



Enhancer–silencer transitions in the human genome

Di Huang and Ivan Ovcharenko

Genome Res. 2022 32: 437-448 originally published online February 1, 2022
Access the most recent version at doi:[10.1101/gr.275992.121](https://doi.org/10.1101/gr.275992.121)

References This article cites 54 articles, 10 of which can be accessed free at:
<http://genome.cshlp.org/content/32/3/437.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

License This is a work of the US Government.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

Research

Enhancer–silencer transitions in the human genome

Di Huang and Ivan Ovcharenko

Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20892, USA

Dual-function regulatory elements (REs), acting as enhancers in some cellular contexts and as silencers in others, have been reported to facilitate the precise gene regulatory response to developmental signals in *Drosophila melanogaster*. However, with few isolated examples detected, dual-function REs in mammals have yet to be systematically studied. We herein investigated this class of REs in the human genome and profiled their activity across multiple cell types. Focusing on enhancer–silencer transitions specific to the development of T cells, we built an accurate deep learning classifier of REs and identified about 12,000 silencers active in primary peripheral blood T cells that act as enhancers in embryonic stem cells. Compared with regular silencers, these dual-function REs are evolving under stronger purifying selection and are enriched for mutations associated with disease phenotypes and altered gene expression. In addition, they are enriched in the loci of transcriptional regulators, such as transcription factors (TFs) and chromatin remodeling genes. Dual-function REs consist of two intertwined but largely distinct sets of binding sites bound by either activating or repressing TFs, depending on the type of RE function in a given cell line. This indicates the recruitment of different TFs for different regulatory modes and a complex DNA sequence composition of these REs with dual activating and repressive encoding. With an estimated >6% of cell type–specific human silencers acting as dual-function REs, this overlooked class of REs requires a specific investigation on how their inherent functional plasticity might be a contributing factor to human diseases.

[Supplemental material is available for this article.]

Transcriptional silencers, exerting negative regulatory impact and counterbalancing positive regulatory elements (REs, such as enhancers), have long been recognized to be essential for fine-tuning gene regulation and precisely responding to cellular signals and environmental stimuli in metazoans (Jacob and Monod 1961; Brand et al. 1985; Johnson et al. 2015; Rojano et al. 2019; Halfon 2020). However, because of the inherent difficulty in assaying and identifying silencers, the research focused on silencers has been greatly overshadowed by studies targeting enhancers. With the scarcity of known examples, knowledge about silencers has lagged behind that of enhancers, promoters, and even that of insulators (Ngan et al. 2020). Recently, the advance of massively parallel reporter assays (MPRAs) has greatly facilitated the large-scale detection of silencers in human cells and ignited interest in silencers (Della Rosa and Spivakov 2020; Doni Jayavelu et al. 2020; Pang and Snyder 2020).

Despite recent progress, large-scale silencer maps are still restricted to very few cell lines in humans, specifically HepG2 liver carcinoma and K562 chronic myelogenous leukemia cell lines (Doni Jayavelu et al. 2020; Pang and Snyder 2020). This only offers a limited characterization of the effects of silencers in various biological contexts. Furthermore, although histone H3 lysine 27 acetylation (H3K27ac) and histone H3 lysine 4 mono-methylation (H3K4me1) have been widely used for enhancer identification, the most promising mark for silencer identification, histone H3 lysine 27 trimethylation (H3K27me3), is not specific enough to accurately separate silencers from enhancers (Della Rosa and Spivakov 2020; Gisselbrecht et al. 2020). In human embryonic stem cells (hESCs), almost two-thirds of H3K27me3 chromatin immunoprecipitation sequencing (ChIP-seq) peaks carry signals of the histone marks associated with transcriptional activation. In

primary T cells, a differentiated cell type, half of the H3K27me3 ChIP-seq peaks show the same property. Furthermore, there is a lack of consistency among silencer maps reported by different groups. In the HepG2 cell line, for example, two independent sets of silencers, which were identified using massively parallel platforms and contain about 4000 elements each, have only four silencers in common (Doni Jayavelu et al. 2020; Pang and Snyder 2020). This small overlap could be partially explained by different silencer pooling strategies and suggests an abundance of silencers active in a human cell type. Inspired by the success of sequence-based deep learning models in the prediction of chromatin states, TF binding occupancies, and enhancers (Alipanahi et al. 2015; Zhou and Troyanskaya 2015; Thibodeau et al. 2018), as well as our prior work on silencer detection (Huang et al. 2019), we trained a sequence-based deep learning model to identify silencers and showed that our model greatly outperforms other methods for silencer detection.

After generating accurate genome-wide maps of cell-specific silencers, we decided to focus on dual-function REs (DFREs), which act either as silencers or enhancers in different cell lines and thus represent the versatile component of gene regulatory machinery. Limited DFREs have been reported in *Drosophila melanogaster* (Stathopoulos and Levine 2005) and mice (Bessis et al. 1997; Kehayova et al. 2011). More recently, DFRE abundance in *D. melanogaster* has been reported (Erceg et al. 2017; Gisselbrecht et al. 2020). Almost 25% of nearly 1000 Pho Recessive Complex (PhoRC)–bound silencers in mesodermal cells function as enhancers at other developmental stages (Erceg et al. 2017), whereas 22 out of 29 (76%) identified mesodermal silencers were shown to act as DFREs in large enhancer screens (Gallo et al. 2011; Bonn et al. 2012) and were later augmented by additional six DFREs picked up by a targeted enhancer analysis (thus, bringing the total to 28/29 [Gisselbrecht et al. 2020]). Although the estimates of the fraction of silencers acting as DFREs differed between these two studies, likely

Corresponding author: ovcharen@nih.gov

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.275992.121>. Freely available online through the *Genome Research* Open Access option.

This is a work of the US Government.

because different silencer subclasses were investigated, both studies clearly showed that DFREs are widespread, at least in *D. melanogaster* embryonic cells. In this study, we focused on human DFREs, investigating distinct functional characteristics of these elements in comparison to regular silencers/enhancers and molecular mechanisms enabling their functional transition across cellular contexts.

Results

Convolutional neural network model accurately predicts silencers

We built a multiclass convolutional neural network (CNN) model with three nodes in the output layer representing silencers, enhanc-

ers, and regulatory neutral DNA sequences (see Methods) (Fig. 1A). Our CNN model achieved a respectable prediction performance for silencer and enhancer identification across multiple human cell types. The area under the curve of the receiver operating characteristic (AUC-ROC) ranged from 0.84 to 0.94 for enhancer detection and from 0.74 to 0.90 for silencer detection across six cell types (H1 hESCs, hematopoietic progenitor cells [HPCs], HepG2s, K562s, monocytes, and T cells) (Fig. 1B; Supplemental Fig. S1).

To perform an independent assessment of the accuracy of silencer predictions, we applied our CNN model to experimentally identified silencers. On two sets of silencers active in K562 cells (see Methods), our CNN model achieved the performance of AUC-ROC ≥ 0.78 and AUC-precision recall curve (AUC-PRC)

