of sequence specificity compared with the LPRs ($P < 1E-4$, permutation test), suggesting that DNA sequence information indeed contributes to the modulation of H3K27me3 variability. The 20 most informative k-mers are GC-rich, as expected (*SI Appendix*, Table S2).

The enrichment of CGIs in HPRs represents an extension of previous studies to multiple cell lines. In previous work, it was shown that a simple model based on the CGI distance and frequency alone was sufficient to predict many Polycomb targets in mouse ES cells (11, 12), although improved accuracy could be achieved by more sophisticated models (13). To test whether this property was applicable to H3K27me3 plasticity, we built a simple model using only the C+G and CpG density sequence features to distinguish HPRs from LPRs (named CG-only model; see *Materials and Methods* for details). As anticipated, the CG-only model already has substantial prediction power [area under the receiver operating characteristic curve (AUC) = 0.74] (Fig. 2*A*).

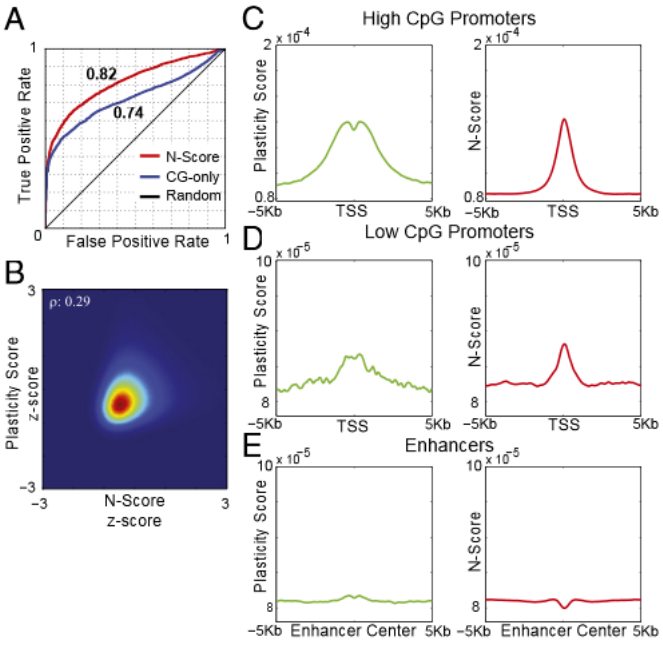

Fig. 2. Prediction of H3K27me3 plasticity using DNA sequence patterns. (*A*) Receiver operating characteristic (ROC) curve for classification of HPRs vs. LPRs using either the N-score model (in red) or the CG-only model (in blue). The corresponding AUC scores are shown next to the ROC curves. (*B*) Genome-wide correlation between the observed and N-score model predicted plasticity scores. (*C*) Average profiles of observed and N-score predicted plasticity score and at high CpG promoters. (*D*) Average profiles of observed and N-score predicted plasticity score and at low CpG promoters. (*E*) Average profiles of observed and N-score predicted plasticity score and at enhancer regions.

We then assessed to what extent the prediction accuracy could be improved by incorporating additional sequence features. By applying the previously developed N-score model (31, 32) (*Materials and Methods*), we obtained an improved AUC score at 0.82 (Fig. 2*A*; see *SI Appendix*, Table S3 for the list of the most informative sequence features). Although this improvement seems moderate, it enabled us to predict not only CGI proximal, but also distal HPRs (*SI Appendix*, Figs. S6 and S7). For example, the improved performance of the N-score model is evident at the flanking regions of the HOXC locus (Fig. 1*A*). The genome-wide correlation between predicted and observed plasticity score is statistically significant (= 0.29, *P* value < 1E-40) (Fig. 2*B*). Importantly, our model correctly predicted the overall pattern at promoters and enhancers (Fig. 2 *C*–*E*), including the significant differences between GC-rich and GC-poor promoters. It is important to note that the overall plasticity score at enhancers was low because of the presence of nonvariable H3K27me3 regions within all of the enhancers. As expected, when using only enhancers overlapping with the HPRs in the analysis, we found that the plasticity scores were much higher (*SI Appendix*, Fig. S8).

Classification of H3K27me3 Variability Within the HPRs. We applied the *k*-means clustering method to identify distinct subpatterns of H3K27me3 variability within the HPRs (Fig. 3*A*). These patterns can be broadly classified into three groups: a "continuous" group (C8 and C9), where the H3K27me3 mark is present in most cell-types but its intensity is highly variable; a "binary" group (C1, C2, C3, C4, and C6), where the H3K27me3 mark is present mainly in a small number of cell types; and a "mixed" group that contains all other clusters. We observed a strong association between the continuous group and the overall GC-content (Fig. 3*A*). The proximal HPRs tend to be spatially extended, whereas the distal HPRs are more focal (Fig. 3*B*). These differences suggest that the CGI proximal and distal HPRs are regulated through distinct mechanisms.

Of note, 44% of the total HPRs are CGI distal (Fig. 3*C*). Although the role of promoter-associated H3K27me3 in gene silencing has been extensively studied, its role in distal regions remains less well understood. Previous studies suggested that H3K27me3 may work together with H3K4me1 in marking poised enhancers, which could be activated by switching to the H3K27ac mark (15, 16), To quantitatively assess the degree of colocalization between distal HPRs and enhancers, we applied chromHMM (33), a recently developed method for chromatin-state segmentation, to identify putative enhancers based on the combinatorial pattern of nine histone marks. Using this approach, we were able to annotate the enhancer regions in 12 of the 19 cell-lines, including 9 cell-lines for which the enhancers have been annotated previously (5). We further excluded regions that are within 2 kb of an annotated transcription start site and CGIs, resulting in a remaining set of 57,815 HPRs that overlapped with chromHMM identified enhancers. Fourteen percent of the distal HPRs are located in the enhancer regions, which is highly statistically significant (9% expected by chance; *P* < 1E-40).

Identification of Distal HPR-Associated Cell-Type–Specific TFs. To systematically identify candidate TFs that may modulate H3K27me3 variability in a cell-type–specific manner, we developed a unique bioinformatic pipeline that consists of the following five steps (Fig. 4*A* and *SI Appendix*, Fig. S9). First, we defined a *z*-score by comparing the H3K27me3 intensity in each cell-type relative to the other 18 cell-types, and selected a subset of HPRs (which we named z-HPRs) with the highest *z*-scores (>2) for further analysis. Second, we generated a genomic background with a matching C+G content. Third, we scanned the DNA sequences at z-HPRs for enrichment of known TF motifs in the TRANSFAC (34), JASPAR (35), and FACTORBOOK (36) databases, by comparing with the matched genomic background (*P* value cutoff = 1E-4).

