BMC Genomics

RESEARCH

Open Access

Cis-regulatory atlas of primary human CD4+T cells

Kurtis Stefan[1,2] and Artem Barski[1,3,4*]

Abstract

Cis-regulatory elements (CRE) are critical for coordinating gene expression programs that dictate cell-specific differentiation and homeostasis. Recently developed self-transcribing active regulatory region sequencing (STARR-Seq) has allowed for genome-wide annotation of functional CREs. Despite this, STARR-Seq assays are only employed in cell lines, in part, due to difficulties in delivering reporter constructs. Herein, we implemented and validated a STARR-Seq–based screen in human CD4+T cells using a non-integrating lentiviral transduction system. Lenti-STARR-Seq is the first example of a genome-wide assay of CRE function in human primary cells, identifying thousands of functional enhancers and negative regulatory elements (NREs) in human CD4+T cells. We find an unexpected difference in nucleosome organization between enhancers and NRE: enhancers are located between nucleosomes, whereas NRE are occupied by nucleosomes in their endogenous locations. We also describe chromatin modification, eRNA production, and transcription factor binding at both enhancers and NREs. Our findings support the idea of silencer repurposing as enhancers in alternate cell types. Collectively, these data suggest that Lenti-STARR-Seq is a successful approach for CRE screening in primary human cell types, and provides an atlas of functional CREs in human CD4+T cells.

Keywords Cis-Regulatory Elements, Enhancers, Negative Regulatory Elements, Silencers, CD4, STARR-Seq

Background

Cis-regulatory elements (CRE) are DNA elements that manage or modulate gene transcription. Putative CREs are frequently identified in structural assays, such as those measuring chromatin accessibility (e.g., DNase sequencing [DNase-Seq] and assay for transposase-accessible chromatin sequencing [ATAC-Seq]), but these assays do not provide information about CRE function. CREs contribute to defining cell-specific gene expression programs through positive and negative regulation, and by insulating genes from inappropriate regulation. Annotation of CRE function is an ongoing challenge using existing datasets; the chromatin and genomic landscape of CREs is highly varied [1, 2]. Multiple attempts to identify CREs from existing data rely on correlating epigenetic chromatin opening and histone enrichment with function [3–6]. Structure-based approaches fail to explain gene expression changes: in our recent study of T cell activation, genes with nearby chromatin opening were found to be both up- and down-regulated [7]. Recently, the ENCODE Project has released an encyclopedia of CREs (SCREEN), which leverages available chromatin modifications, chromatin immunoprecipitation sequencing (ChIP-Seq) binding experiments, and expression quantitative trait loci (eQTLs) to annotate putative regulatory elements [8, 9]. Other attempts at annotating the regulatory genome were attempted with genome-wide Cas9 editing [10], and more narrowly with targeted Massively Parallel Reporter

*Correspondence:
Artem Barski
Artem.Barski@cchmc.org
[1] Division of Allergy & Immunology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 7028, Cincinnati, OH 45229-3026, USA
[2] Medical Scientist Training Program (MSTP), University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
[3] Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229-3026, USA
[4] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA

Assays [11]. Measuring CRE activity using self-transcribing active regulatory region sequencing (STARR-Seq) has proved to be adaptable and sensitive, making it a powerful assay in the functional genomic toolkit [12]. However, CRE imputation is limited in functional predictive power, as screening approaches, including STARR-Seq, largely overlook negative regulatory elements (NREs), such as silencers, attenuators, and insulators [13].

Enhancer functional screening has been the subject of numerous genome-wide investigations, including in some immune cell populations [14–17]. NREs, however, are historically the subjects of individual experimentation. Notably, NREs are uniquely important during the development of some immune cell populations, in which they suppress CD4 expression during thymic development [18]. Only recently have silencers been identified through novel genome-wide screening methods [11, 14, 19]. Such screens are limited by the use of a synthetic library, or the ability to identify only NREs and not enhancers. One recent report suggests that ATAC-STARR is able to detect both activating and repressing fragments, but this study was performed in a cell line and does not include functional validation [20]. There remains a need to functionally assess the regulatory potential of CREs in cell types of interest, particularly in mature, primary, non-transformed, human cells.

CD4+ T cells are adaptive immune cells with important roles in human health and disease. It is known that CD4+ T cells are transcriptionally plastic, responding to cytokine stimuli with diverse and distinct transcriptional programs [21]. These large transcriptional changes are accompanied by precise opening and closing of chromatin; however, the functional status of these putative CREs remains unknown [7]. A limited number of studies have attempted to define CREs in CD4+ T cells, including using a synthetically constructed massively parallel reporter assay (MPRA) to identify small pathogenic single-nucleotide polymorphisms (SNPs) [22] and using pseudo genome-wide assays like CapSTARR-Seq (DNA hypersensitive site enrichment with capture) in P5424 murine thymocytes [23]. Others have proposed a CD4+ subtype specific enhancer atlas based on histone 3, lysine 27 acetylation (H3K27ac) and histone 3, lysine 4 monomethylation (H3K4me1) distribution [24]. Thus, there is an active need to comprehensively profile the function of the non-coding genome of relevant human cells, in particular human CD4+ T cells.

We implemented a STARR-Seq–based screen in human resting total CD4+ T cells using a non-integrating lentiviral transduction system. Our screen identifies nearly 8000 functional enhancers and 6000 functional NREs from a library of open chromatin. This assay is the first example of a genome-wide assay examining the accessible chromatin elements in human primary cells. Herein, we demonstrate that a modified STARR-Seq screening method can identify both enhancer and NREs with high specificity as validated by luciferase assays. We find that the STARR-Seq enhancers and NREs are marked with distinct profiles of histone modifications, including both canonically activating and repressive histone marks. Interestingly, enhancers and NREs display characteristic nucleosome positioning in their endogenous locations, and provide regulation of target genes via chromatin looping to target promoters. We provide supporting evidence that NREs may function as enhancers in other cell types, whereas functional enhancers are largely specific to hematopoietic lineages. We also provide a catalogue of transcription factors that may regulate enhancers and NREs in CD4+ T cells, providing hitherto known and unknown factors for future study.

Results

Lenti-STARR-Seq in Primary Human CD4+ T Cells

We adapted a screening vector from the promoter-less STARR-Seq vector, which uses a bacterial origin of replication (ORI) as a cryptic promoter to initiate transcription (Fig. 1A) [25]. This STARR-Seq screening vector was further cloned into a lentiviral backbone for viral packaging [pLenti-STARR]. An open chromatin library was prepared from a single donor's resting total human CD4+ T cells in 16X reactions with 100,000 cells each to ensure sufficient library diversity. The OMNI-ATAC protocol was employed to reduce the number of mitochondrial reads without the need for negative selection [26]. ATAC-Seq library fragments were gel purified (150–500 bp including adapters), cloned into linearized pLenti-STARR using NEBuilder HiFi assembly, and amplified in bacteria (Fig. 1B). We elected to use lentiviral particles with an HIV-1 envelope to allow for transduction of resting CD4+ cells [27, 28]. A lentivirus packaging vector with mutated integrase, psPAX-D64V, was used to ensure that the STARR-Seq plasmid stays episomal, thus avoiding positional regulatory effects [28]. The Lenti-STARR-Seq library was transduced into human CD4+ T cells isolated from four separate adult, healthy donors. At least 50 million CD4+ cells were transduced per donor to ensure high fragment diversity in each biological replicate. Twenty-four hours after transduction, total RNA was harvested, reverse transcribed with a STARR transcript specific primer, PCR amplified, and next-generation sequenced. An input control library was amplified from the pLenti-STARR-Seq library.