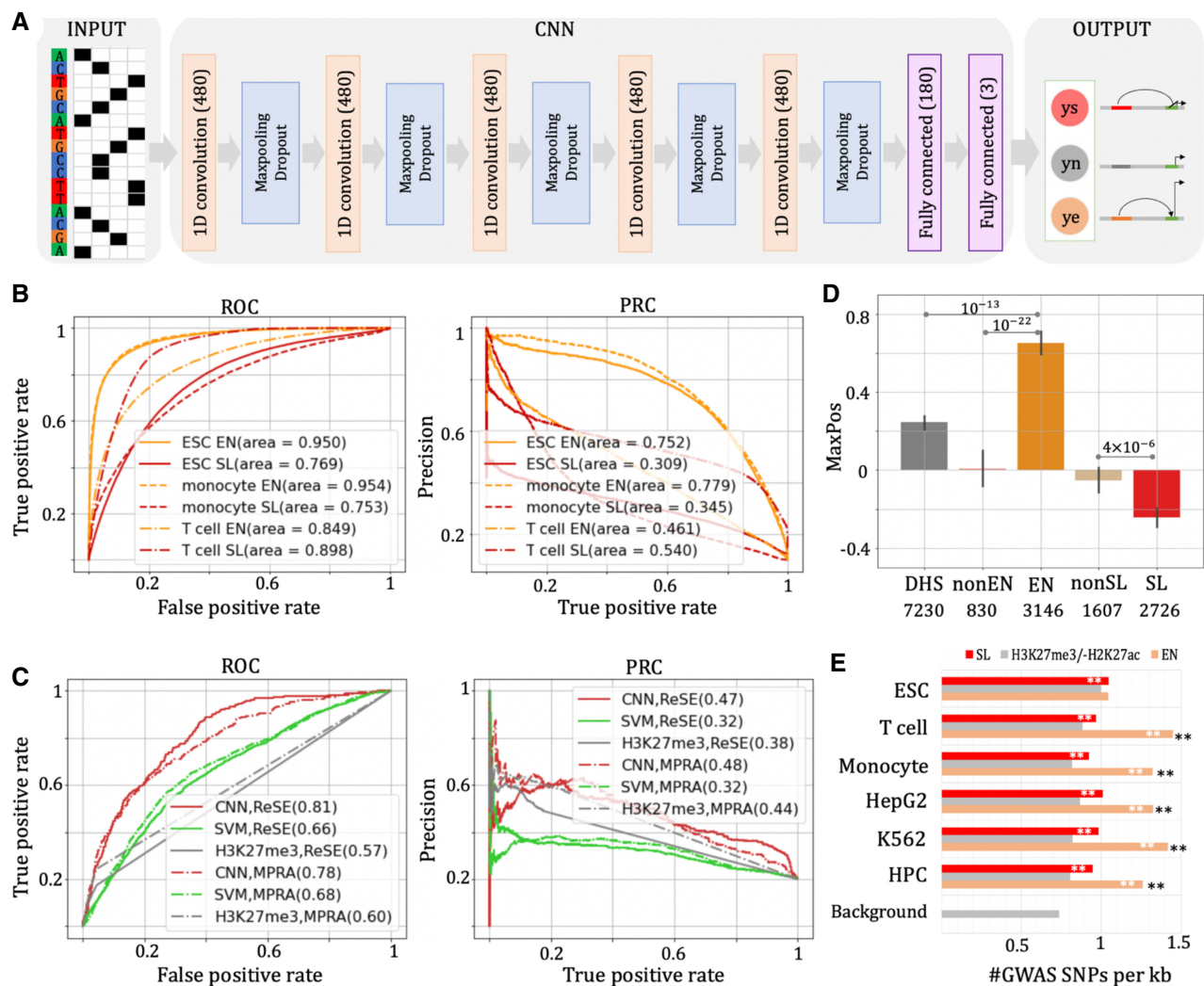


Figure 1. A deep learning model to predict silencers. (A) A schematic of the deep learning model used to predict cell type-specific silencers and enhancers. The number of kernels or neurons in a layer is listed in parentheses. The input of the model consists of 1-kb genomic sequences, and the output is a set of three probabilities of the input sequence being a silencer (ys), an enhancer (ye), or a nonfunctional sequence (yn). (B) ROCs and PRCs of the prediction models in three human cell types. The results in the additional three cell lines are presented in Supplemental Figure S1. (C) ROC and PRC classification accuracy in prediction of ReSE (Pang and Snyder 2020) and MPRA (Doni Jayavelu et al. 2020) experimentally characterized silencers. (D) Sharpr-MPRA MaxPos scores in K562 cells. EN and SL are the predicted enhancers and silencers, respectively. NonEN represents the DNase-seq peaks that overlap with H3K27ac ChIP-seq peaks but have not been predicted as enhancers in this study. NonSL represents the H3K27me3 ChIP-seq peaks not predicted as silencers, and the DHS column corresponds to all DNase-seq peaks. The count under a group label (x-axis) is the size of the corresponding sequence set. (E) The density of GWAS SNPs in the predicted silencers (SL), the sequences with H3K27me3 peaks but no H3K27ac peaks (H3K27me3/-H3K27ac), and the predicted enhancers (EN). The background consists of the genomic sequences having the GC content and repeat density matching to T cell SLs. (***) $P < 10^{-10}$. White asterisks are the significance enrichments of GWAS SNPs compared with H3K27me3/-H3K27ac, and black asterisks are the significant enrichments in enhancers compared with silencers.

≥ 0.47 (Fig. 1C). It outperforms the linear support vector machine (SVM) model we developed previously (AUC-ROC ≥ 0.66 and AUC-PRC = 0.32) (Huang et al. 2019). The improved accuracy of the CNN model in comparison to the SVM model can be attributed to the superiority of CNNs in capturing spatial patterns in sequences and retrieving nonlinear connections among these patterns (Alipanahi et al. 2015; Zhou and Troyanskaya 2015; Koo and Ploenzke 2020). It is important to note that our method accurately predicts experimentally identified ReSE (Pang and Snyder 2020) and MPRA (Doni Jayavelu et al. 2020) silencers (AUC ROC = 0.81 and 0.78 for ReSE and MPRA silencers, respectively) (Fig. 1C). Despite a small overlap between these two data sets (<0.2% of matching sequences), the silencers from these data sets share common DNA sequence motifs, such as the enrichment in TFBSs of the repressive TFs REST and SETDB1 (Supplemental Fig. S2). Silencers predicted by the CNN model also feature similar TFBS enrichment (Supplemental Fig. S2), indicative of this model capturing the motifs shared across different silencer sets and showing a reliable performance on two otherwise distinct silencer sets. There are also differences in the TFBS enrichment profiles of these three silencer sets, which could reflect different types of silencers within them.

To profile human silencers, we applied the trained CNN model to all DNA sequences carrying open chromatin signals (i.e., DNase-ChIP peaks) or repressive histone marks (i.e., H3K27me3 ChIP-seq peaks) and labeled the sequences having a silencer prediction score large enough to correspond to a less-than-0.1 false-positive rate (FPR) on test samples as silencers (see Methods). We also applied this scheme to identify enhancers. To this end, we predicted approximately 130,000 silencers and 120,000 enhancers per cell line (Supplemental Table S1; Supplemental Fig. S3). Among the false-positive silencer predictions, that is, those that carry no H3K27me3 signals but are predicted as silencers, 7.7% correspond to enhancer candidates, that is, the DNase-seq peaks having H3K27ac but no H3K27me3 signals in the corresponding cell type. The remaining 92.3% of false-positive silencer predictions correspond to the background, that is, genomic regions with no H3K27ac, H3K27me3, or DNase signals in the corresponding cell type. In comparison to the test sample sets in which 5.6% are enhancer candidates (Supplemental Fig. S4), these results suggest a bias of false-positive silencer predictions to enhancer candidates. This likely reflects shared sequence motifs of enhancers and silencers active in a particular cell type that correspond to the binding sites of multifunctional TFs (Berest et al. 2019). For example, the NFkB2 complex functions as both activator and repressor in T cells (Senftleben et al. 2001; Grinberg-Bleyer et al. 2018), and its TFBSs are enriched in both T cell silencers and enhancers (see the section “Distinct sequence syntax of the DFREs”).

In a systematic high-resolution activation and repression profiling with MPRA (Sharpr-MPRA), the activating/repressive capabilities of individual nucleotides within REs have been measured by MaxPos scores (Ernst et al. 2016). A low/high MaxPos score indicates a strong repressive/activating regulatory effect. In K562 cells, predicted silencers show the lowest average MaxPos score among all elements, which is 4.7 times lower than the average MaxPos score of H3K27me3 ChIP-seq peaks not predicted as silencers (dubbed nonSLs, Student's *t*-test $P = 4 \times 10^{-6}$) (Fig. 1D). Similarly, predicted enhancers display a higher average MaxPos score than other elements, including H3K27ac ChIP-seq peaks not predicted as enhancers (dubbed nonEN, Student's *t*-test $P = 10^{-22}$) (Fig. 1D). These experimental results provide an independent support of our silencer and enhancer predictions.

In the five tested cell lines (monocytes were not included in this analysis owing to the lack of gene expression data for this cell type), the predicted silencers were enriched in the proximity of low-expression genes (Supplemental Fig. S5). Additionally, the density of single-nucleotide polymorphisms (SNPs) detected in genome-wide association studies (GWASs) is greater in predicted silencers and enhancers than in the sequences carrying H3K27me3 but no H3K27ac ChIP-seq peaks for all examined cell lines but hESCs (binomial test, $P < 10^{-10}$) (Fig. 1E). Furthermore, we observed that silencers show higher genomic mappability than H3K27me3 regions overall (Supplemental Fig. S6). To exclude a potential ascertainment bias arising from mappability differences between different sequence sets, we restricted the GWAS SNP analysis to the elements with at least 50% of certainly mappable nucleotides (i.e., mappability score = 1) (Supplemental Fig. S7). As there is no qualitative change in the results, this further confirms the functional importance of the predicted silencers, at least in comparison to H3K27me3 regions and randomly selected background sequences with matching length, GC content, and repeat density. For simplicity, we refer to these predicted silencers/enhancers as silencers/enhancers for the remainder of the manuscript.

DFREs are associated with TFs and chromatin remodeling genes

In our search for REs capable of alternating between enhancer and silencer function, we focused on H1 hESC enhancers that transition to silencers in primary T cells, a type of differentiated lymphoid cells. Based on overlap with H1 hESC enhancers, the silencers in primary T cells have been split into DFRE and non-DFRE silencers (see Methods). The latter were termed regular silencers (SLRs). To address the function of DFREs, we compared DFREs with SLRs and hESC enhancers that do not function as enhancers or silencers in T cells (named as ENrs below).

In primary T cells, 11,888 silencers (6% of all silencers) are DFREs (Fig. 2A). Although DFREs and SLRs show similar H3K27me3 intensities (Supplemental Fig. S8), 8% of these DFRE sequences are conserved across placental species (Siepel et al. 2005), which is significantly higher than the 4.1% of SLRs and 3.6% of the background (binomial test $P < 10^{-10}$) but lower than the 10.6% of ENrs ($P < 10^{-10}$) (Fig. 2B). Also, the DFREs harbor 5.06 common SNPs per kilobase, which is significantly lower than the 5.44 common SNPs per kilobase in SLRs ($P < 10^{-10}$) and is similar to that of ENrs (Fig. 2B). In addition, 7.96% of DFRE SNPs have a derived allele frequency (DAF) of greater than 0.9, which is significantly lower than the 9.23% of the SLR SNPs ($P < 10^{-10}$) (Supplemental Fig. S9). These results suggest that DFREs have been and are still evolving under strong purifying selection, which is indicative of their functional importance.