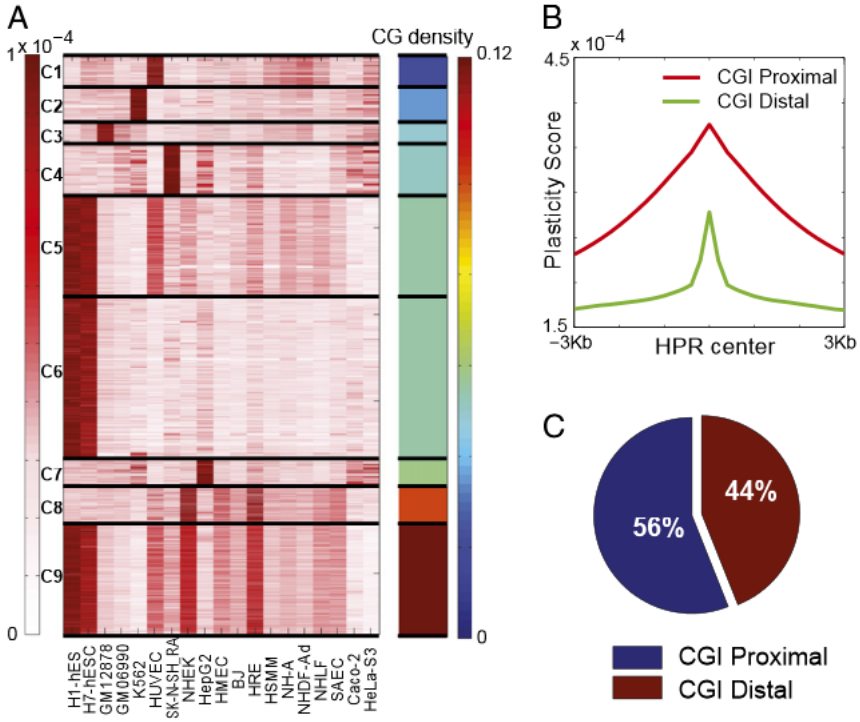

Fig. 3. Distinct variation patterns between subsets of HPRs. (*A*) *k*-means clustering of HPRs shows distinct patterns of variation. The color bar on the left shows H3K27me3 ChIP-seq read counts, and the color bar on the right shows CpG density. (*B*) Average plasticity scores around CGI proximal and distal HPRs. (*C*) Decomposition of the total HPRs into CGI proximal and distal subsets.

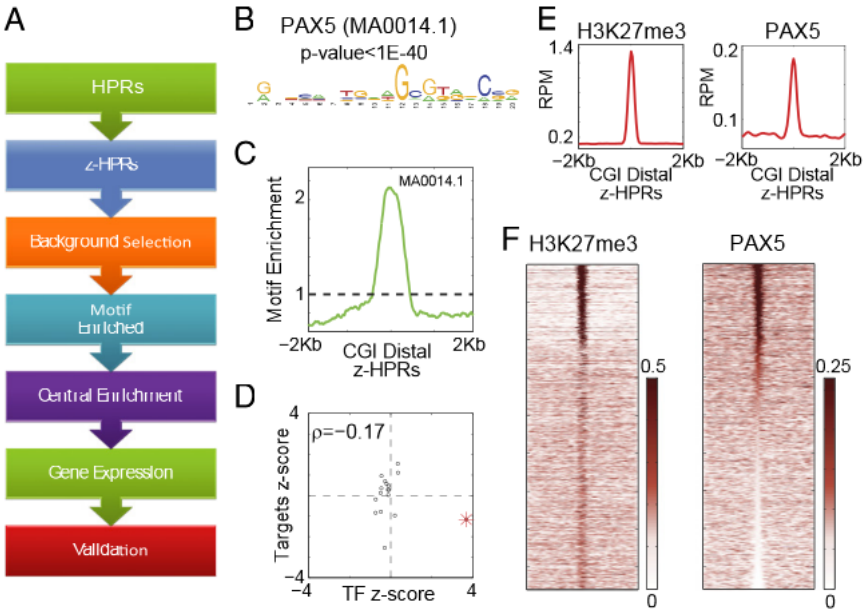

Fig. 4. Identification of cell-type–specific z-HPR regulators. (*A*) A flowchart of the computational pipeline for identification of cell-type specific z-HPR regulators (details shown in the main text and *SI Appendix*, Fig. S5). (*B*–*F*) Application of the above pipeline identifies PAX5 as a candidate regulator for GM12878/GM06990 z-HPRs. (*B*) The PAX5 motif logo in the JASPAR database. (*C*) Average enrichment profile of PAX5 motif sites around distal z-HPRs. (*D*) Correlation between the expression levels of PAX5 and target genes neighboring z-HPRs across different cell lines. Expression levels are normalized by z-scores. The red star corresponds to the GM12878/GM06990 cell lines. Average profile (*E*) and heatmap (*F*) of ChIP-seq signal for H3K27me3 and PAX5 around distal z-HPRs.

Fourth, we compared the motif site frequency in the center of z-HPRs with the flanking regions, and selected the motifs that were enriched in the center regions (fold-change > 1.2). Fifth, we further integrated gene expression data to select for candidate TFs that were specifically expressed in the cell-type (*z*-score > 0.75) whereas its target genes neighboring the z-HPRs were repressed (*z*-score < -0.75).

In total, our analysis identified 41 cell-type–specific associations between TFs and z-HPRs (Table 1). Several of these associations were implicated in the literature. For example, our analysis identified paired box 5 (PAX5) as a candidate Polycomb recruitment factor in B-lymphoblastoid cells (based on GM12878/GM06990 cell lines) (Fig. 4 *B*–*D*). PAX5 is known to be essential for normal B-cell development by activating B-cell commitment genes, while concomitantly repressing non–B-cell lineage genes (37). Depletion of PAX5 in pro-B cells resulted in significantly reduced levels of H3K27me3 at V_H genes, suggesting a role of PAX5 in the recruitment of PRC2 complex (38). To validate our predictions, we analyzed the colocalization between z-HPRs and PAX5 binding sites, using a ChIP-seq dataset generated by the ENCODE consortium. This analysis demonstrated that PAX5 binding signal was indeed highly enriched at z-HPRs (Fig. 4 *E* and *F*).