Enhancers were called using MACS2, with the input plasmid library used as a control. Only enhancers overlapping peaks called from the input control were

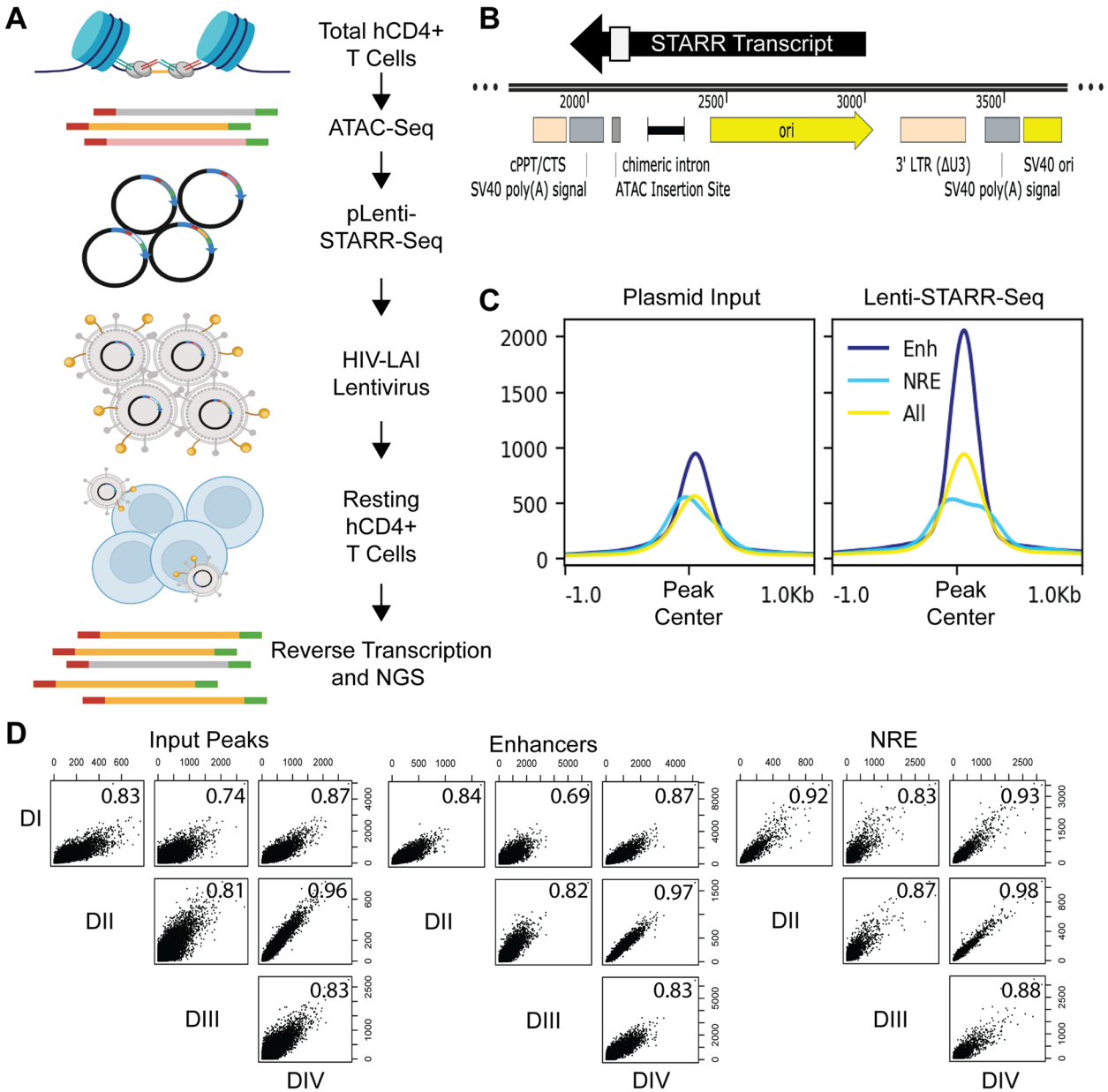

Fig. 1 Lenti-STARR Approach. **A** Putative regulatory sequences enriched by ATAC-Seq are used as input into Lenti-STARR-Seq. ATAC-Seq was performed on total human CD4+T cells from peripheral blood obtained from healthy adult donors. The ATAC-Seq library is cloned into a promoter-less STARR-Seq screening vector that uses a bacterial origin of replication as a cryptic promoter [25]. The screening library is packaged in an HIV-enveloped non-integrating lentivirus, and transduced into resting CD4+T cells. After 24 h in culture, STARR-Seq RNA is collected, reverse-transcribed using a transcript-specific primer, PCR-amplified, and prepared for next-generation sequencing (NGS). Created with *BioRender.com* **B** The Lenti-STARR-Seq screening vector carries an ATAC insertion site downstream of chimeric intron. **C** Enhancers are called with MACS2 [-log10(P) > 75], comparing STARR-Seq–obtained RNA reads to input plasmid reads [29]. Negative regulatory elements (NREs) are called using MACS2 [-log10(P) > 30] comparing input plasmid to STARR-Seq-obtained RNA reads [29]. Mean tag density is displayed, compared to 'All' accessible peaks from input. Reads are centered at CRE peak center. **D** Pearson correlation of tags across input ATAC-Seq peaks, Enhancers, or NREs for each STARR-Seq human CD4+ donor ($N = 4$)

considered for downstream analysis. Although the weaker ATAC sites may function as enhancers within a STARR-Seq assay, we reason that the lower accessibility makes them less likely to be functional in their native genomic context. NREs, likely including silencers and insulator sequences, were detected using MACS2 with the plasmid as treatment and STARR-seq RNA as control. All genome-wide analysis was conducted with only

highly significant enhancers [-log10(P) > 75] and NREs [-log10(P) > 30] which intersected strongly accessible chromatin from peaks which overlap MACS peaks called from the input plasmid library alone.

STARR-Seq functional enhancers demonstrate strong central STARR RNA production, whereas NREs are depleted in the STARR-Seq library relative to all peaks from input plasmid control (Fig. 1C). Although the Lenti-STARR screening library is shared between donors, STARR RNA regulation appears consistent across the hCD4+ donors (Fig. 1D), as demonstrated by high Pearson correlation coefficients of read counts across peaks.

Enhancer and NRE element validation

We next sought to validate CRE function in traditional luciferase assays using randomly selected CRE across a variety of MACS2 significance and input accessibility levels (Fig. 2A). Functional NREs in STARR-Seq were cloned upstream of a strong promoter (pCMV-ENH-LUC) and transfected into resting hCD4+ T cells. Luciferase activity was calculated relative to a Renilla luciferase (pRL-TK) transfection control. [Mann–Whitney test, N = 3 human CD4+ donors]; We find that the majority of putative NREs were functional in traditional luciferase assays, across varying significance and activity (fold change) levels, and across a wide array of endogenous genomic locations (intron, exon, near promoter, intergenic).