Next, we used genes flanking DFREs to assess their potential function. Compared with all silencers, the DFREs show a strong preference for the loci of genes participating in transcriptional regulation (such as mRNA transcription, hypergeometric test $P = 2 \times 10^{-8}$) (Supplemental Fig. S10) and/or genes associated with binding activity (such as T cell receptor binding, $P = 3 \times 10^{-5}$) (Fig. 2C). Overall, 19.2% of DFREs are located in the loci of genes encoding TFs and/or chromatin modifiers (CMs), which is significantly higher than the 16.7% of SLRs (binomial test $P = 10^{-11}$; see Methods) (Fig. 2D). This trend strengthens in the loci of the top 1000 high-expression TFs and/or CMs by 1.3 times ($P = 6 \times 10^{-5}$, high-expression vs. all TFs/CMs). Furthermore, compared with ENrs, DFREs show a similar enrichment in the loci of all TFs/CMs ($P = 0.03$) but a contrasting strong preference to TFs/CMs

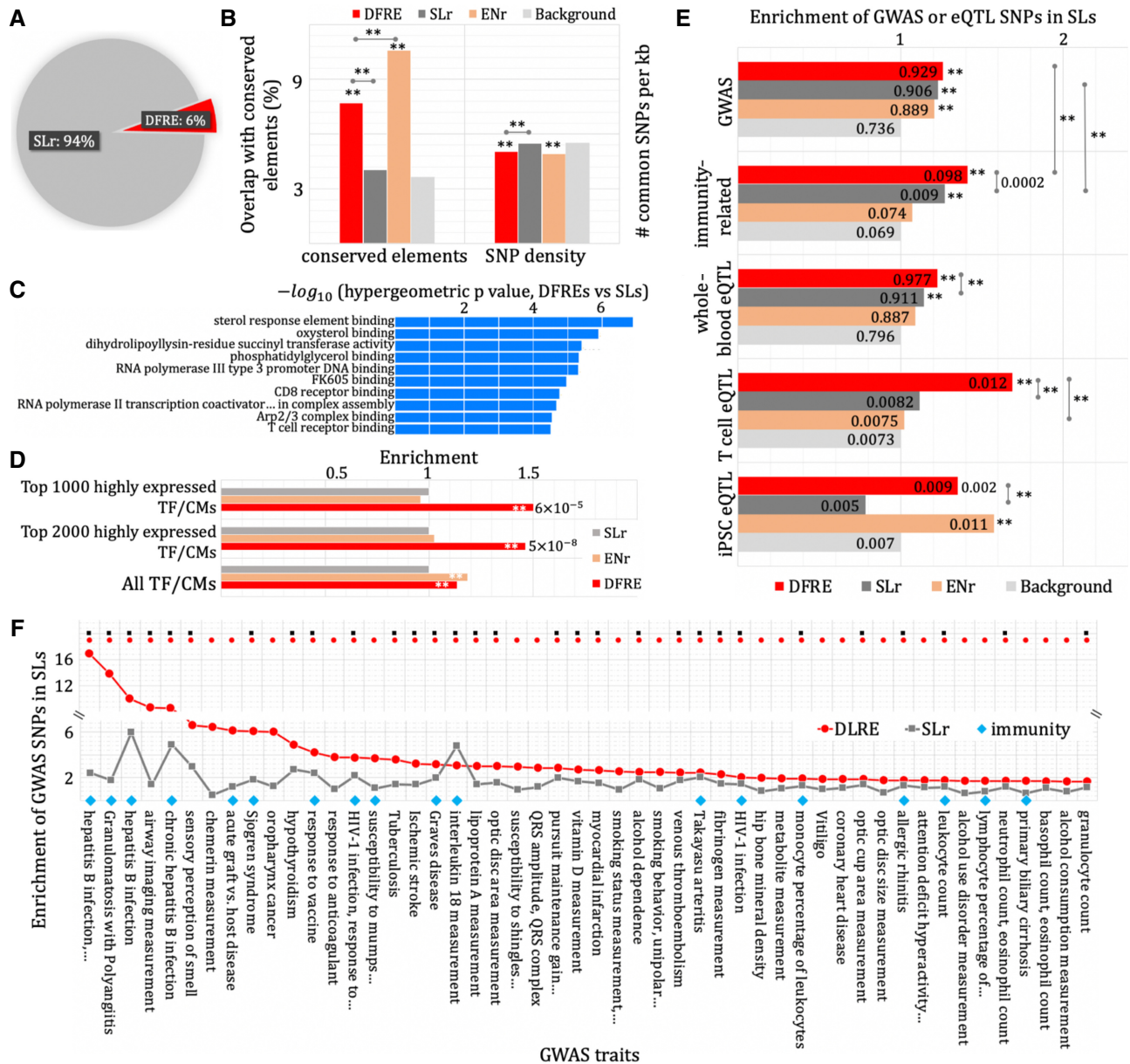


Figure 2. DFREs are consequential. (A) A distribution pie chart of the silencers (SLRs) and DFREs for primary T cells in peripheral blood. (B) Evolutionary sequence conservation of DFREs (red), SLRs (gray), and ENRs (orange) computed using the overlap with conserved segments and the density of common SNPs. The asterisks above the bars indicate the significant enrichments compared with the background consisting of a group of the randomly sampled genomic regions with the GC and repeat contents matching the silencers. (***) $P < 10^{-10}$. (C) Molecular functions significantly associated with the DFREs. The results are from the web tool GREAT using all silencers as background. (D) Enrichment of DFREs in the loci of the TFs and CM genes. (***) $P < 10^{-5}$. White asterisks in the bars represent significant enrichment compared with the SLRs. The presented P -values are the enrichments compared with all TF and CMs. (E) Enrichment of GWAS SNPs and eQTLs in DFREs, SLRs, and ENRs. The density of SNPs per kilobase is listed in the bars. The asterisks (and the values) next to the bars quantify the significance of enrichment compared with the background. (***) $P < 10^{-5}$. (F) GWAS traits with which the associated SNPs are enriched in the DFREs compared with ENRs. Only the GWAS traits significantly enriched in the DFREs are presented. The azure diamonds highlight immunity-related traits. Red and black dots on the top indicate the significant enrichments ($P < 0.05$) of GWAS SNPs in DFREs and SLRs, respectively. The details about these results are listed in Supplemental Table S2.

highly expressed in T cells (binomial test $P = 10^{-7}$) (Fig. 2D), suggesting a distinct association of DFREs with the regulation of the TFs/CMs specific to T cells. The TFs are the core of gene regulation networks, and CM enzymes, which reshape the chromatin topologies globally or locally to facilitate or impede the binding of TFs, are crucial for regulating the responses to cellular and external signals, especially during cell differentiation and in the context of

immunity (Dixon et al. 2015; Tarte and Takeuchi 2015; Chen et al. 2020). The strong association of DFREs with these genes suggests a primary contribution of DFREs to governing the activity and identity of T cells.

Next, we used RNA-seq gene expression (Roadmap Epigenomics Consortium et al. 2015) and Hi-C contact data in hESCs and T cells (Jung et al. 2019; Yang et al. 2020) to address if a transition

in DFRE function might result in change of the DFRE target gene. Among intergenic DFREs that contact only one of their flanking genes (named A) and not the other flanking gene (named X) in hESCs, we identified DFREs that also feature an expression level of A at least fivefold greater than the corresponding X and the average gene expression level in hESCs. This set of DFREs thus constituted the most pronounced DFRE enhancer effects on a specific target gene in hESCs. We then observed that 84.6% of these A genes in T cells are expressed at a less than twofold greater level than both the corresponding X genes and the average gene expression in T cells, and 100% of these A genes are expressed at a less than fivefold greater level than the corresponding X genes and the average gene expression in T cells. This indicates a pronounced drop in the level of A gene expression upon DFRE transition consistent with the enhancer–silencer DFRE change with no change in the target gene. Therefore, these results strongly suggest that the target gene of DFREs remains largely unchanged upon enhancer-to-silencer transition. In addition, using the Hi-C data (Jung et al. 2019; Yang et al. 2020), we identified eight intergenic DFREs targeting a single flanking gene in both hESCs and T cells. All (8/8) of these single target gene contacts remain intact upon DFRE transition from being an enhancer in hESCs to becoming a silencer in T cells (for two examples, see Supplemental Fig. S11), further supporting preservation of DFRE targets post-transition. These particular target genes include well-known T cell differentiation regulators *TPD52* (Kang et al. 2021) and *CHD7* (Koh et al. 2015). The lack of change in chromatin contacts to proximal genes further highlights that the function transition of DFREs is pivotal for precisely regulating the transcription of these T cell differentiation genes during the development of T cells.

DFREs are enriched in the T cell–related genetic variants

To evaluate the influence of DFREs, we next explored GWAS SNPs and their linkage-disequilibrium (LD) counterparts ($r^2 > 0.8$; see Methods). The DFREs and SLRs harbor 0.93 and 0.91 GWAS SNP per kilobase, respectively, which represents a significant enrichment over the 0.74 GWAS SNP per kilobase in the background sequences (i.e., randomly sampled genomic sequences with the length, GC content, and repetitive element density matching T cell silencers; binomial test $P < 10^{-10}$) (Fig. 2E). Furthermore, 10.6% of GWAS SNPs in DFREs are associated with immunity-related traits and 9.8% of SLR SNPs are immunity associated. Both percentages are higher than the 8.5% of ENr SNPs and 9.4% of background SNPs ($P < 10^{-4}$) (Fig. 2E), highlighting the contribution of these silencers, especially DFREs ($P = 0.0002$, DFREs vs. SLRs), to the regulation of the immune system genes. Analyzing individual GWAS traits, we observed that T cell DFREs are specifically enriched in SNPs linked to immune system disorders, such as polyangiitis with granulomatosis, Sjogren syndrome, and type I diabetes mellitus (binomial test $P < 0.05$) (Fig. 2F; Supplemental Table S2). Eighteen out of the top 50 DFRE-associated GWAS traits (36%) are immunity related (Fig. 2F, azure diamonds), which is significant given that there are ~8% of immunity-related traits among all GWAS traits (hypergeometric test $P = 10^{-7}$).

In addition, T cell DFREs host 0.98 whole-blood expression quantitative trait loci (eQTLs) per kilobase, which is significantly higher than the 0.91 in the SLRs, 0.89 in ENrs, and 0.8 in the background sequences (binomial test $P < 10^{-5}$) (Fig. 2E). Furthermore, the density of T cell eQTLs in the DFREs is 0.012 per kilobase, which is 1.5 times that in SLRs ($P = 2 \times 10^{-6}$) and 1.7 times that in ENrs ($P = 2 \times 10^{-8}$) and the background sequences ($P = 9 \times 10^{-9}$).