Furthermore, our analysis identified GATA binding protein 1 (GATA1) to be strongly associated with z-HPRs in the human erythroleukemia cell line K562 (*SI Appendix*, Fig. S10 *A*–*C*). GATA1 is a master regulator for erythroid and megakaryocyte development (39) and its mutation has been linked to Down syndrome and acute megakaryoblastic leukemia (40). Previous work showed that GATA1 employs the PRC2 complex in gene repression in erythroid cells (41). To validate our predictions, we obtained a GATA1 ChIP-seq dataset in K562 cells generated from the ENCODE consortium. Similar to PAX5 in B cells, we found that the GATA1 ChIP-seq signal was highly enriched at the z-HPRs in K562 cells (*SI Appendix*, Fig. S10 *D* and *E*).

In addition, our analysis suggested that AP2-□ [or transcription factor AP-2 alpha (TFAP2A)] may recruit Polycomb in normal human epidermal keratinocytes (NHEK) (*SI Appendix*, Fig. S11), consistent with a role of AP2-□ in regulating keratinocyte-specific gene expression (42). Similarly, ZFX (zinc finger protein, X-linked) was associated with z-HPRs in SK-N-SH_RA cells (a neuroblastoma cell line) (*SI Appendix*, Fig. S12), consistent with an implicated role of ZFX in aggressive gliomas (43). These analyses not only validated the computational predictions of our bioinformatics pipeline, but also provided links between lineage-restricted TFs and the Polycomb regulators that may be functionally important for their roles in respective cell lineages.

Computational Analysis Identifies TAL1 in Regulating H3K27me3 Variability. Recently we have applied genomic methods to query the regulatory mechanism underlying developmental-stage changes of gene activities, using ProE cells as a model system (6). We found that these changes were mainly mediated by the enhancer activities, whereas the promoter activities were nearly identical between developmental stages. In addition, we observed a small fraction but still significant number of enhancers to be transcriptionally "poised" (1,249 of 12,960 total enhancers), as defined by the presence of both H3K4me1 and H3K27me3 marks (15, 16).

To gain mechanistic insights into the cell-type–specific H3K27me3 activity in the ProE cells, we applied the computational pipeline described above to determine ProE-specific H3K27me3 modulators. To this end, we adjusted the choice of HPRs by merging the H3K27me3 data in ProE and the other 19 cell lines. Of the total HPRs, 78% were unaffected by this adjustment, suggesting that incorporating a new dataset had only a moderate effect on the overall HPR selection. On the other hand, the adjustment is useful for identifying variability signal that is specific to the ProE cells. Among the 20,174 total ProE z-HPRs, 33% are located in distal regions. Importantly, we observed a statistically significant overlap between the ProE-specific z-HPRs and poised enhancers (160 common regions, as opposed to 5 expected by chance, $P <$ 1E-4 permutation test) (*Materials and Methods*). We then identified

Table 1. List of distal z-HPR associated TFs predicted by our computational pipeline

Gene name	Motif ID	Presence in target (%)	Presence in background	Ratio	*P* value	*Z*-scores	Central enrichment	*P* value
BJ								
CRX	V_CRX_Q4	24.60	17.50	1.38	5.23E-07	(2.61; □1.04)	1.92	<1E-40
NHLH1	V_HEN1_01	22.20	16.10	1.36	2.68E-05	(1.82; □1.05)	1.88	<1E-40
Caco-2								
EGR1	V_EGR_Q6	24.30	20.20	1.19	8.06E-05	(1.53; □1.45)	1.56	<1E-40
GM12878/GM06990								
SPZ1	V_SPZ1_01	35.90	22	1.61	<1E-40	(2.41; □1.22)	4.17	<1E-40
PAX5	MA0014.1	21.40	11.20	1.84	<1E-40	(3.67; □1.16)	3.62	<1E-40
PAX4	V_PAX4_04	20.20	11.40	1.7	<1E-40	(1.85; □0.978)	2.85	<1E-40
ZBTB7B	V_CKROX_Q2	39.10	28.10	1.38	<1E-40	(1.66; □1.05)	6.7	<1E-40
IKZF1	V_IK_Q5	43.20	32.40	1.32	<1E-40	(3.08; □1.68)	2.01	<1E-40
EWSR1	MA0149.1	34.60	25.10	1.36	<1E-40	(1.09; □1.51)	3.79	<1E-40
MZF1	MA0057.1	22.30	14.40	1.51	<1E-40	(1.4; □1.1)	4.51	<1E-40
SREBF2	V_SREBP2_Q6	26.60	19.20	1.37	<1E-40	(1.13; □1.43)	2.09	<1E-40
GLI1	V_GLI_Q2	26.40	21.40	1.22	3.91E-39	(0.974; □1.76)	1.58	<1E-40
INSM1	MA0155.1	19	15.60	1.21	4.70E-22	(1.48; □0.908)	1.6	<1E-40
PRDM1	PRDM1	14.10	11.70	1.18	2.38E-12	(1.77; □1.75)	1.29	<1E-40
ZNF263	ZNF263	29	25.90	1.11	1.59E-11	(0.945; □1.62)	3.42	<1E-40
E2F1	V_E2F_Q6_01	3.08	2.12	1.31	3.52E-09	(1.13; □1.5)	2.17	<1E-40
MYB	V_CMYB_01	6.92	5.66	1.19	3.69E-06	(1.34; □0.836)	1.6	<1E-40
EGR2	V_EGR2_01	9.24	7.82	1.16	8.66E-06	(2.7; □0.967)	1.65	<1E-40
HeLa-S3								
HOXA7	I_ANTP_Q6_01	17.70	14.50	1.2	4.19E-09	(0.944; □1.39)	1.48	<1E-40
HepG2								
GFI1	V_GFI1_Q6	8.89	6.91	1.25	8.87E-12	(2.68; □1.54)	1.3	<1E-40
HRE								
MAZ	V_MAZ_Q6	49.20	41	1.19	1.97E-25	(0.81; □0.806)	1.56	<1E-40
REST	V_NRSF_Q4	37.50	29.80	1.25	3.13E-25	(0.815; □0.809)	1.71	<1E-40
HUVEC								
FOXD3	MA0041.1	18.30	15.60	1.16	3.10E-12	(0.876; □1.69)	1.36	<1E-40
MECOM	V_EVI1_01	8.47	7.06	1.17	5.78E-06	(1.53; □0.958)	1.25	<1E-40
ZBTB16	V_PLZF_02	10.50	9.01	1.15	2.80E-05	(1.2; □1.77)	1.36	<1E-40
K562								
GATA1	V_GATA1_01	6.34	5.43	1.14	5.39E-08	(3.69; □2.45)	1.22	<1E-40
TBP	V_TBP_01	4.29	3.55	1.16	1.02E-07	(2.68; □2.23)	1.5	<1E-40
NHEK								
E2F4	E2F4	26.10	18.60	1.38	<1E-40	(0.834; □1.94)	2.14	<1E-40
ZNF219	V_ZNF219_01	45.10	36.20	1.24	<1E-40	(1.25; □1.76)	1.67	<1E-40
SREBF1	V_SREBP_Q6	41.40	33.20	1.24	<1E-40	(2.12; □1.71)	1.46	<1E-40
KLF4	MA0039.2	40.40	33.90	1.19	1.95E-29	(1.17; □1.73)	1.46	<1E-40
SREBF1	V_SREBP1_Q5	39.70	33.30	1.19	7.95E-29	(2.12; □1.69)	1.47	<1E-40
OVOL1	V_MOVOB_01	26.40	21.50	1.22	1.40E-20	(0.897; □1.84)	1.55	<1E-40
SREBF2	V_SREBP2_Q6	32.30	27	1.19	2.03E-20	(1.18; □1.74)	1.5	<1E-40
TFAP2A	AP2	33.30	28	1.18	2.76E-20	(1.61; □2.15)	1.38	<1E-40
VDR	V_VDR_Q6	14	10.90	1.27	5.28E-14	(2; □1.93)	1.3	<1E-40
BBOX1	B-Box	14.90	12.30	1.2	5.52E-08	(0.83; □1.35)	1.52	<1E-40
ZBTB7A	V_LRF_Q2	25	21.80	1.14	1.16E-07	(1.05; □1.88)	1.35	<1E-40
SK-N-SH_RA								
ZFX	MA0146.1	46.30	43.50	1.06	5.91E-15	(0.817; □2.62)	1.49	<1E-40
ZNF350	V_ZBRK1_01	23.40	21.20	1.1	4.16E-14	(1.14; □2.59)	1.44	<1E-40
KLF12	V_AP2REP_01	13.90	12.80	1.08	2.86E-04	(1.25; □2.68)	1.28	<1E-40