STARR-Seq–identified enhancers were cloned upstream of the minimal promoter (pGL4.26) luciferase-expressing vector. Enhancers demonstrated strong luciferase activity, which was sometimes donor dependent (Fig. 2B). Similar to NREs, functional enhancers identified by STARR-Seq came from across various genomic locations (intron, intergenic, exon, near promoter). Some fragments displayed in Fig. 2 were tested in luciferase assays despite not reaching genome-wide significance.

We next estimated the true proportion of enhancers and NREs that are functional within the original screening library. The true proportion of CRE activity in the screen was estimated from the success rate of luciferase validation using binomial exact tests (Fig. 2C). It has been suggested that enhancers comprise between ~ 10% [30] and 35% [31] of accessible chromatin sites. We found that 27% of ATAC-Seq peaks in CD4+ T cells exhibited statistically significant enhancer activity by STARR. The estimated true proportion of these enhancers that are also luciferase active with a maximum-likelihood estimated at nearly 80%. STARR functional NREs comprise nearly 21% of the accessible genome in CD4+ T cells and also yielded an estimated true proportion of functional NREs of nearly 75% by luciferase validation. Although it is unknown what fraction of the accessible chromatin landscape NREs comprise, we observed overall that STARR-Seq performed similarly for enhancers and NREs, but may marginally outperform in detecting functional enhancers over functional NREs.

STARR-Seq CREs regulate transcription within their endogenous location

As expected, STARR-Seq enhancers and NREs display a distinct relationship with RNA transcriptional initiation and RNA polymerase. Assays, including Precision Run-On Nuclear Sequencing (PRO-Seq), capture nascent RNA production (bidirectional non-polyadenylated RNA) characteristic of active enhancer elements [32–34]. Shown in Fig. 3A, we find functional enhancers in STARR-Seq displayed strong central enrichment of nascent RNA production in PRO-Seq, whereas NREs were centrally depleted of such transcripts. Intriguingly, there was enrichment of the PRO-Seq signal within 100–200 bp around NREs, suggesting that enhancers may be located in surrounding regions. This suggests that although NREs are located in genomic areas generally permissive to eRNA transcription, transcriptional initiation is centrally repressed. Although the NREs may include silencers and insulators, the clear repression observed in the PRO-Seq signal suggests that this group of CRE is distinctly less transcriptionally permissive. NREs and enhancers both demonstrated RNA Polymerase II binding, though NREs exhibited weaker enrichment of S5-Phosphorylated PolII. These findings suggest that PolII may be recruited to enhancer elements bordering NREs but is less likely to be functionally active (Fig. 3B).

We next sought to characterize the effect of enhancers and NREs on expression of their target genes. First, we attempted to assign target genes to CREs on the basis of distance. Each CRE was assigned to the nearest TSS if within 10 kb (Fig. S1). We found that genes assigned enhancers, or both enhancers and NREs, had significantly higher TPM than genes associated with only open peaks from input.

Although assignment of genes to CREs by distance was used traditionally, multiple reports demonstrate that both enhancers and NREs loop from long distances to activate or repress their target promoters [19, 38]. We assigned CRE gene targets leveraging Promoter Capture HiC (PC-HiC) performed in total CD4+ T cells [37]. Genes that are physically contacted by NREs are significantly repressed compared to those contacted by enhancers (Fig. 3C). Genes that are the targets of both enhancers and NREs or enhancers alone had significantly higher expression than genes that were targets of non-annotated ATAC-Seq peaks. We next assessed whether the regulation by CRE was dose-dependent on the number of CREs that target each promoter. We found that amongst genes only contacted by NRE, increasing the number of NRE to promoter contacts is not a significant negative predictor of gene expression