Compared with whole-blood eQTLs, T cell eQTLs show elevated enrichment in the T cell silencers, especially the DFREs (chi-squared test $P < 10^{-5}$) (Fig. 2E), suggesting strong tissue specificity and pronounced regulatory impact of these silencers on target gene expression in T cells. Also, the DFREs harbor 0.009 eQTL detected in induced pluripotent stem cells (iPSCs) per kilobase, which is similar to that in ENrs (binomial test $P = 0.13$) but is 1.7 times that in SLRs ($P = 2 \times 10^{-7}$) and 1.4 times that in the background ($P = 0.002$). Combined, the enrichment in immunity-associated SNPs and stem-cell-associated SNPs supports the functional duality of the DFREs, a distinct feature among all tested elements.

DFRE mutations are more likely to negatively affect silencer activity than SLR mutations

To further characterize the effects of silencer mutations, we compared the output of our CNN models for wild-type (WT) alleles to those for mutant alleles, defining the CNN-based silencing alteration score (CNN-SAS; see Methods) (Supplemental Fig. S12). A large positive value of CNN-SAS suggests that the corresponding mutation causes a strong decrease in repressive activity. To evaluate CNN-SASs, we first turned to the experimental results in reporter assay quantitative trait locus (raQTL) studies where the regulatory effect of a SNP mutation was quantified by the expression change of the reporter gene (van Arensbergen et al. 2019). Positive raQTL scores represent mutations causing a decrease in reporter gene expression. Using the HepG2 CNN model, we calculated the CNN-SASs of HepG2 raQTL mutations and observed that the absolute values of the CNN-SASs of raQTLs were significantly higher than those of non-raQTLs (Wilcoxon rank-sum test $P = 4 \times 10^{-164}$; see Methods) (Fig. 3A; Supplemental Fig. S13). We also observed a gradual decrease in raQTL scores with an increase in CNN-SASs (Fig. 3B). Among silencer raQTLs with the 5% highest CNN-SASs, 94% had negative raQTL scores, which is a significant enrichment given that 54% of all tested raQTLs had negative raQTL scores (binomial test $P = 3 \times 10^{-25}$) (Fig. 3B). The fraction of negative raQTLs dwindled with the decrease of CNN-SASs, ending at 12% among the raQTLs with the 5% lowest CNN-SASs ($P = 10^{-25}$ compared with the expected on all tested raQTLs). The strong negative correlation between CNN-SAS and raQTL scores validates the ability of our method to quantify the strength of silencer-disrupting mutations.

In T cells, for which raQTL data are not available, we used eQTL data to evaluate CNN-SASs. T cell CNN-SASs are correlated with T cell eQTLs (Spearman's correlation $r = 0.064$, $P = 0.008$) (Fig. 3C), potentially confirming the ability of CNN-SASs to predict the regulatory impact of mutations in T cells. The weak correlation between CNN-SASs and whole-blood eQTLs ($r = 0.003$, $P = 0.35$) (Supplemental Fig. S14) could be attributed to the indirect measurement of regulatory activity using eQTLs, as eQTLs reflect association and not causality of noncoding mutations with gene expression owing to (1) LD in loci containing regulatory variants and (2) a cumulative effect of multiple REs on target gene expression (as opposed to a direct readout of single mutation effects in raQTL experiments).

We next used the T cell CNN-SASs to assess the impact of silencer mutations. A CNN-SAS was considered significant when its absolute value was larger than 1% of those of all possible single-nucleotide mutations in the silencers (Methods) (Supplemental Fig. S15). Compared with the ENrs and background sequences, DFREs and SLRs host more significant CNN-SASs among GWAS SNPs and eQTLs (binomial test $P \leq 0.002$), reflecting the functional importance of the DFREs and SLRs to the regulation of

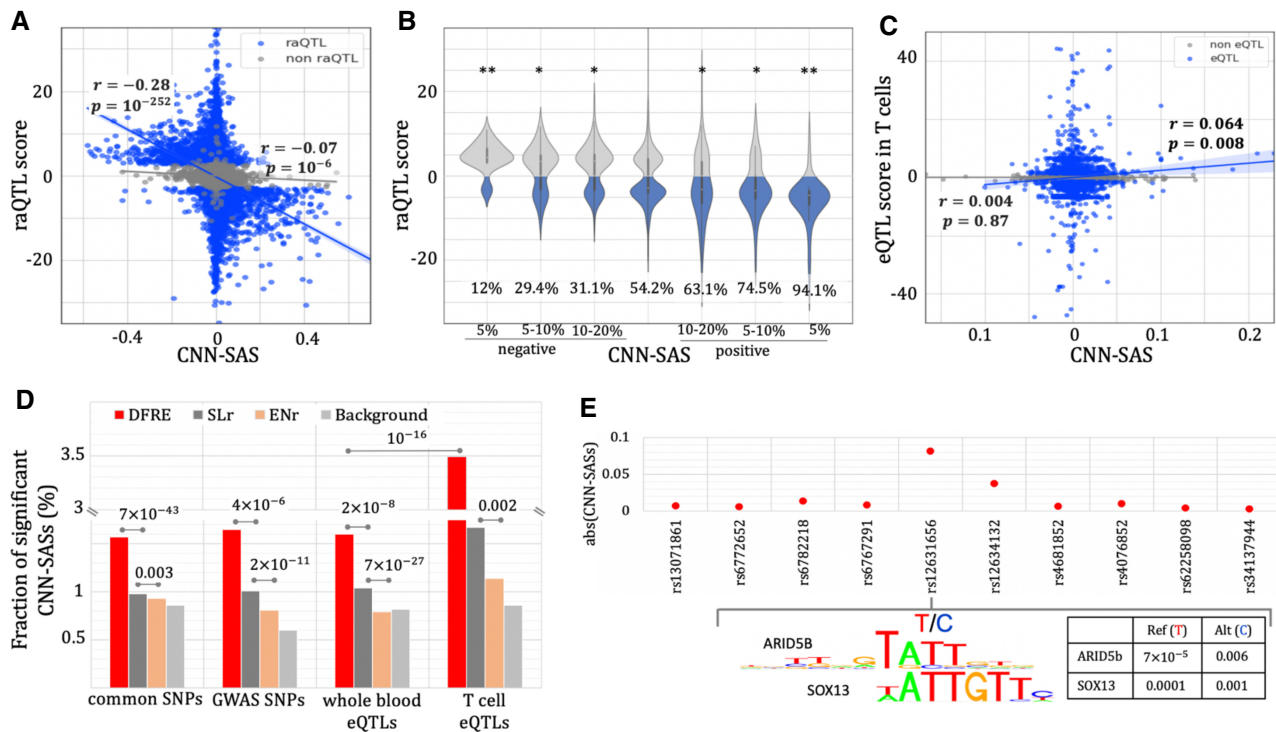


Figure 3. Enrichment of significant CNN-SASs in DFREs. (A) Correlation between CNN-SASs and raQTL scores on the raQTL SNPs (blue dots) and non-raQTL SNPs (gray dots). (B) Distribution of raQTL scores in mutation groups binned based on CNN-SASs. The number under each box is the fraction of negative raQTL scores. A negative raQTL score represents the increase in silencing activity. (***) $P < 10^{-10}$ and (*) $P < 10^{-2}$ represent the enrichment significance in negative raQTL scores compared with all tested mutations. (C) Correlation between CNN-SASs and eQTL scores in T cells. (D) Fraction of significant CNN-SASs across GWAS and eQTL sets of SNPs. (E) The CNN-SASs of the SNPs in a LD block associated with systemic lupus erythematosus and systemic sclerosis. Rs12631656, a DFRE SNP, is the only one having a significant CNN-SAS score. The table lists the mapping significance levels (i.e., FIMO significance P -values) of the ARID5B and SOX13 binding motifs for the sequences carrying the reference or alternative allele at rs12631656.

T cell phenotypes and gene expression. In addition, compared with SLRs, the DFREs are enriched for significant CNN-SASs (i.e., potential damaging mutations) across a panel of mutation sets (binomial test $P < 10^{-5}$) (Fig. 3D). For example, in DFREs, 3.5% of T cell eQTLs SNPs correspond to a significant CNN-SAS (that is 3.5-fold greater than expected by chance; $P < 10^{-10}$). These results validate the utility of significant CNN-SASs in detecting most likely silencer-damaging mutations.

An example DFRE mutation having a significant CNN-SAS is rs12631656. Its T-to-C (i.e., major to minor allele) mutation corresponds to a CNN-SAS of 0.0816 ($P = 0.003$) (Fig. 3E). This SNP is located within a tight LD block ($r^2 > 0.96$) of rs6445975 and rs4681851, the two SNPs that have been reported to be significantly associated with systemic lupus erythematosus and systemic sclerosis (Harley et al. 2008; Mayes et al. 2014). This LD block hosts five DFRE SNPs, four SLr SNPs, and one enhancer SNP in T cells. Among these SNPs, rs12631656 is the only one with a significant CNN-SAS (Fig. 3E). The TF motif mapping shows that the mutation at rs12631656 potentially weakens the binding affinity of ARID5B and SOX13, two known repressors in T cells (Lefebvre 2010; Wang et al. 2020).