eight TFs that are associated with these distal z-HPRs (Table 2). Of note, among these identified TFs, four were previously reported to physically interact [TAL1-LIM domain only 2 (LMO2), nescient helix loop helix 1 (NHLH1)-LMO2, and TAL1-epididymal sperm binding protein 1 (ELSPB1)], suggesting that they cooperate in regulating H3K27me3 variability and Polycomb activities (44). Importantly, our analysis identified TAL1, a principal regulator of hematopoietic development (39), as a candidate regulator of ProE-specific H3K27me3 variability (Fig. 5 *A*□*C*). TAL1 is commonly known as a transcriptional activator in various hematopoietic lineages (39), although a role for gene repression has recently been described in a different cellular context (embryonic endothelium) (45). Our analysis suggested that TAL1 might also contribute to transcriptional repression in hematopoietic cells by modulating the activity of poised enhancers. To test this hypothesis, we examined genome-wide TAL1 chromatin occupancy in ProE cells by ChIP-seq analysis. Consistent with the predicted association between TAL1 motifs and distal HPRs (Fig. 5 *A*□*C* and Table 2), we observed a significant enrichment of TAL1 ChIP-seq signal within distal HPRs in ProE cells (Fig. 5 *D* and *E*).

Table 2. List of TFs predicted to be associated with ProE distal z-HPRs

Gene name	Motif ID	Presence in target (%)	Presence in background	Ratio	*P* value	*Z*-scores	Central enrichment	*P* value
NHLH1	MA0048.1	36	21.70	1.63	2.41E-16	(3.37; □3.02)	4.57	<1E-40
LMO2	V_LMO2COM_01	47.70	32.90	1.43	4.24E-15	(3.45; □3.08)	3.5	<1E-40
TFAP4	V_AP4_Q6_01	42	28.30	1.47	9.40E-14	(1.52; □3.02)	4.05	<1E-40
UBP1	V_LBP1_Q6	28.50	16.70	1.66	5.15E-13	(1.75; □3.02)	5.96	<1E-40
ELSPBP1	V_E12_Q6	38.70	26.50	1.45	1.27E-11	(3.96; □3)	3.9	<1E-40
TAL1	V_TAL1_Q6	39	27	1.43	4.18E-11	(3.24; □3.02)	4.05	<1E-40
E2F1	V_E2F_Q2	8.88	4.35	1.85	3.59E-06	(1.66; □3.2)	12.03	<1E-40
ESR1	MA0112.2	52.70	44.50	1.18	1.94E-05	(1.22; □3.1)	2.77	<1E-40

To gain functional insights, we applied GREAT (27) to identify enriched Gene Ontology (GO) categories associated with the poised enhancers. One of the most enriched GO biological processes was "regulation of heart contraction" [false-discovery rate (FDR) q-value = 1.1E-3], and one of the strongly associated mouse phenotypes was "congestive heart failure" (FDR q-value = 7.4E-4) (*SI Appendix*, Table S4). Interestingly, in the aforementioned study (45), the genes repressed by TAL1 were also associated with cardiomyogenesis. Despite the differences between the two model systems, we observed a significant overlap (58 of 965, P = 5.2E-6) between the genes harboring a poised enhancer in ProE cells and those up-regulated by TAL1 depletion in embryonic endothelium (45). In contrast, TAL1-H3K27ac genes were enriched with biological processes such as "regulation of erythrocyte" (P = 1.5E-16), "negative regulation of cell-cell adhesion" (P = 5.4E-11), and "response to iron ion" (P = 6.2E-6), consistent with the role of TAL1 in activating genes required for normal erythroid development (46, 47).