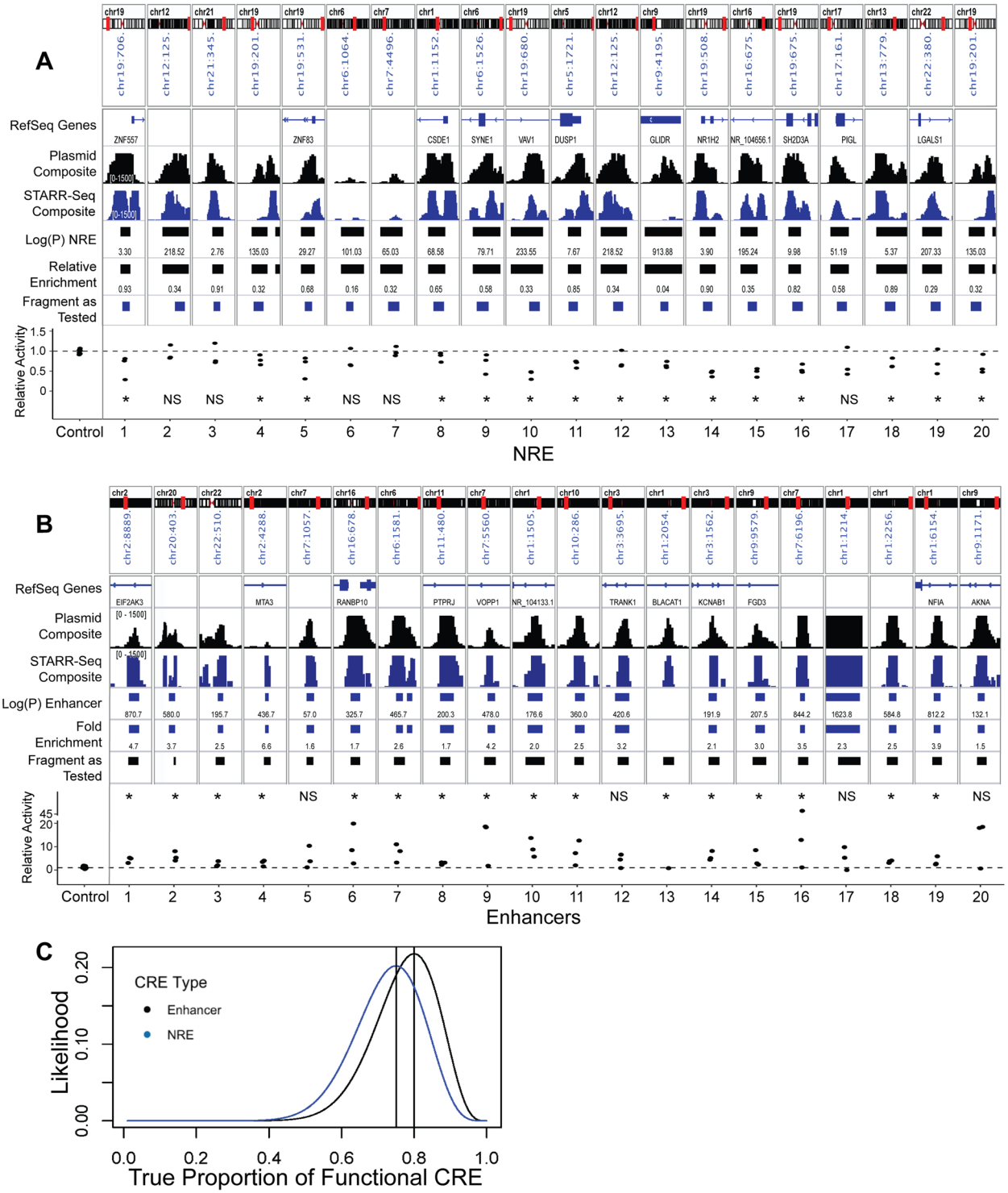

Fig. 2 CRE Validation by Luciferase Assays. MACS2 fold enrichment (for enhancers) and MACS2 relative enrichment (for NREs) are displayed in the genome browser. Fragment Tested represents the DNA sequence cloned for luciferase testing. **A** Twenty putative NREs are tested using CMV-Enhancer-LUC in DualGLO Luciferase assays (Promega). Relative luciferase activity of each CRE as tested in $N = 3$ independent human CD4+ donors is displayed. Relative Activity compares NRE activity to empty-vector CMV-Enhancer-LUC control activity. [(*) P-value ≤ 0.05 in one sided Mann–Whitney test] **B** Twenty putative enhancers are tested using pGL4.26 (minP promoter) in DualGLO Luciferase assays as described above. **C** Estimation of the true proportion of tested CREs which are functional in luciferase assays. The maximum likelihood is estimated from the binomial distribution given the number of significant luciferase CRE as tested

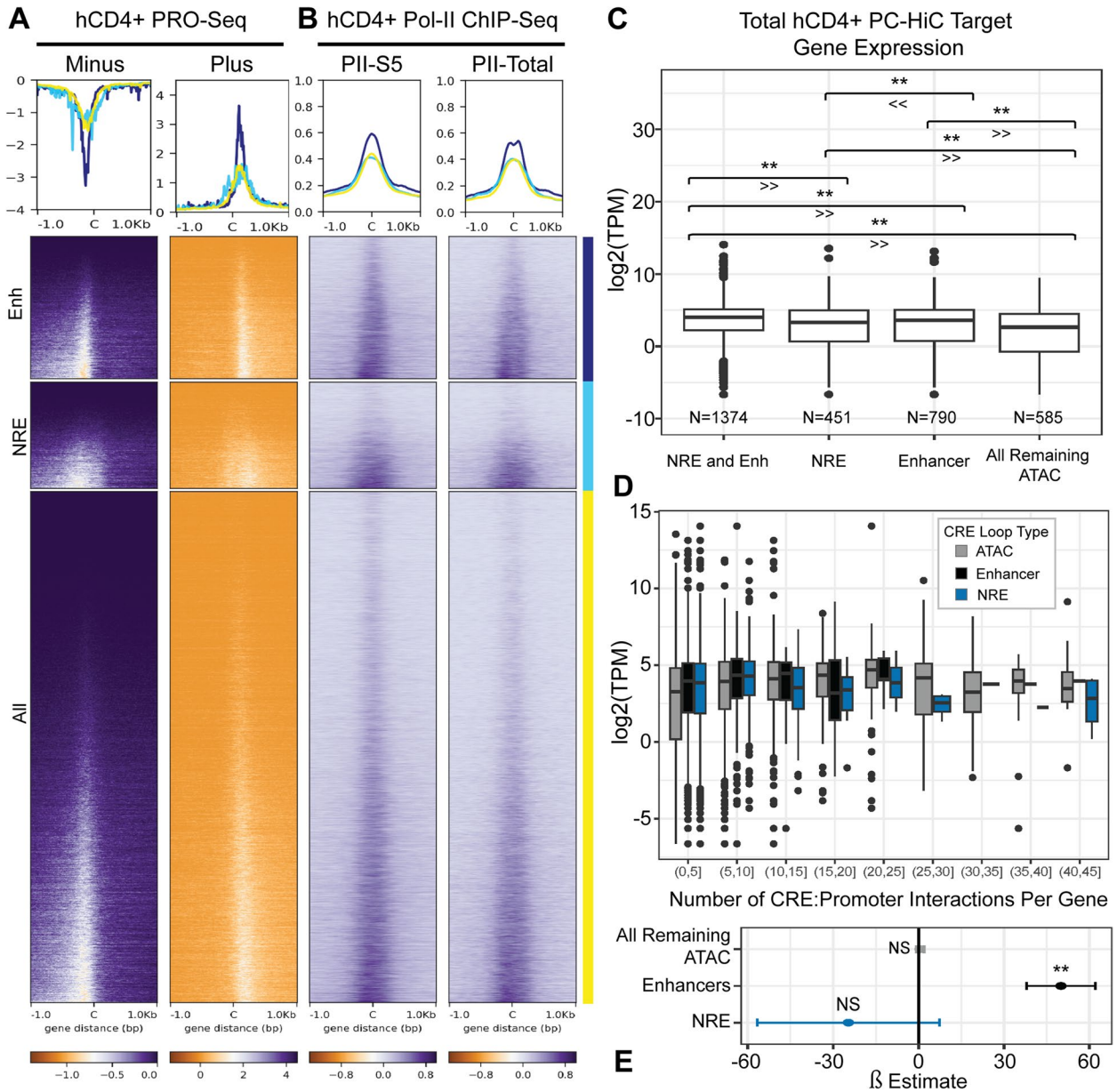

Fig. 3 CRE Regulation of Transcription. **A** PRO-Seq tags in hCD4+T cells at STARR-Seq Enhancers, NREs, and 'All' open peaks from input plasmid library [35]. Reads are centered at CRE peak center 'C'. **B** ChIP-Seq experiments of RNA Polymerase II Phospho-S5 and total RNA Polymerase II conducted in human CD4+T cells [36]. **C** Chromatin looping targets of CRE to promoter with TPM from hCD4+ PolyA RNA-Seq [35, 37]. [** Adjusted *P*-value ≤ 0.05 by non-parametric Kruskal–Wallis one-sided test with Holm adjustment for multiple comparisons]. **D** Number of CRE:Promoter PC-HiC contacts to each target gene, with TPM from hCD4+ PolyA RNA-Seq [35]. **E** Linear model explaining TPM by number of CRE:Promoter interactions to each promoter within each group of C [35] [** *P*-value ≤ 0.05]

($ß_{NREs}$: -23.66 ± 31.89, *P*-value = 0.44) (Fig. 3D and E). The number of enhancers that contacted a gene was a positive predictor of expression ($ß_{ENH}$: 50.03 ± 12.06, *P*-value = 3.75E-05). It is well known that enhancers are often cryptic and redundant [39, 40], which supports our finding that genes with at least one enhancer association displayed significantly higher expression in Fig. 3C.