Distinct sequence syntax of the DFREs

To address function encryption in DFRE sequences, we considered two primary hypotheses: (1) there is a single set of TF binding sites (TFBSs) within each DFRE bound by a stable set of TFs with alternating repressor and activator functions, and (2) there are two dis-

tinct sets of TFBSs within DFREs—one encoding silencer function and another encoding enhancer function. To test these hypotheses, we mapped the TFBSs in TF ChIP-seq peaks (see “TFBS prediction in TF ChIP-seq peaks” in the Supplemental Notes). With these TFBSs, we first characterized TFBS signatures of the DFREs for their activating and repressive functions. In T cells, although inactive ENrs are depleted of ChIP-seq TFBSs of all TFs, silencers (DFREs and SLRs) are depleted of ChIP-seq TFBSs of activators (such as JUND, IRF3, and IRF4) but are enriched in ChIP-seq TFBSs of ubiquitous repressors, including REST and repressors acting predominantly in blood cells, such as EBF1 (Fig. 4A; Györy et al. 2012), which is in line with their repressive function. Functioning as enhancers in hESCs, DFREs and ENrs show a similar TFBS profile. They are depleted of ChIP-seq TFBSs of REST ($P \leq 0.02$) but are enriched in ChIP-seq TFBSs of hESC-specific TFs, such as POU5F1 and NANOG (Fig. 4B). These results suggest that DFREs recruit a distinct set of TFs for activating and repressive functions. Furthermore, 70% of silencer TFBSs are silencer specific, and 45% of enhancer TFBSs are enhancer specific (Fig. 4C), indicating a presence of distinct DFRE fragments that establish activating and repressive functions.

The enrichment of CTCF ChIP-seq TFBSs in DFREs prompted us to compare DFREs and insulators. In hESCs, where the topologically associated domains (TADs) have been reported (Liu et al. 2019), 0.84% of the DFREs are located at the TAD boundaries, which are known to be enriched for insulators (Ong and Corces 2014). This fraction is comparable to 0.94% of all DNase-seq peaks ($P = 0.25$) and is significantly lower than the 1.12% of CTCF ChIP-

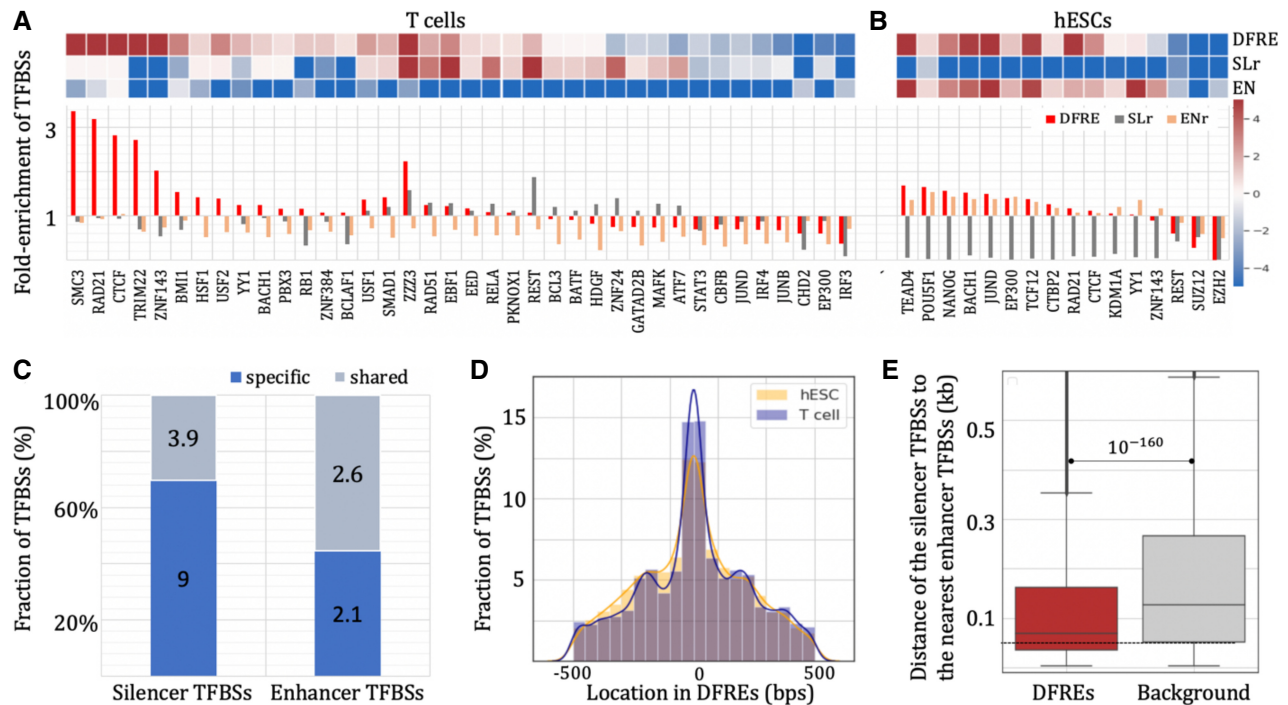


Figure 4. Distinct binding syntax of the DFREs. Enrichment of ChIP-seq TFBSs in T cells (A) and hESCs (B). The heatmap on top illustrates enrichment/depletion significance ($-\log_{10}$ binomial test P) in DFREs, SLRs, and ENRs compared with all DNase-seq and H3K27me3 ChIP-seq peaks in the corresponding cell type. The red and blue shades represent significance levels of enrichment and depletion, respectively. (C) Functional specificity of silencer and enhancer TFBSs in these DFREs. According to their activity in T cells and H1 hESCs, these TFBSs are categorized into two groups: function specific (i.e., unique to one cell type) and shared (i.e., shared by two cell types). The numbers in the bars are the average numbers of TFBSs per DFRE. (D) Distribution of silencer TFBSs (the blue bars and line) and enhancer TFBSs (the orange bars and line) in DFREs. (E) Distance of silencer TFBSs to their nearest enhancer TFBSs. The background distribution was generated through randomly scattering silencer and enhancer TFBSs within DFREs.

seq TFBSs ($P = 10^{-11}$) (Supplemental Fig. S16A). These trends were also observed in T cells, where 0.95% of DFREs, 0.93% of all DNase-seq peaks ($P = 0.34$ vs. DFREs), and 1.31% of CTCF ChIP-seq TFBSs ($P = 10^{-11}$ vs. DFREs) (Supplemental Fig. S16B) are located within the T cell TAD boundaries detected in Yang's study (Yang et al. 2020). Also, in T cells, DFREs are enriched in the loci of the 1000 lowest expressed genes (9.1% of DFREs vs. 6.3% of the whole genome, binomial test $P = 5 \times 10^{-20}$) (Supplemental Fig. S16C) and are depleted in the loci of the 1000 highest expressed genes ($P = 2 \times 10^{-5}$). A reverse trend was observed for DFREs in hESCs ($P \leq 0.001$) (Supplemental Fig. S16D). Similarly, the DFREs frequently target the lowly expressed genes in T cells ($P = 10^{-6}$) (Supplemental Fig. S16E) but highly expressed genes in hESCs ($P = 4 \times 10^{-15}$) (Supplemental Fig. S16F). These results are consistent with the silencer function of DFREs in T cells and the enhancer function of DFREs in hESCs. On the other hand, CTCF ChIP-seq TFBSs show no consistent correlation with either highly or lowly expressed genes in both cell types (Supplemental Fig. S16), supporting a functional distinction between DFREs and CTCF-defined insulators.

We also observed that the majority of DFRE TFBSs, specifically 60% of enhancer TFBSs and 62% of silencer TFBSs, are located within the central component (± 200 bp from the midpoint) of the DFREs, which is significantly $>43\%$ of randomly scattered TFBSs (binomial test $P < 10^{-10}$) (Fig. 4D). Furthermore, 37% of the silencer TFBSs are located within 50 bp to their nearest enhancer TFBSs, which is much higher than 24% of randomly scattered TFBSs in DFREs (binomial test $P = 10^{-160}$) (Fig. 4E). Combined, these results advocate for intertwined and spatially proximal distribution of distinct activating and repressive TFBSs within DFREs.

Next, to account for TFs not yet profiled in ChIP-seq experiments, we used CNN-SASs to predict TFBSs as short segments over-represented in high CNN-SASs (see "TFBS prediction using a CNN model" in the Supplemental Notes). More than 43% of the CNN-predicted TFBSs overlap the reported TF ChIP-seq peaks in both T cells and hESCs, which is more than 1.4 times that of randomly scattered TFBSs (Supplemental Fig. S17A) and, therefore, justifies the use of this predictive approach for additional TFBS discovery. Predicted TFBSs in T cell (silencer) DFREs are enriched with the binding motifs of repressors, such as SNAI1/2, REST, and LMO2 (Supplemental Fig. S17B). Predicted TFBSs in hESC (enhancer) DFREs host binding motifs of developmental activators, including POU5F1, NANOG, and SOX4 (Supplemental Fig. S17C). The CNN-predicted TFBSs confirm and further strengthen the trends observed using ChIP-seq TF data. Fifty-nine percent of these predicted TFBSs reside within the central region (± 200 bp from the midpoint) of DFREs, which is significantly more than the expected 45% of randomly scattered TFBSs (binomial test $P = 10^{-323}$). Also, 43% of silencer TFBSs are located within 50 bp of enhancer TFBSs, which is significantly more than the expected 32% of randomly scattered TFBSs ($P = 10^{-323}$) (Supplemental Fig. S17). Again, these results suggest a tight intertwining between TFBSs active for opposite functions within DFREs as well as a "co-encription" within the central regions of DFREs.