TAL1 Modulates H3K27me3 Variability at Poised Enhancers in Erythroid Cells. The above findings suggest that TAL1 plays a dual role in regulating gene-expression programs in primary human erythroid cells. We next investigated the context differences between TAL1–z-HPR and TAL1-H3K27ac regions using motif analysis. The most enriched motif in TAL1–z-HPR regions corresponds to the consensus binding sequence for GFI1B (Fig. 6*A*), a transcriptional repressor critical for normal hematopoietic development and oncogenesis (48, 49). The transcriptional corepressor complex lysine (K)-specific demethylase 1A (LSD1)/CoREST was previously shown to be required for GFI1B-mediated transcriptional repression in erythroid cells (50). In contrast, the most significantly enriched motif in TAL1-H3K27ac regions was the TAL1:

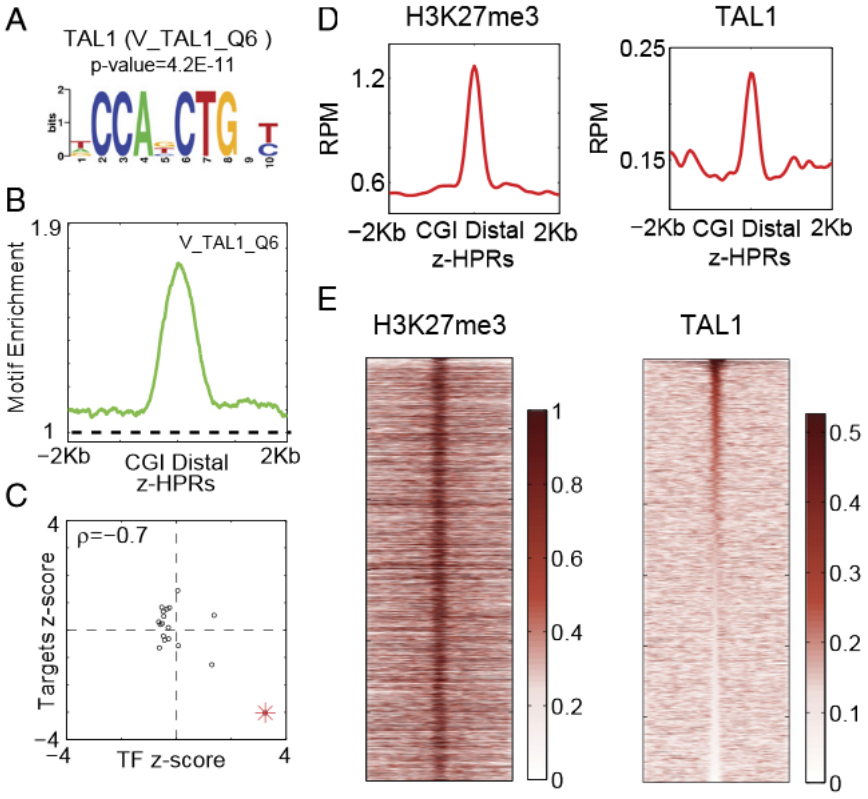

Fig. 5. TAL1 is predicted to be associated with z-HPRs in ProE cells. (*A*) TAL1 motif logo from the TRANSFAC database. (*B*) Average enrichment profile of TAL1 motif sites around distal z-HPRs. (*C*) Correlation between the expression level of TAL1 and target genes neighboring distal z-HPRs. The red star corresponds to the ProE cells. (*D*) Average ChIP-seq signals surrounding H3K27me3 and TAL1 around distal z-HPRs. (*E*) Heatmaps of ChIP-seq signals surrounding H3K27me3 and TAL1 around distal z-HPRs. Each row represents a different z-HPR region. The color bar represents the level of ChIP-seq signal.

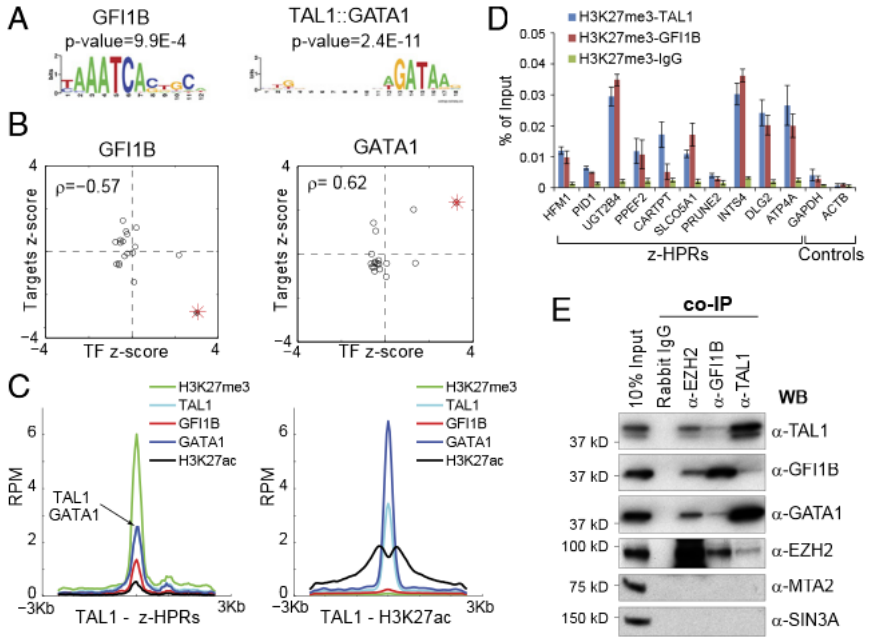

Fig. 6. GFI1B is a cofactor for TAL1-mediated H3K27me3 variation in ProE cells. (*A*) GFI1B and TAL1::GATA1 motif logos from the TRANSFAC database. (*B*) Correlation between the expression level of GFI1B (*Left*) or GATA1 (*Right*), and target genes neighboring TAL1+z-HPRs (or TAL1+H3K27ac). (*C*) Average ChIP-seq signal for multiple factors surrounding TAL1+z-HPR (*Left*) or TAL1+H3K27ac (*Right*) regions. (*D*) Sequential ChIP analysis in ProE cells with the indicated pairs of antibodies. ChIP DNA was analyzed by quantitative PCR for selected TAL1-GFI1B-H3K27me3 cooccupied z-HPRs and negative control regions (GAPDH and ACTB). The percent (%) of input is shown. Results are mean ± SD from at least three technical replicates. (*E*) Co-IP experiments for EZH2, TAL1, GATA1, and GFI1B using nuclear extracts from ProE cells. MTA2 and SIN3A are analyzed as negative controls.

GATA1 composite motif (Fig. 6*A* and *SI Appendix*, Fig. S13*A*). Gene-expression analysis showed that GFI1B was up-regulated, whereas its target genes were down-regulated in ProE cells (Fig. 6*B*). An opposite trend was associated with GATA1/TAL1 common targets (Fig. 6*B*). These differences are consistent with our computational prediction based on sequence analysis.