Lenti-STARR NREs and enhancers exhibit distinct patterns of chromatin modifications

We next sought to compare histone modification landscape across functional enhancers and NREs identified by STARR-Seq. Previous functional silencer screens report enrichment with chromatin modifications H3K27me3, H3K9ac, and H3K79me2, which are also relatively

enriched at functional NREs compared to enhancers in this study (Fig. S2) [19, 20, 41, 42]. Enhancers classically display enrichment with H3K27ac, H3K4me1, and sometimes H3K4me3, which we also observe at our STARR functional enhancers as well. Reports from other silencer screens of open chromatin find some enrichment for marks of heterochromatin [19, 42]. Functional NREs identified in our assay were weakly enriched for H3K27me3 (Fig. 4A). We found that the traditionally active histone marks H3K27ac and H3K4me3 demonstrated divergent enrichment patterns at functional enhancers and NREs. H3K4me1, another mark of active enhancers, appeared similarly enriched at both NREs and enhancers. The distribution of H3K27ac and H3K4me3 histone ChIP-Seq tags suggested that NREs might be occupied by a nucleosome. To test this hypothesis, we imputed positions of nucleosomes from ATAC-Seq data using NucleoATAC [43]. Functional NREs were occupied by nucleosomes (or possibly nucleosome-size protein complexes), whereas enhancers seemed to be located between nucleosomes.

To test where functional enhancers and NREs reside within previously imputed chromatin states, we overlapped these elements with genomic segments identified by ChromHMM imputation based on 25 histone marks (Fig. 4B) [3]. In CD4+ T cells, STARR-Seq functional enhancers were more likely to be segmented as enhancer subtypes, or as active TSS. We also did not observe differences in CpG island enrichment within promoters between NREs and enhancers (Fig. S7B). NREs functional in STARR-Seq were more likely to be adjacent to promoters, particularly in downstream of TSS DNase accessible regions, which are likely intronic; enhancers are more likely found within annotated promoter regions (Fig. S7A). NREs displayed weaker enrichment in putative enhancer classes than did STARR-Seq functional enhancers. Though STARR-Seq enhancers were most likely to be segmented as enhancers in imputation models in hCD4+ T cells themselves, we found that these enhancers were also predicted as enhancer groups across hematopoietic cells, suggesting that some CD4+ enhancers are functional within a broader class of hematopoietic cells. Neither STARR-Seq enhancers nor NREs frequently segmented in regions of heterochromatin or ZNF repeats in mature CD4+ or hematopoietic lineages; mature CD4+ T cells exhibited weak heterochromatin chromatin signatures at sites of open chromatin [36]. Intriguingly, functional NREs also appear within accessible and classically activated chromatin environments. We also observe that the percentage of both NREs and enhancer elements that were devoid of histone modifications (quiescent) increased in non-hematopoietic cell types and stem cell lines, suggesting a progressive opening and activation of these CREs during hematopoietic development.

Both STARR-Seq identified enhancers and NREs were enriched within related cell types' enhancer predictions in FANTOM5 (Fig. 4C) [45]. Consistent with other silencer reports, we found that NREs could convert into enhancers in unrelated cell types on the basis of FANTOM5 enhancer annotations, DNase accessibility, and enhancer chromatin marks H3K4me1 and H3K27ac (Fig. 4D, E and F). Though we found that functional NREs are annotated as enhancers using their chromatin environment, additional experiments are needed to demonstrate that NREs are in fact cross-functional as enhancers in other cell types.

Transcription factor binding across Lenti-STARR cis-regulatory elements

In order to identify putative transcription factors that interact with the regulatory elements that we identified, we examined the overlap between these elements and a collection of ChIP-Seq experiments from various cell types available in GEO database using the RELI algorithm [44]. Functional enhancers were likely to be bound by numerous, classic enhancer activating factors (Fig. 5A). For example, CCAAT/enhancer binding protein zeta (CEBPZ) demonstrated stronger enrichment in enhancers versus NREs. We found that lymphoid enhancer binding factor 1 (LEF1) had strong enrichment at enhancers compared to NRE in epithelial carcinoma cells. LEF1 also has lymphocyte-specific functions: it promotes expression of some conventional and regulatory Th cell genes, while repressing CD8+ T cell–specific cytokine gene expression [46, 47]. We observe that STARR-Seq functional enhancers are enriched with binding by some activation-inducible transcription factors: NFATC1, NFATC2, FOS (AP-1), and NFκB1 (Fig. S3A and B). This suggests that at least some enhancers are capable

(See figure on next page.)

Fig. 4 Chromatin Environment of Functional Regulatory Elements. **A** Tag density plots of mean intensity and heatmaps of ChIP-Seq or ATAC-Seq performed in human CD4+ T cells plotted against STARR-Seq–identified Enhancers, NREs, and 'All' open peaks from input [35]. Nucleosome location in resting CD4+ T cells imputed from NucleoATAC [43]. Tag density of mean signal intensity displayed for Histone enrichment, and of imputed nucleosome position. Reads are centered at CRE peak center 'C'. **B** Proportion of functional CD4+ STARR-Seq Enhancers and NREs within ChromHMM Segmentation of 25 imputed groups [4]. Other cell groups displayed as percent of total. **C** RELI enrichment logP-value of available enhancer predictions in FANTOM5 database [44]. Coverage ratio is the proportion of target sites overlapping all sites. **D** RELI enrichment logP-value of STARR-Seq identified elements across available DNase-Seq experiments, **E** H4K1me1 experiments, and **F** H3K27ac experiments [44]