Developmental dynamics of DFRE formation

To understand the progression in DFRE formation during development, we applied our analysis pipelines for identifying DFREs in primary HPCs, which represent an intermediate step in

differentiation from embryonic stem cells to T cells. Twenty-seven percent of HPC silencers are DFREs, acting as enhancers in hESCs (Fig. 5A). Compared with HPC SLrs, these DFREs are more conserved across placental species (10.2% of DFREs vs. 7.4% of SLrs, binomial test $P < 10^{-10}$) and feature a lower density of common SNPs (4.93 SNPs/kb in DFREs vs. 5.45 SNPs/kb in the SLrs, $P < 10^{-10}$) (Fig. 5B). HPC DFREs also harbor 0.98 GWAS SNPs and 1.3 whole-blood eQTLs per kilobase, which is significantly higher than the 0.87 GWAS SNPs and 0.94 whole-blood eQTLs in the SLrs, 0.91 GWAS SNPs and 0.96 whole-blood eQTLs in the ENrs (i.e., the hESC enhancers acting neither as silencers nor as enhancers in HPCs), and 0.74 GWAS SNPs and 0.8 whole-blood eQTLs in the background sequences ($P < 10^{-5}$) (Fig. 5C). HPC DFREs are more highly enriched in GWAS and eQTL SNPs than the SLrs, ENrs, and background in both tested cell lines, suggesting their functional importance across cellular contexts. Overall, HPC DFREs feature functional characteristics very similar to T cell DFREs, with the main exception of approximately four times as many HPC silencers being DFREs compared with T cell silencers (binomial test $P < 10^{-5}$). Furthermore, among 18,965 HPC DFREs, 11.6% (2201) are the DFREs in T cells, which is significantly lower than the 18.6% of HPC SLrs and 46.5% of HPC enhancers shared by T cells (binomial test $P < 10^{-5}$) (Fig. 5D). The shrinking fraction of DFREs during differentiation and the strong cell specificity of DFREs in a mature cell suggest the loss of functional plasticity in REs upon differentiation and a relatively small fraction of REs preserving multifunctional activity after the differentiation.

Discussion

Advances in high-throughput sequencing and MPRA vastly increase the knowledge of gene regulation in human cells. However, the detection and functional characterization of active

silencers remain challenging (Della Rosa and Spivakov 2020; Halfon 2020). In this study, we developed a deep learning silencer model that accurately detects experimentally identified silencers and quantifies the impact of silencer mutations. Identified silencers predominantly populate the loci of low-expression genes and are enriched in GWAS SNPs associated with the cell type matching disease etiology. In principle, silencers exert the repressive impact mainly through reducing the activity of basal promoters or smothering the activity of proximal enhancers (Gisselbrecht et al. 2020). H3K27me3 modification has been evidenced to take part in both repressive mechanisms (Ogiyama et al. 2018; Cai et al. 2021). As such, the CNN models, which were built with H3K27me3 ChIP-seq peaks as training silencer samples, predict both types of silencers. These predicted silencers can be further categorized by using chromatin interaction maps with the knowledge that the silencers that reduce promoter activity interact with promoters more frequently than the other silencers. Also, our models, centered on the H3K27me3-associated silencers, might be less accurate in identifying silencers not marked by H3K27me3. Further studies are needed for investigate the differences and similarities between H3K27me3-based and H3K27me3-independent silencers.

By combining silencer detection with enhancer detection, we observed enhancer–silencer transitions and found that 6% of the T cell silencers and 28% of HPC silencers are DFREs, functioning as enhancers in hESCs and changing their regulatory function during differentiation. Compared with regular silencers, DFREs feature greater evolutionary sequence conservation and are enriched in GWAS SNPs and eQTLs. Moreover, DFREs are preferentially distributed in the proximity of genes governing transcriptional regulation. DFRE mutations are more than 1.5 times as likely to significantly damage silencer activity (as measured using CNN-SASs) as regular silencer mutations. These results support the essentiality and a rather overlooked important role of DFREs

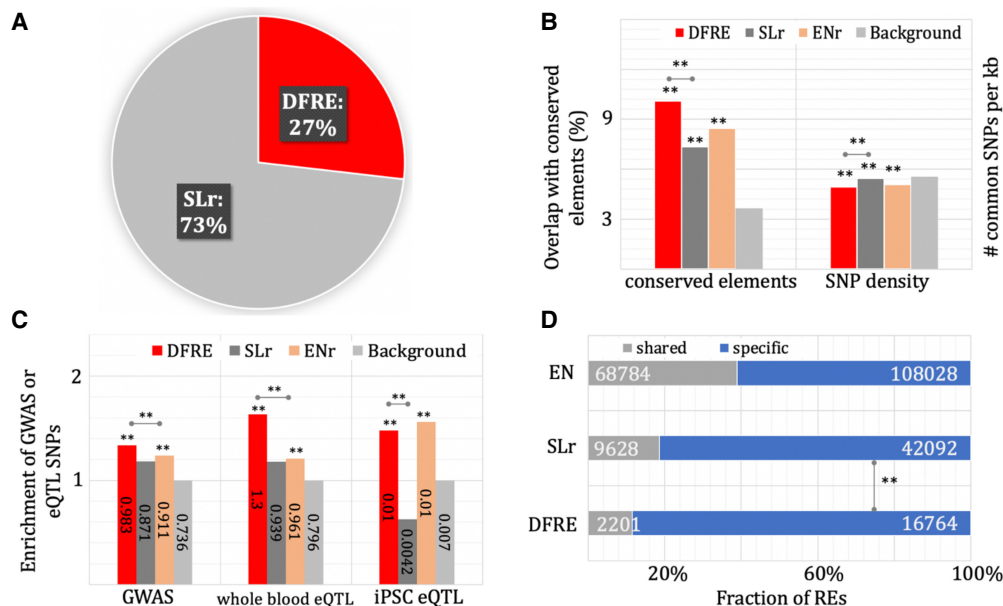


Figure 5. Characterization of DFREs detected in HPCs. (A) Fraction of HPC silencers working as DFREs. (B) Evolutionary sequence conservation measured using the overlap with the conserved segments (*left*) and common SNP density (*right*) of HPC DFREs, SLrs, and ENrs. The background in Figure 2 is also included here (represented by “background”) for consistency. (C) Enrichment of GWAS SNPs and eQTLs in the HPC. The number of SNPs per kilobase is listed in the bars. (B,C) The asterisks *above* the bars represent the significance levels compared with the background. (D) Developmental specificity of DFREs, SLrs, and enhancers in HPCs. The numbers in bars are the numbers of REs. (shared) REs shared in HPCs and T cells, (specific) REs acting in HPCs but not in T cells. (**) $P < 10^{-5}$.

in regulating the development and maintaining the fate of blood cell types (and, likely, other differentiated cell types not included in this study). Earlier studies in *D. melanogaster* embryonic cells, coupled with the scope of our study targeting a single differentiation path, suggest that a large fraction of developmental enhancers might be wired to act as silencers upon and after differentiation (Erceg et al. 2017; Gisselbrecht et al. 2020).

To gain insights into the mechanisms of DFRE functional duality, we compared the distribution of active TFBSs corresponding to silencer and enhancer functions of DFRE sequences. We observed largely distinct but proximal DFRE regions occupied by binding sites of opposing functions. Our findings suggest a “double” genomic regulatory encryption within DFRE sequences with independent active regulatory binding sites, reflective of the use of different TFs for establishing activating and repressive functions after effectively rejecting an alternative hypothesis of the same pool of TFs bound to DFREs and switching their function based on cellular context.

Silencers are an integral part of the regulatory machinery in metazoans shaping and fine-tuning gene regulatory programs. The difference in the function of DFREs reported in *D. melanogaster* was primarily attributed to different cell types. In our study, we focus on the functional switch from enhancer to silencer activity, which is associated with the development and maturation of T cells. Therefore, it is logical to assume that the presented DFREs are more likely to be associated with developmental abnormalities and carcinogenesis. It is also possible that a subset of reported DFREs might underlie the environmental response, and their mutations are more likely to have a phenotypic effect owing to their role in development. Here we report thousands of DFREs during development of T cells and uncover unique features and profound phenotypic/pathogenic contributions of these elements. These findings together suggest the wide spread of DFREs in various biological processes across species. Follow-up studies extending to other biological contexts and additional chromatin data will further characterize human DFREs and help understanding how the same parts of the genome are reused for multiple regulatory functions at different developmental time points and in different cellular contexts. This is a crucial step toward revealing how gene regulation swiftly and precisely adapts to changing of cellular contexts in various biological processes.

Methods

Building CNN models to predict silencers

We designed a deep learning model composed sequentially of five convolutional layers and two fully connected network layers (Fig. 1A). Each CNN layer is followed by a max-pooling and a dropout layer. The input of this model is 1000-bp-long DNA sequences, whereas the output layer contains three nodes, each representing one of the sequence classes: silencer, enhancer, and background. Each output node predicts the probability of a given sequence belonging to the corresponding sequence classes. We used the Python library Keras version 2.4.0 (<https://github.com/keras-team/keras>) to implement our model. To deal with this multiclass classification task, categorical cross-entropy was used as the cost function. By minimizing this cost function, the parameters of the models were adjusted by using the gradient-based algorithm Adam and then Adagrad as implemented in Keras (see “CNN model” in the Supplemental Notes; Supplemental Fig. S18).

We built a CNN model for each of six cell types, including H1 hESCs, the K562 cell line, the HepG2 cell line, primary HPCs, pri-

mary monocytes, and primary T cells from peripheral blood. During training, we set aside Chromosome 6 for validation and Chromosomes 7 and 8 for testing. All other autosomes and Chromosome X were used for training.

To identify silencers, we calculated the cutoff in the silencer prediction scores (silencer scores) generated by the CNN model, which corresponds to the FPR of 0.1 in test samples, among which background samples were randomly selected so that the ratio of background samples to enhancers/silencers was 9:1 in each tested cell type. Sequences that carry a DNase-seq peak or a H3K27me3 ChIP-seq peak and have a silencer score greater than this cutoff were labeled as silencers. A similar approach with the FPR of 0.1 in CNN model enhancer prediction scores (enhancer scores) was used for enhancer identification. DNase-seq peaks that carry a H3K27ac signal and have an enhancer score greater than this cutoff established the set of predicted enhancers. After applying this scheme to all tested cell types, we identified approximately 130,000 silencers and 120,000 enhancers per cell type (Supplemental Table S1). The size of these enhancer sets is comparable to those identified by ChromHMM (Ernst and Kellis 2017) as well as those collected in the Benton and Gao databases (Benton et al. 2019; Gao and Qian 2020).