To experimentally test whether TAL1 and GFI1B cooperate to regulate H3K27me3 at poised enhancers, we generated GFI1B and GATA1 genome-wide binding profiles in ProE cells by ChIP-seq analysis. Importantly, the ChIP-seq signal for GFI1B was substantially higher at TAL1–z-HPRs than the TAL1-H3K27ac regions, whereas the GATA1 ChIP-seq signal was stronger at the TAL1-H3K27ac regions than TAL1–z-HPRs (Fig. 6*C*). This difference was highly consistent with the computational predictions, suggesting that TAL1 and GFI1B may function cooperatively within z-HPRs. Additionally, the target genes of GFI1B-TAL1-H3K27me3 cooccupied regions were expressed at lower levels compared with those corresponding to TAL1-H3K27me3 or TAL1-H3K27ac, respectively (*SI Appendix*, Fig. S13*B*). To further determine whether TAL1, GFI1B and H3K27me3 cooccupy on the same allele in ProE cells, we performed sequential ChIP analysis. Sequential ChIP analysis of TAL1 and H3K27me3 demonstrated colocalization at nine out of ten representative z-HPRs (Fig. 6*D* and *SI Appendix*, Fig. S14*A*). Similarly, sequential ChIP analysis of GFI1B and H3K27me3 showed colocalization at 8 of 10 z-HPRs (Fig. 6*D*). The reverse direction of TAL1-H3K27me3 and GFI1B-H3K27me3 sequential ChIP also indicated colocalization at the majority of z-HPRs (*SI Appendix*, Fig. S14 *B* and *C*). Taken together, these results are consistent with the colocalization of TAL1, GFI1B, and H3K27me3 at a genomic scale (Fig. 6*C*), suggesting that TAL1 and its cofactor GFI1B cooperate on the same alleles to modulate H3K27me3 in ProE cells.

To further establish the causality of TAL1 occupancy and H3K27me3 variability, we ectopically expressed TAL1 in the human lymphoid (REH) and embryonic kidney (293T) cell lines, followed by ChIP-quantitative PCR analysis of H3K27me3 at the representative ProE-specific z-HPRs (*SI Appendix*, Fig. S15). Notably, a modest increase of H3K27me3 and a significant increase of TAL1 binding were observed at three loci (UGT2B4, INTS4, and ATP4A), respectively. Similarly, ectopic expression of TAL1 in 293T cells led to modest increase of H3K27me3 at two loci (PID1 and INTS4). Taken together, these results suggest that TAL1 can gain access to at least a subset of ProE-specific z-HPRs in other cellular context, accompanied by increased H3K27me3 at the same genomic location (*SI Appendix*, Fig. S15).

The colocalization of TAL1 and GFI1B within z-HPRs suggests that TAL1, GFI1B, and the PRC2 complex may physically interact to modulate H3K27me3 in a context-specific manner. To test this hypothesis, we performed a coimmunoprecipitation (co-IP) experiment using nuclear extracts prepared from ProE cells. Importantly, EZH2, the histone methyltransferase subunit of the PRC2 complex, could efficiently pull-down both TAL1 and GFI1B in ProE cells (Fig. 6*E*). Similarly, both TAL1 and GFI1B could pull-down the endogenous EZH2 protein, respectively, whereas no detectable interactions were found for MTA2 (a core subunit of the Mi-2□/NuRD complex) and SIN3A (a core subunit of the transcriptional corepressor and histone deacetylase complex SIN3A) (Fig. 6*E*). These results suggest that TAL1, GFI1B, and PRC2 physically interact and cooperate within the same complex that functions to regulate H3K27me3 marks at a subset of gene-distal regulatory elements.

Taken together, these data strongly support a model in which TAL1 may mediate gene repression by modulating H3K27me3 variation through interaction with transcriptional cofactors GFI1B and the Polycomb regulators. Such context-dependent regulation enables a single master regulator to simultaneously control multiple gene-expression programs, which may in turn enhance the precision of cell-fate decisions during development.

Discussion

With the increasing amount of genomic data, a pressing challenge is to understand how the chromatin states are regulated.

Although some computational methods have been developed to model chromatin states across cell-types, such as epi-MARA (14), these studies are only applicable to promoter regions. Here, we developed and experimentally validated a systematic approach to investigate genome-wide epigenetic changes, with a focus on gene distal regions. Specifically, we analyzed extensive public data to obtain a genome-wide profile of chromatin-state plasticity, and focused on the HPRs to gain mechanistic insights. Of note, although we have only analyzed H3K27me3 data, our approach is generally applicable to other epigenetic marks.

Our analysis has provided further evidence supporting a close link between genomic and epigenomic variation (reviewed by refs. 8, 9, and 51). By definition, epigenetic changes occur without change of the genomic sequence. However, distinct genomic sequences may recruit TFs, which in turn may recruit chromatin regulators through protein–protein interactions. From an evolutionary point of view, epigenetic plasticity is advantageous for maintaining phenotypic diversity, but it remains unclear whether epigenetic patterns can be inherited through germ lines. An attractive model is that specific genomic sequence features have evolved that can modulate epigenetic plasticity (52). In support of this model, a recent comparative genomic/epigenomic study has found a correlation between epigenomic and genomic conservation (53). Interestingly, the degree of epigenomic conservation is high at regions characterized by either a high or low substitution rate (quantified by the PhyloP score), but not at regions with moderate substitution rate (53).

The function of the H3K27me3 mark is traditionally associated with promoter silencing, whereas its role in distal regions remain largely unexplored. Here we show that a significant fraction of HPRs are associated with distal regions, and that these regions are strongly enriched in enhancer elements, suggesting that H3K27me3 variation in distal regions contributes to mediating global gene-expression patterns in diverse cell types.

One unexpected outcome of our analysis is the association between ProE-specific z-HPRs with TAL1. Although a role of TAL1 in transcriptional repression was previously implicated in other cellular contexts (45), the mechanistic underpinning remains unknown. Our analysis suggests that TAL1 plays a dual role in transcriptional regulation by activating erythroid lineage genes, while at the same time repressing the expression of genes important for alternative lineage decisions such as cardiomyocytes. Our analysis also suggests that TAL1 may participate in two functionally opposite complexes, an activator complex that contains GATA1 and a repressor complex contains GFI1B and EZH2/PRC2, thereby regulating gene activities in a highly context-dependent manner. Such mechanisms may greatly facilitate the synergy among multiple transcriptional regulators in lineage specification during development. Context-dependent recruitment of Polycomb group complexes may be a general mechanism used by master regulators to activate and repress gene expression within the same cell type.