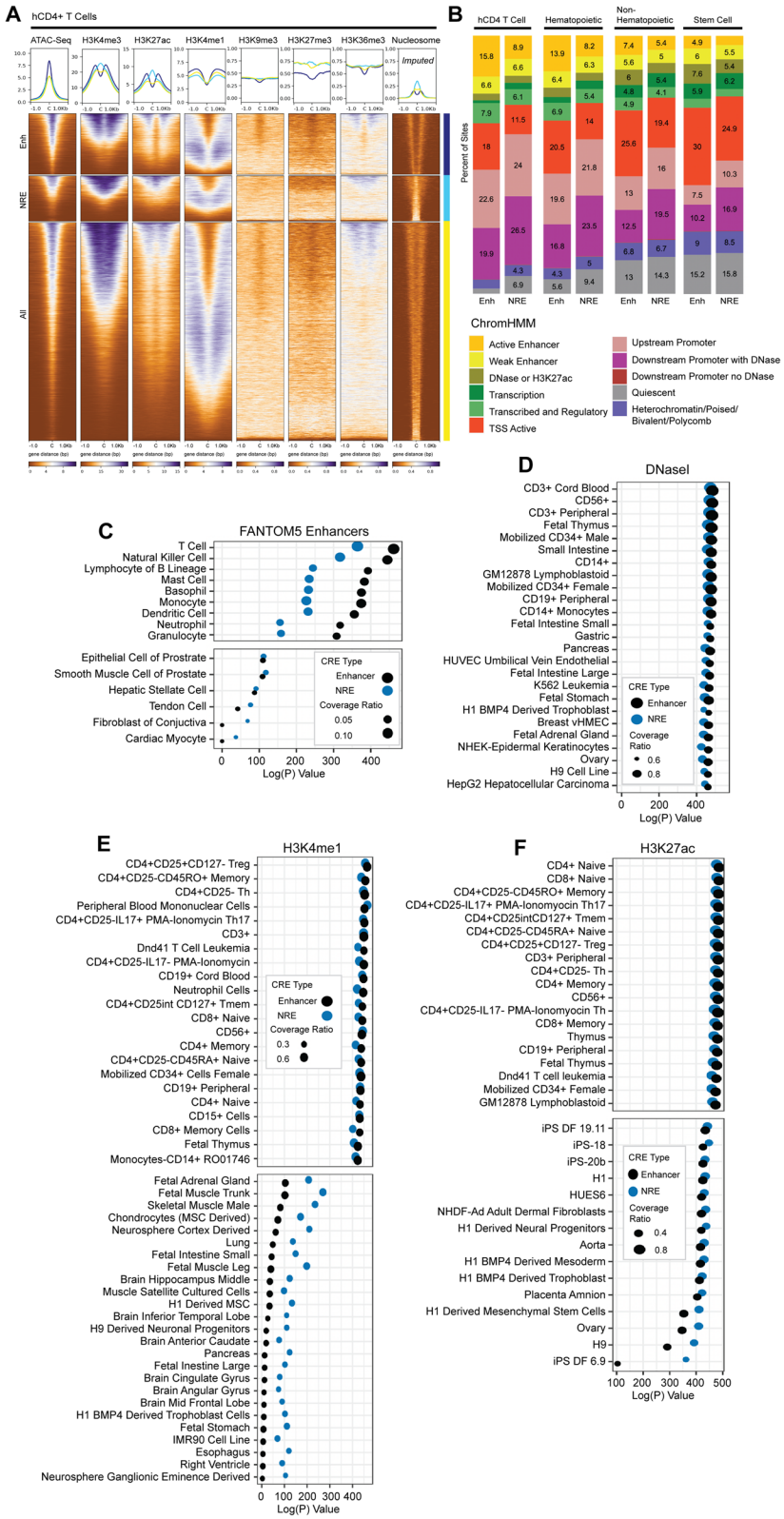

Fig. 4 (See legend on previous page.)

of participating in, or become subsequently turned-on during CD4 activation. Constitutively expressed factors, including ETS family member FLI1, also were enriched within enhancer motifs (Fig. 5C). FLI is known to bind enhancers in T cells [48], and other immune cell subtypes [megakaryocytes [49]], and can induce T cell leukemia if overexpressed [50]. The FLI portion of FLI-Ewing's Sarcoma fusion proteins was characterized as an activator of enhancers, and activated reporters in luciferase assays [51]. Another putative enhancer binder that we identified was CD74, a cell surface marker which was recently shown to also possess transcriptional activation activity [52, 53].

We explored putative co-regulators of NREs by leveraging a catalogue of ChIP-Seq experiments, again performed in various cell types (Fig. 5B). Numerous histone modifiers are likely to play a role in NRE function; NREs demonstrated enrichment for lysine demethylases, including KDM2B and KDM4A, which are recruited to promote demethylation at H3K4, H3K9, and H3K36 residues. KDM4A demethylates both H3K36me3 [55] and H3K9me3, and recruits NCOR (nuclear core repressor) [56]. Other canonical repressors were also enriched in NREs, including members of polycomb repressor complex 1 (RNF2, RYBP) and polycomb repressor complex 2 (EZH2, SUZ12, EED).

NREs may also be recruiters of chromatin remodelers. INO80 member SRCAP was recruited to NREs in human CD4+ T cells (Fig. S3), and is a chromatin remodeler recruited by YY1 and other TFs. These findings support a model in which the accessibility changes and remodeling nucleosome position seen in Fig. 4A may be a result of chromatin remodeler recruitment to NREs. Next, we conducted motif analysis to find transcription factor motifs enriched at CREs (Fig. 5). Similar to other reports, we found that STARR-Seq identified putative insulators in the negatively regulated fraction of screened sequences [14]. We observed that the CTCF paralogue, CTCFL (Boris), motif was the most enriched motif at NREs, followed by the ETS family motif, which was highly enriched across all open chromatin sites in CD4+ T cells. (Fig. 5D). The function of CTCF is likely dual purpose; enhancers were also enriched for the CTCF paralogue, CTCFL (Boris), which is known to facilitate long-range enhancer and super-enhancer looping to target gene promoters [57, 58]. Other enriched motifs within both enhancers and NREs belonged to families with mixed activating and repressing function, including the RUNX and ETS families (Fig. 5C) [59, 60]. Enhancers were also enriched for multiple activating transcriptional activators, including NFY, which frequently binds cell-specific enhancers [61], and SP5. We also found motif enrichment for transcription factors related to CD4 activation, suggesting that some of these enhancers are functional during CD4 activation (Jun).

Discussion

We describe a screen identifying functional enhancers and NREs from a library of open chromatin in resting human CD4+ T cells. Using a STARR-Seq approach, we identified and validated the function of CREs in human primary CD4+ T cells. We found that Lenti-STARR-Seq can identify enhancers and NREs, each with nearly 80% sensitivity, as verified by luciferase assays. We described the chromatin landscape of both enhancers and NREs, in which enhancers were marked with H3K27ac and H3K4me1. They were also enriched for binding of enhancer activating proteins (CEBPZ). Intriguingly, we found that NREs tend to have central nucleosome placement, unlike enhancers which are located between nucleosomes. NREs and enhancers also function to regulate gene expression via long-range chromatin interactions. In short, we report thousands of functional enhancers and NREs from a screen of CRE function in primary human CD4+ T cells.

Enhancers have long been known to bind activating factors, including P300 and CEBP(Z), among others. We also found in other cell types that CD4 enhancers could bind LEF, TCF, and novel, recently identified enhancer activators like CD74. Integration of chromatin accessibility and chromatin modification data across cell types suggests that many CD4 functional enhancers are broadly active across hematopoietic lineages. Indeed, many hematopoietic cells have overlapping repertoires of transcription factors that may bind these elements. Although many enhancers are likely shared between hematopoietic lineage cells, these findings invite further investigation to refine enhancer specificity within lymphoid cell subtypes, particularly during developmental transitions (e.g. CD8+ CD4+ double-positive to CD4+ CD8- single-positive cells), and

(See figure on next page.)

Fig. 5 Transcription Factor Binding at CRE. RELI enrichment logP-value of CRE enrichment across available transcription factor ChIP-Seq experiments in GEO [44]. Heatmaps of RNA expression [transcripts per million (TPM)] in resting human CD4+T cells are displayed furthest right [35]. Transcription factor family is displayed adjacent to RNA expression. **A** The ChIP-Seq experiments with strongest relative enrichment in STARR-Seq enhancers are displayed, whereas **B** displays the top experiments with strongest relative enrichment in NRE. **C** Motif enrichment performed in HOMER de novo motifs in enhancers (predicted motif, *p*-value, and transcription factor name) and **D** in NREs [54]