Data for training and validating CNN models

We trained a CNN enhancer/silencer model independently for each cell type. Cell type-specific DNase-seq and histone modification ChIP-seq data sets have been used for training (all ChIP-seq data used in this study were downloaded from the NIH Roadmap Epigenomics Project) (Roadmap Epigenomics Consortium et al. 2015). H3K27me3 ChIP-seq peaks with neither H3K27ac nor H3K4me1/3 ChIP-seq peaks overlapping their 400-bp central region constituted the set of training silencers. DNase-seq peaks with an overlapping H3K27ac peak but no overlapping H3K27me3 ChIP-seq peaks within their 400-bp central region constituted the set of training enhancers. Background sequences were randomly selected from DNase-seq peaks and H3K4me1, H3K4me3, H3K27ac, and H3K27me3 histone modification ChIP-seq peaks detected in any cell line other than the tested cell line. CNN models were trained on 1-kb regions centered on enhancer, silencer, and background sequences. For T cells, we collected 416,773 DNase-seq peaks and 392,442 H3K27me3 and 129,500 H3K27ac “broad” peaks from the NIH Roadmap Epigenomics Project (Roadmap Epigenomics Consortium et al. 2015). This translated into 239,925 and 158,887 nonredundant putative silencer and enhancer sequences, respectively. The CNN models identified 219,971 T cell silencers and 156,451 T cell enhancers from these sequences.

In the K562 cell line, experimentally validated silencers were used to evaluate the accuracy of the silencer prediction model. Those silencers were acquired from two resources: 3796 K562 silencers identified with self-transcribing active regulatory region sequencing platform (STARR-seq) (Doni Jayavelu et al. 2020) and 3909 K562 silencers reported using the repressive ability of silencer element assay (ReSE) (Pang and Snyder 2020). After redundancy exclusion, we assembled 7701 distinct silencers experimentally validated in K562 cells. To apply the CNN model in which the input is 1000-bp sequences on these silencers, we extended these silencers into 1000-bp genomic sequences centered at the midpoint of tested segments. To prevent a bias from prior knowledge on the estimate of the model accuracy, we excluded all 1406 experimentally validated silencers that overlap training or validation sequences (either for building the SVM or the CNN model) from the computational model evaluation.

In addition, the results of Sharpr-MPRA were used to evaluate silencer/enhancer predictions (Ernst et al. 2016). A low Sharpr-

MPRA MaxPos score represents a negative RE. In this study, a genomic sequence was assigned with a MaxPos score only when this sequence hosted the whole corresponding Sharp-mpRA sequence.

Enhancer–silencer transitions during development

To identify enhancer–silencer transitions during development, we compared silencer profiles in mature cell types with the enhancer map in H1 hESCs. A T cell silencer overlapping an H1 hESC enhancer by >600 bp was considered a DFRE. During embryonic development, the differentiation of T cells is a stepwise process from ESCs to mesodermal cells, HPCs, common lymphoid progenitors, and finally T cells (Kumar et al. 2018). In adulthood, T cells are differentiated from HPCs.

Evaluating the repressive effect of mutations

To evaluate the impact of silencer mutations, we focused on ys and ye in a built CNN model (Fig. 1A) as they predict the capability of genomic sequences being silencers and activators, respectively. Given the WT and mutant allele (MU) of a mutation, the difference of ys and ye between these alleles directly measures the alteration of activating and repressing power of the host sequence. That is, the CNN-SAS of a given mutation is defined as

$$CNN - SAS = (ys_{WT} - ys_{MU}) - (ye_{WT} - ye_{MU}), \quad (1)$$

where ys_{WT} and ys_{MU} are the probabilities of silencer activity in the WT and the MU sequences, respectively. ye_{WT} and ye_{MU} are the probabilities of enhancer activities of the corresponding sequences. In addition, we measured the silencing alteration caused by a mutation with the odd ratios of $(ys_{WT} - ye_{WT})$ to $(ys_{MU} - ye_{MU})$ after deriving the probability function of $(ys - ye)$ and noticed the similar performance between CNN – SAS scores and odd-ratio scores (see “CNN-based silencing odds ratio of mutations” in the Supplemental Notes; Supplemental Fig. S12).

The CNN – SAS distribution of all possible single-nucleotide silencer mutations forms a T distribution with 69% of mutations having an absolute CNN – SAS value smaller than 0.01 (Supplemental Fig. S15). Using this distribution as a background, we evaluated the significance of observed CNN – SAS scores. A low value of the probability represents a mutation having a strong impact on the activity of the host silencer. A CNN – SAS score was considered significant if this probability was <1%.

Data and tools used for analysis

We downloaded the GWAS SNPs curated in the National Human Genome Research Institute (NHGRI) catalog as of January 2019 (Buniello et al. 2019). There were 70,578 unique SNPs in the catalog at that time. To account for the fact that GWAS SNPs might not be directly responsible for phenotypic alterations but might be in LD with untested causal genetic variants (Cantor et al. 2010), we further expanded the SNP set by adding SNPs in a strong LD ($r^2 > 0.8$) block with GWAS SNPs in at least one population from the 1000 Genomes Project Consortium. To that end, we acquired 1,409,462 GWAS SNPs associated with 2722 traits. SNPs showing a strong phenotypic influence are likely to be associated with multiple traits according to independent studies (Zhou et al. 2018). To account for this, GWAS SNPs were linearly weighted by the number of the traits linked to them when we evaluated the GWAS SNP density. The GWAS traits were identified as immunity related when they contain keywords such as allergy, arthritis, asthma, Graves disease, infection, inflammation, lupus, multiple sclerosis, Sjogren syndrome, type I diabetes mellitus, etc.

Whole-blood eQTLs used in this study were downloaded from the Genotype-Tissue Expression Project, GTEx database version 8

(The GTEx Consortium 2015). To further evaluate the tissue specificity of the predicted DFREs, we collected eQTLs detected in T cells from the Blueprint epigenome project (Chen et al. 2016). eQTLs in iPSC were downloaded from the study of DeBoever et al. (2017).

We downloaded 14,183 HepG2 raQTLs from the study (van Arensbergen et al. 2019). These SNPs are associated with a significant change in reporter gene activity in HepG2 cells. We also downloaded additional 14,183 SNPs that correspond to an insignificant change in the expression of reporter genes (i.e., non-raQTL mutations) as analysis background.

We used the Genomic Regions Enrichment of Annotation Tool (GREAT) (McLean et al. 2010) to evaluate Gene Ontology (GO) biological processes and molecular functions associated with different RE groups. The element–gene association setting was selected as “two nearest genes.”

We explored conserved elements predicted using the 46-way placental phastCons scores (Siepel et al. 2005). A notable overlap of a genome sequence with conserved elements is indicative of an evolutionary constraint imposed on that sequence. The density of SNPs in genomic sequences was used to assess the evolutionary pressure acting on genomic segments in the human lineage. A low SNP density corresponds to strong selective pressure. The SNPs reported by The 1000 Genomes Project Consortium et al. (2015) were used for this analysis.

HESC TAD boundaries were downloaded from the Topologically Associating Domain Knowledge Base (TADKB) database in which the boundaries, in the resolution of 10 kb, were detected by using the Gaussian mixture model and proportion test (Liu et al. 2019). The chromatin contacts in hESCs were downloaded from the database assembled in the study of Jung et al. (2019). The TADs and genomic contacts in T cells were downloaded from the report of Yang et al. (2020). We used the detections for unstimulated T cells. The TAD boundaries were defined as the 10-kb-long genomic regions centering at all the ends of TADs.

Software availability

Custom Python scripts (training CNN models and predicting enhancers/silencers with a built CNN model) are available on GitHub (<https://github.com/ncbi/SilencerEnhancerPredict>) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Dorothy L. Buchhagen, Chris Hill, Sanjarbek Hudaiberdiev, and Wei Song for critical reading of the manuscript. This research was supported by the Intramural Research Program of the National Library of Medicine and the National Human Genome Research Institute, National Institutes of Health (NIH). This work used the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

Author contributions: D.H. performed computational analysis and analyzed the data. I.O. supervised computational work. D.H. prepared figures and supplemental tables. D.H. and I.O. wrote the manuscript.