Chromatin regulators typically interact with each other rather than act alone. It will be interesting to use similar strategies to investigate the covariation patterns among multiple epigenetic marks. As demonstrated in this study, analysis of cross cell-type variation patterns may offer insights that cannot be obtained from examination of a single cell type. Future studies along this direction may help uncover the origin of epigenetic aberrations in pathologic conditions, such as cancers.

Materials and Methods

Quantifying Cross Cell-Type Plasticity of a Histone Mark. H3K27me3 ChIP-seq data in 19 cell lines were obtained from University of California at Santa Cruz ENCODE genome browser (http://genome.ucsc.edu/ENCODE). Raw sequence reads were initially processed by FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) for quality control, and then aligned to the Feb 2009 reference human genome assembly (GRCh37/hg19) by using Bowtie (54) with the "-m 3–strata–best" parameter setting. The aligned sequence reads were then processed by using SAMTOOLS (55) and BEDTOOLS (56), counting reads in 200-bp nonoverlapping bins. Raw sequence read counts were normalized by the total number of reads followed by arcsine transformation (57), defined as $\arcsin\sqrt{P}$ for any value P between 0 and 1, to enhance variance stability. Cross cell-type plasticity of H3K27me3 levels was quantified by using the IOD, defined as the variance divided by the mean value across different cell lines. We selected the 1% bins with highest IOD values and merged adjacent bins into contiguous regions. These regions were referred to as the HPRs. The LPRs were defined in a similar manner using the 1% bins with the lowest IOD values.

Classification of HPRs Using DNA Sequences. Three-thousand HPRs and 3,000 LPRs were randomly selected from the genome for the purpose of model training. Centered at the midpoint of each region, a 256-bp sequence was extracted. Motif scanning was carried out by using the FIMO package of the MEME suite (58) using the standard threshold ($P < 1E\text{-}4$). The background distribution was estimated by randomly selecting regions with matching C+G content. Known TF motifs were obtained from the JASPAR (35), TRANSFAC (34), and FACTORBOOK (36) databases. Statistical significance of motif enrichment was determined using Fisher's exact test (59).

We considered two classification models. First, we built a logistic regression model including quadratic terms, using the C+G and CpG frequencies as predictors. This model is referred to as the CG-only model. Second, to comprehensively integrate information from other sequence features, we applied the N-score model, which was originally developed for prediction of nucleosome positioning (31, 32). In brief, the model integrates three types of sequence features, including sequence periodicities (31), word counts (60), and structural parameters (61), a total of 2,920 candidate features. Model selection was done by stepwise logistic regression. The final model was used for target prediction. These sequences were used as the training set to build classification models. To evaluate classification accuracy, we applied each model to a testing test, which contained 3,000 HPRs and 3,000 LPRs independently chosen as described above. Prediction accuracy was evaluated by using the AUC. To predict genome-wide patterns, we applied the N-score model to moving windows of 256-bp sequences across the genome. The predicted scores were assigned to the center position of each window.

The N-score predicts the log-odds of a DNA sequence being chosen from HPR as opposed to LPR, therefore it has a different dynamic range than the plasticity scores. It is necessary to normalize the N-scores to directly compare with plasticity scores. To maintain the overall shape, we applied the linear transformation: $N\text{-}score_{norm} = 6.8\times10^{-6}\times N\text{-}score_{raw} - 8.2\times10^{-5}$ to normalize the N-score. This normalization does not affect the genome-wide correlation between N-score and the plasticity score.

Definition of High and Low CpG Promoters. Promoters were defined as the transcription start site ± 2,000 bp and were divided into subgroups according to the CpG content as described in the literature (1), with minor modifications. In particular, high CpG promoters were defined as those containing at least one 500-bp contiguous window in which the GC content is greater than 0.55 and the CpG observed/expected ratio is greater than 0.6. All other promoters were called low CpG promoters.

ChIP, ChIP-seq, and Data Analysis. ChIP-seq analyses of H3K27ac, H3K27me3, TAL1, GATA1, and GFI1B were performed using chromatin prepared from primary human ProE cells, as previously described (6). Sequential ChIP analysis was performed as previously described (16, 62). The following antibodies were used: H3K27ac (ab4729, Abcam), H3K27me3 (07-449, Millipore), TAL1 (sc-12984; Santa Cruz Biotechnology), GATA1 (ab11852; Abcam), GFI1B (ab26132; Abcam), and rabbit IgG (12-370; Millipore). Raw ChIP-seq sequence reads were filtered by quality checking and aligned to the reference genome as described above. TF binding peaks were detected by using the MACS software (63), with the default parameter setting. The ChIP-seq data were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession no. GSE52924.

Co-IP and Western Blot. Co-IP experiments were performed using nuclear extracts from ProE cells as described previously (64). The following antibodies were used for co-IP or Western blot analysis: TAL1 (sc-12984; Santa Cruz Biotechnology), GATA1 (ab11852; Abcam), GFI1B (ab26132; Abcam), EZH2 (612666; BD Transduction Laboratories), MTA2 (sc-9447; Santa Cruz Biotechnology), SIN3A (sc-994; Santa Cruz Biotechnology), and rabbit IgG (12-370; Millipore).

Gene Expression. mRNA gene-expression levels in the 19 cell lines, measured by using Affymetrix Human Exon 1.0 GeneChip exon arrays, were obtained from ENCODE (http://genome-preview.cse.ucsc.edu/cgi-bin/hgTrackUi?g=wgEncodeUwAffyExonArray). Probe sequences were mapped to the

current transcription annotations (Refseq hg19) by using a custom cdf file obtained from the BRAINARRAY (http://brainarray.mbni.med.umich.edu). Raw data were background-corrected, normalized, and quantified by using the robust multiarray average procedure (as implemented in the Bioconductor "affy" package) (65). For the ProE cells, the gene-expression levels were measured by using a different microarray platform (Affymetrix human genome U133 Plus 2.0). Therefore, the results were not directly comparable with the ENCODE data. To normalize the array-platform–specific differences, we compared the expression level ranking within each array, and calculated the *z*-scores based on these rankings.