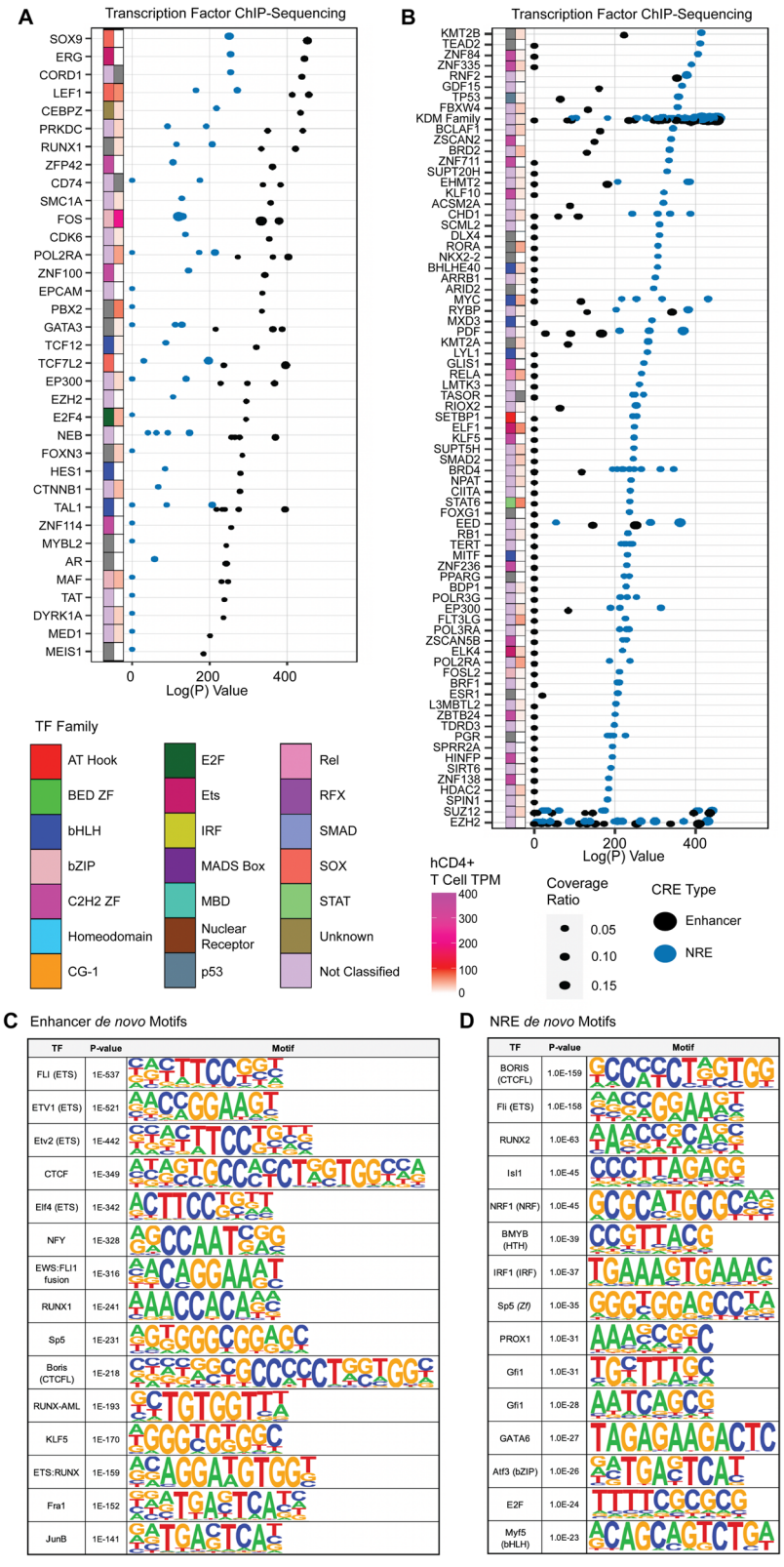

Fig. 5 (See legend on previous page.)

during activation events in immune cells. We also find that CD4+ enhancers are capable of being bound by inducible transcription factors following activation (Fig. S3). Given that our experiments were performed in resting cells, it remains to be shown whether these enhancers are repurposed during activation by these inducible factors or whether lentiviral transduction may be weakly activating.

We conducted this STARR-Seq screen using preparations of total CD4+ T cells. This cell isolation pool is comprised of numerous CD4+ T cell subtypes, many of which we expect exhibit unique CRE function. Though it is possible to glean subtype-specific CRE function based on accessibility and chromatin modifications, we cannot rule out that the STARR-Seq functional CRE identified herein are functional in only some CD4+ subtypes. This limitation is an ongoing technical challenge for screening methods, including STARR-Sequencing, as it requires high read-density coverage and tens of millions of cells per biologic replicate. With the advent of single-cell sequencing and single-cell CRISPR screens, these limitations may become less technically and fiscally burdensome.

Broadly, two mechanisms of silencer action have been proposed: 1) competing DNA binding between activating and repressing proteins and 2) recruitment of negative regulators, such as HDACs, with subsequent formation of heterochromatin domains [62, 63]. Although we do observe local heterochromatin mark enrichment at functional NREs, use of an accessible chromatin library as input precludes analysis of functional NREs within expansive heterochromatin marked domains. Rather, we demonstrate that functional NREs can exist in open chromatin and even within 'activated' chromatin environments. These observations suggest some competition between transcription factors with opposing functions. The presence of PolII at NREs, but weaker enrichment of activated S5-PolII at NREs compared to enhancers, suggests that regulation of polymerase may be a silencing mechanism used by at least some NREs. Though we did not identify a central unifying repressor in CD4+ T cells, we found that these NREs were likely to recruit transcription factors and co-activators with known repressor and nucleosome-repositioning function. Other silencer screens corroborate this widely varied transcription factor and motif utilization across the functional silencer elements; no unifying chromatin marks or transcription factor binding profile of functional NREs has been identified to date [11, 19, 20, 42, 63].

DNA methylation is an important epigenetic regulator of transcription [64, 65]. It is known that both activators and repressors are recruited to sites of DNA methylation including methyl CpG binding protein 2 (MECP2), which can function as both an activator and repressor of gene expression [66]. Intriguingly, both enhancers and silencers were reported to lie within CpG islands [42, 67, 68]. We also demonstrate that overlap with CpG islands is not different between NREs and enhancers (Fig. S7B). We cannot rule out that some elements identified in our screen mechanistically rely upon DNA methylation for either enhancer or NRE function, which may be an area for future study.

We unexpectedly observe that classically activating histone marks (H3K27ac, and H3K4me3) are enriched at NRE. Due to central enrichment of H3K4me3 at NRE, we assessed the proximity of NREs with promoters. NREs are more likely to be adjacent to promoters, particularly downstream of TSS within DNase accessible regions (intronic); enhancers are more likely found within annotated promoter-TSS regions (Fig. S7A). Enhancers, in fact, are demonstrably *active* promoters, which leads to detection of bi-directional RNA transcription originating from enhancer CRE by PRO-Seq. We demonstrate that PRO-Seq signal is enriched at the identified enhancers, and centrally *depleted* at functional NREs. NREs also function to actively repress target gene expression and are not merely inactive promoters (Fig. 3).