References

The 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean

- GA, Abecasis GR, et al. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74. doi:10.1038/nature15393
- Alipanahi B, Delong A, Weirauch MT, Frey BJ. 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* **33**: 831–838. doi:10.1038/nbt.3300
- Benton ML, Talipineni SC, Kostka D, Capra JA. 2019. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC Genomics* **20**: 511. doi:10.1186/s12864-019-5779-x
- Berest I, Arnold C, Reyes-Palomares A, Palla G, Rasmussen KD, Giles H, Bruch P-M, Huber W, Dietrich S, Helin K, et al. 2019. Quantification of differential transcription factor activity and multiomics-based classification into activators and repressors: diffTF. *Cell Rep* **29**: 3147–3159.e12. doi:10.1016/j.celrep.2019.10.106
- Bessis A, Champiaux N, Chatelin L, Changeux J-P. 1997. The neuron-restrictive silencer element: a dual enhancer/silencer crucial for patterned expression of a nicotinic receptor gene in the brain. *Proc Natl Acad Sci* **94**: 5906–5911. doi:10.1073/pnas.94.11.5906
- Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**: 148–156. doi:10.1038/ng.1064
- Brand AH, Breeden L, Abraham J, Sternglanz R, Nasmyth K. 1985. Characterization of a “silencer” in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**: 41–48. doi:10.1016/0092-8674(85)90059-5
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**: D1005–D1012. doi:10.1093/nar/gky1120
- Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, Raju A, Lieberman Aiden E, Li S, Manikandan L, et al. 2021. H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat Commun* **12**: 719. doi:10.1038/s41467-021-20940-y
- Cantor RM, Lange K, Sinsheimer JS. 2010. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am J Hum Genet* **86**: 6–22. doi:10.1016/j.ajhg.2009.11.017
- Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S, et al. 2016. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**: 1398–1414.e24. doi:10.1016/j.cell.2016.10.026
- Chen S, Yang J, Wei Y, Wei X. 2020. Epigenetic regulation of macrophages: from homeostasis maintenance to host defense. *Cell Mol Immunol* **17**: 36–49. doi:10.1038/s41423-019-0315-0
- DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, Olson KM, Huang H, Biggs W, Sandoval E, D’Antonio M, et al. 2017. Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells. *Cell Stem Cell* **20**: 533–546.e7. doi:10.1016/j.stem.2017.03.009
- Della Rosa M, Spivakov M. 2020. Silencers in the spotlight. *Nat Genet* **52**: 244–245. doi:10.1038/s41588-020-0583-8
- Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**: 331–336. doi:10.1038/nature14222
- Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD. 2020. Candidate silencer elements for the human and mouse genomes. *Nat Commun* **11**: 1061. doi:10.1038/s41467-020-14853-5
- Erceg J, Pakozdi T, Marco-Ferrerres R, Ghavi-Helm Y, Girardot C, Bracken AP, Furlong EEM. 2017. Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. *Genes Dev* **31**: 590–602. doi:10.1101/gad.292870.116
- Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**: 2478–2492. doi:10.1038/nprot.2017.124
- Ernst J, Melnikov A, Zhang X, Wang L, Rogov P, Mikkelsen TS, Kellis M. 2016. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol* **34**: 1180–1190. doi:10.1038/nbt.3678
- Gallo SM, Gerrard DT, Miner D, Simich M, Des Soye B, Bergman CM, Halfon MS. 2011. REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res* **39**: D118–D123. doi:10.1093/nar/gkq999
- Gao T, Qian J. 2020. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res* **48**: D58–D64. doi:10.1093/nar/gkz980
- Gisselbrecht SS, Palagi A, Kurland JV, Rogers JM, Ozadam H, Zhan Y, Dekker J, Bulky ML. 2020. Transcriptional silencers in *Drosophila* serve a dual role as transcriptional enhancers in alternate cellular contexts. *Mol Cell* **77**: 324–337.e8. doi:10.1016/j.molcel.2019.10.004
- Grinberg-Bleyer Y, Caron R, Seeley JJ, De Silva NS, Schindler CW, Hayden MS, Klein U, Ghosh S. 2018. The alternative NF- κ B pathway in regulatory T cell homeostasis and suppressive function. *J Immunol* **200**: 2362–2371. doi:10.4049/jimmunol.1800042
- The GTEx Consortium. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660. doi:10.1126/science.1262110
- Györy I, Boller S, Nechanitzky R, Mandel E, Pott S, Liu E, Grosschedl R. 2012. Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. *Genes Dev* **26**: 668–682. doi:10.1101/gad.187328.112
- Halfon MS. 2020. Silencers, enhancers, and the multifunctional regulatory genome. *Trends Genet* **36**: 149–151. doi:10.1016/j.tig.2019.12.005
- Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, et al. 2008. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. *Nat Genet* **40**: 204–210. doi:10.1038/ng.81
- Huang D, Petrykowska HM, Miller BF, Elnitski L, Ovcharenko I. 2019. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res* **29**: 657–667. doi:10.1101/gr.247007.118
- Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318–356. doi:10.1016/S0022-2836(61)80072-7
- Johnson WC, Ordway AJ, Watada M, Pruitt JN, Williams TM, Rebeiz M. 2015. Genetic changes to a transcriptional silencer element confers phenotypic diversity within and between *Drosophila* species. *PLoS Genet* **11**: e1005279. doi:10.1371/journal.pgen.1005279
- Jung I, Schmitt A, Diaoy Y, Lee AJ, Liu T, Yang D, Tan C, Eom J, Chan M, Chee S, et al. 2019. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat Genet* **51**: 1442–1449. doi:10.1038/s41588-019-0494-8
- Kang JW, Kim Y, Lee Y, Myung K, Kim YH, Oh CK. 2021. AML poor prognosis factor, TPD52, is associated with the maintenance of haematopoietic stem cells through regulation of cell proliferation. *J Cell Biochem* **122**: 403–412. doi:10.1002/jcb.29869
- Kehayova P, Monahan K, Chen W, Maniatis T. 2011. Regulatory elements required for the activation and repression of the protocadherin- α gene cluster. *Proc Natl Acad Sci* **108**: 17195–17200. doi:10.1073/pnas.1114357108
- Koh FM, Lizama CO, Wong P, Hawkins JS, Zovein AC, Ramalho-Santos M. 2015. Emergence of hematopoietic stem and progenitor cells involves a Chd1-dependent increase in total nascent transcription. *Proc Natl Acad Sci* **112**: E1734–E1743. doi:10.1073/pnas.1424850112
- Koo PK, Ploenzke M. 2020. Deep learning for inferring transcription factor binding sites. *Curr Opin Syst Biol* **19**: 16–23. doi:10.1016/j.coisb.2020.04.001
- Kumar BV, Connors TJ, Farber DL. 2018. Human T cell development, localization, and function throughout life. *Immunity* **48**: 202–213. doi:10.1016/j.immuni.2018.01.007
- Lefebvre V. 2010. The SoxD transcription factors—Sox5, Sox6, and Sox13—are key cell fate modulators. *Int J Biochem Cell Biol* **42**: 429–432. doi:10.1016/j.biocel.2009.07.016
- Liu T, Porter J, Zhao C, Zhu H, Wang N, Sun Z, Mo Y-Y, Wang Z. 2019. TADKB: family classification and a knowledge base of topologically associating domains. *BMC Genomics* **20**: 217. doi:10.1186/s12864-019-5551-2
- Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, Assassi S, Ying J, Tan FK, Arnett FC, et al. 2014. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet* **94**: 47–61. doi:10.1016/j.ajhg.2013.12.002
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495–501. doi:10.1038/nbt.1630
- Ngan CY, Wong CH, Tjong H, Wang W, Goldfeder RL, Choi C, He H, Gong L, Lin J, Urban B, et al. 2020. Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* **52**: 264–272. doi:10.1038/s41588-020-0581-x
- Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang J-M, Cavalli G. 2018. Polycomb-dependent chromatin looping contributes to gene silencing during *Drosophila* development. *Mol Cell* **71**: 73–88.e5. doi:10.1016/j.molcel.2018.05.032
- Ong C-T, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**: 234–246. doi:10.1038/nrg3663
- Pang B, Snyder MP. 2020. Systematic identification of silencers in human cells. *Nat Genet* **52**: 254–263. doi:10.1038/s41588-020-0578-5

- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst JBM, Yen A, Kheradpour P, Wang J, Whitaker JW, Schultz MD, Ward LD, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330. doi:10.1038/nature14248
- Rojano E, Seoane P, Ranea JAG, Perkins JR. 2019. Regulatory variants: from detection to predicting impact. *Brief Bioinformatics* **20**: 1639–1654. doi:10.1093/bib/bby039
- Senftleben U, Cao Y, Xiao G, Gretchen Florian R, Krähn G, Bonizzi G, Chen Y, Hu Y, Fong A, Sun S-C, et al. 2001. Activation by IKK α of a second, evolutionary conserved, NF- κ B signaling pathway. *Science* **293**: 1495–1499. doi:10.1126/science.1062677
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050. doi:10.1101/gr.3715005
- Stathopoulos A, Levine M. 2005. Localized repressors delineate the neurogenic ectoderm in the early *Drosophila* embryo. *Dev Biol* **280**: 482–493. doi:10.1016/j.ydbio.2005.02.003
- Tartey S, Takeuchi O. 2015. Chromatin remodeling and transcriptional control in innate immunity: emergence of Akirin2 as a novel player. *Biomolecules* **5**: 1618–1633. doi:10.3390/biom503161
- Thibodeau A, Uyar A, Khetan S, Stitzel ML, Ucar D. 2018. A neural network based model effectively predicts enhancers from clinical ATAC-seq samples. *Sci Rep* **8**: 16048. doi:10.1038/s41598-018-34420-9
- van Arensbergen J, Pagie L, FitzPatrick VD, de Haas M, Baltissen MP, Comoglio F, van der Weide RH, Teunissen H, Vösa U, Franke L, et al. 2019. High-throughput identification of human SNPs affecting regulatory element activity. *Nat Genet* **51**: 1160–1169. doi:10.1038/s41588-019-0455-2
- Wang P, Deng Y, Yan X, Zhu J, Yin Y, Shu Y, Bai D, Zhang S, Xu H, Lu X. 2020. The role of ARID5B in acute lymphoblastic leukemia and beyond. *Front Genet* **11**: 598. doi:10.3389/fgene.2020.00598
- Yang J, McGovern A, Martin P, Duffus K, Ge X, Zarrineh P, Morris AP, Adamson A, Fraser P, Rattray M, et al. 2020. Analysis of chromatin organization and gene expression in T cells identifies functional genes for rheumatoid arthritis. *Nat Commun* **11**: 4402. doi:10.1038/s41467-020-18180-7
- Zhou J, Troyanskaya OG. 2015. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* **12**: 931–934. doi:10.1038/nmeth.3547
- Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. 2018. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet* **50**: 1171–1179. doi:10.1038/s41588-018-0160-6

Received July 12, 2021; accepted in revised form January 27, 2022.