ACKNOWLEDGMENTS. We thank Drs. Zhen Shao and Kimberly Glass for helpful discussion. This work was supported in part by National Institute of Diabetes and Digestive and Kidney Diseases Career Development Award K01DK093543 (to J.X.); and National Institutes of Health Grants R01HL32259 and P01HL03262 (to S.H.O.) and R01HG005085 (to G.-C.Y.). S.H.O. is a Howard Hughes Medical Institute investigator.

1. Mikkelsen TS, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153):553–560.
2. Mohn F, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30(6):755–766.
3. Hawkins RD, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6(5):479–491.
4. Wei G, et al. (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity* 30(1):155–167.
5. Ernst J, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43–49.
6. Xu J, et al. (2012) Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell* 23(4): 796–811.
7. Rada-Iglesias A, et al. (2012) Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* 11(5):633–648.
8. Moazed D (2011) Mechanisms for the inheritance of chromatin states. *Cell* 146(4): 510–518.
9. Yuan GC (2012) Linking genome to epigenome. *Wiley Interdiscip Rev Syst Biol Med* 4(3):297–309.
10. Mendenhall EM, et al. (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* 6(12):e1001244.
11. Orlando DA, Guenther MG, Frampton GM, Young RA (2012) CpG island structure and trithorax/polycomb chromatin domains in human cells. *Genomics* 100(5):320–326.
12. Ku M, et al. (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4(10):e1000242.
13. Liu Y, Shao Z, Yuan GC (2010) Prediction of Polycomb target genes in mouse embryonic stem cells. *Genomics* 96(1):17–26.
14. Arnold P, et al. (2013) Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res* 23(1):60–73.
15. Creyghton MP, et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107(50):21931–21936.
16. Rada-Iglesias A, et al. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470(7333):279–283.
17. Bernstein BE, et al.; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
18. Pal S, et al. (2011) Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res* 21(8):1260–1272.
19. Luco RF, et al. (2010) Regulation of alternative splicing by histone modifications. *Science* 327(5968):996–1000.
20. Mercer TR, et al. (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat Genet* 45(8):852–859.
21. Nag A, et al. (2013) Chromatin signature of widespread monoallelic expression. *eLife* 2:e01256.
22. Khalil AM, et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 106(28):11667–11672.
23. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322(5902):750–756.
24. Martens JH, et al. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24(4):800–812.
25. Ohm JE, et al. (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 39(2): 237–242.
26. Hansen KD, et al. (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43(8):768–775.
27. McLean CY, et al. (2010) GREAT improves functional interpretation of *cis*-regulatory regions. *Nat Biotechnol* 28(5):495–501.
28. Zhu J, et al. (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152(3):642–654.
29. Kanhere A, et al. (2010) Short RNAs are transcribed from repressed Polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell* 38(5):675–688.
30. Pinello L, Lo Bosco G, Hanlon B, Yuan GC (2011) A motif-independent metric for DNA sequence specificity. *BMC Bioinformatics* 12:408.
31. Yuan GC, Liu JS (2008) Genomic sequence is highly predictive of local nucleosome depletion. *PLOS Comput Biol* 4(1):e13.
32. Yuan GC (2009) Targeted recruitment of histone modifications in humans predicted by genomic sequences. *J Comput Biol* 16(2):341–355.
33. Ernst J, Kellis M (2012) ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 9(3):215–216.
34. Matys V, et al. (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1):374–378.
35. Portales-Casamar E, et al. (2010) JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 38(Database issue): D105–D110.
36. Wang J, et al. (2013) Factorbook.org: A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res* 41(Database issue):D171–D176.
37. O'Brien P, Morin P, Jr., Ouellette RJ, Robichaud GA (2011) The Pax-5 gene: A pluripotent regulator of B-cell differentiation and cancer disease. *Cancer Res* 71(24): 7345–7350.
38. Xu CR, Schaffer L, Head SR, Feeney AJ (2008) Reciprocal patterns of methylation of H3K36 and H3K27 on proximal vs. distal IgVH genes are modulated by IL-7 and Pax5. *Proc Natl Acad Sci USA* 105(25):8685–8690.
39. Orkin SH, Zon LI (2008) Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* 132(4):631–644.
40. Wechsler J, et al. (2002) Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* 32(1):148–152.
41. Yu M, et al. (2009) Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* 36(4):682–695.
42. Leask A, Byrne C, Fuchs E (1991) Transcription factor AP2 and its role in epidermal-specific gene expression. *Proc Natl Acad Sci USA* 88(18):7948–7952.
43. Zhou Y, et al. (2011) The Zfx gene is expressed in human gliomas and is important in the proliferation and apoptosis of the human malignant glioma cell line U251. *J Exp Clin Cancer Res* 30:114.
44. Orii N, Ganapathiraju MK (2012) Wiki-pi: A web-server of annotated human protein-protein interactions to aid in discovery of protein function. *PLoS ONE* 7(11):e49029.
45. Van Handel B, et al. (2012) Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. *Cell* 150(3):590–605.
46. Kassouf MT, et al. (2010) Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells. *Genome Res* 20(8): 1064–1083.
47. Wilson NK, et al. (2010) Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 7(4):532–544.
48. Tong B, et al. (1998) The Gfi-1B proto-oncoprotein represses p21WAF1 and inhibits myeloid cell differentiation. *Mol Cell Biol* 18(5):2462–2473.
49. Osawa M, et al. (2002) Erythroid expansion mediated by the Gfi-1B zinc finger protein: Role in normal hematopoiesis. *Blood* 100(8):2769–2777.
50. Saleque S, Kim J, Rooke HM, Orkin SH (2007) Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. *Mol Cell* 27(4):562–572.
51. Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. *Cell* 128(4): 669–681.
52. Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA* 107(Suppl 1):1757–1764.
53. Xiao S, et al. (2012) Comparative epigenomic annotation of regulatory DNA. *Cell* 149 (6):1381–1392.
54. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
55. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
56. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
57. Zar J (1998) *Biostatistical Analysis* (Prentice Hall, Englewood Cliffs, NJ), 4th Ed.
58. Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.
59. Fisher RA (1922) On the interpretation of □2 from contingency tables, and the calculation of *P*. *JR Stat Soc* 85(1):87–94.
60. Peckham HE, et al. (2007) Nucleosome positioning signals in genomic DNA. *Genome Res* 17(8):1170–1177.
61. Lee W, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* 39(10):1235–1244.
62. Furlan-Magaril M, Rincón-Arano H, Recillas-Targa F (2009) Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. *Methods Mol Biol* 543:253–266.
63. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9): R137.
64. Xu J, et al. (2010) Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev* 24(8):783–798.
65. Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31(4):e15.