This screen demonstrates striking differences in patterns of epigenetic enrichment between enhancers and NREs. Other screens have identified functional silencer activity to be associated with H3K36me3 and H3K9ac, marks of active transcription and enhancers. We also find H3K9ac enrichment at NREs (Fig. S3C); analysis of functional silencers in K562 cells by Pang et al. 2020 also found that silencers were marked with H3K27ac and alternate histone H2A.Z. Intriguingly, some STARR-Seq NREs are positioned within active chromatin environments, including super-enhancers, although only STARR-Seq enhancers are significantly enriched within super-enhancers (Fig. S4). Previously, a detailed analysis of one super-enhancer found that some of its constituent elements function as silencers [69]. These multimodal ATAC-Seq sites may be multifunctional, providing both positive and negative regulatory information to nearby genes in different cellular contexts. Further study is required to refine the in situ function of these complex enhancer:NRE relationships at the level of individual genes.

We observe that a large number of functional NRE are occupied centrally by nucleosomes, as imputed by NucleoATAC, and confirmed by histone modification ChIP-Seq [43]. These nucleosome occupied fragments

are isolated in the input ATAC-Seq library; this is not surprising given that Tn5 is likely to cleave target DNA in the areas between nucleosomes which produces nucleosome-containing fragments upon PCR. Furthermore, each ATAC-Seq peak represents *composite accessibility:* the isolated DNA is prepared from millions of cells which possess a variety of nucleosome positions. Central nucleosome placement can facilitate repressive TF recruitment to DNA motifs; indeed, many TF families demonstrate preferred binding to nucleosomes over free DNA [70].

Our approach queried the function of accessible CREs for enhancer and negative regulatory activity from a library of open chromatin; this assay excluded interrogation of CRE within heterochromatin and inaccessible chromatin environments. Screens for silencers from libraries of repressed chromatin, performed in Drosophila and other Hi-C experiments, suggest that at least some CREs within H3K27me3 domains are also functional as NREs [11, 38, 42]. Intriguingly, we also found some weak central enrichment of heterochromatin mark H3K27me3 at functional NREs in CD4+ T cells, despite a paucity of this mark at sites of open chromatin in mature CD4+ T cells. We also demonstrate that NREs and enhancers display key differences in local chromatin accessibility. STARR-Seq functional NREs possess a centrally positioned nucleosome, which may drive the strong central enrichment that we observed for the histone marks H3K27ac and H3K4me3 (Fig. S5). It is known that nucleosome repositioning by transcription factors can be repressive when placed in unfavorable locations for transcription [71]. A direct investigation of silencers that form heterochromatin found that silencing requires precise nucleosome positioning and can be sensitive to nucleosome repositioning [72]. Other investigations find that nucleosome positioning is important for silencing of heterochromatically marked genes more broadly [73, 74]. As our assay tests putative CREs outside of their native chromatin environment, some NREs may contain intrinsic nucleosome-positioning information, or be permissive to nucleosome occupancy. These possibilities could support a potential silencing mechanism related to local chromatin accessibility changes that prevent transcriptional elongation. Alternatively, STARR-Seq may uncover some functional silencers that are endogenously occupied by nucleosomes. Intriguingly, we demonstrate that these NREs are functional from numerous genomic locations: intronically (STARR-Seq), upstream of a promoter (Luciferase), and in long-range chromatin looping in situ (Fig. 3E). Together this suggests that the silencing mechanism in hCD4+ cells is not limited to unfavorable nucleosome placement blocking intron transcription.

Others have speculated that NREs may have alternate enhancer functions in unrelated cell types. We also demonstrate that hCD4+ functional NREs resemble the enhancer phenotype in other cells due to their enrichment in FANTOM5, and enrichment of the enhancer histone marks H3K4me1 and H3K27ac. Though functional evidence is needed to support this hypothesis, we posit on the basis of our own findings that NRE function may also be incorrectly assumed as enhancing. Although it is tempting to ascribe enhancer activity to histone mark enrichment alone, we and others demonstrate that functional NREs exist within chromatin environments traditionally associated with enhancers. This challenge extends to many enhancer and silencer imputation attempts, which mainly leverage chromatin state. Indeed, we found that imputed silencer location in CD4+ T cells from SilencerDB had poor enrichment of Lenti-STARR NREs and enhancers (Fig. S6) [75]. We emphasize the need for further investigation into NRE behavior across developmental time periods and transitions, and specifically into whether NREs are repurposed as enhancers or simply maintain repressive function.

Conclusions

We implemented a STARR-Seq–based screen in human resting total CD4+ T cells using a non-integrating lentiviral transduction system. This screen annotates thousands of newly identified functional cis-regulatory elements in a primary human cell, and was validated using both traditional luciferase assays, and by assessing their regulation of target gene expression. We also explore the chromatin and transcription factor landscape of functional CRE, finding that NREs display characteristic central nucleosome positioning and can also function within histone environments marked by H3K27ac.

This study is an important step in characterizing the functional cis-regulatory genome in human primary cells, given that cell specific patterns of gene expression are a result of cell specific CRE function. Whereas previous assays for enhancer and silencer function are performed in cell lines or other model organisms, we demonstrate a STARR-Seq approach is useful even in difficult to transfect primary cells. These data are of great interest to study of lymphocyte function, where there is longstanding interest in identifying CRE responsible for control of gene expression.

Materials and Methods

Reagents

Catalogue Number	Kit	Manufacturer	Location
11346D	Human Untouched CD4+ Isolation Kit	Invitrogen	Waltham, MA, USA
A1048501	CTS Optimizer	Gibco	Waltham, MA, USA
G7513	L-Glutamine Supplement	Sigma-Aldrich	Burlington, MA, USA
P4333	Penicillin–Streptomycin Solution	Sigma-Aldrich	Burlington, MA, USA
C3040H	Stable E. Coli	New England Biolabs	Ipswich, MA, USA
E0554S	Q5 Site-Directed Mutagenesis	New England Biolabs	Ipswich, MA, USA
D4203	Plasmid Endo-Free Maxi-Prep	Zymo	Irvine, CA, USA
28206	MinElute Reaction Cleanup Kit	Qiagen	Germantown, MD, USA
28104	PCR Cleanup Kit	Qiagen	Germantown, MD, USA
M0544S	NEBNExt Ultra-II Q5 PCR Master Mix	New England Biolabs	Ipswich, MA, USA
T1020L	Gel Purification Kit	New England Biolabs	Ipswich, MA, USA
E5520S	NEBuilder HiFi DNA Assembly Cloning	New England Biolabs	Ipswich, MA, USA
75144	RNeasy Midi	Qiagen	Germantown, MD, USA
T2040L	Monarch RNA Cleanup Kit	New England Biolabs	Ipswich, MA, USA
18090010	SuperScript IV	Invitrogen	Waltham, MA, USA
28206	Reaction Cleanup Kit	Qiagen	Germantown, MD, USA
C640003	MegaX DH10B T1-R Electrocompetent Cells	Invitrogen	Waltham, MA, USA
1652089	Gene Pulser Cuvette 0.1 cm	BioRad	Hercules, CA, USA
BP1426	LB Broth, Miller	Fisher	Fair Lawn, NJ, USA
3471	Ultra Low Bind 6-Well Plate	Corning	Glendale, AZ, USA
11995065	DMEM, High Glucose, Pyruvate	Gibco	Waltham, MA, USA
S11150	Fetal Bovine Serum	R&D Systems	Flowery Branch, GA, USA
23966	Polyethylenimine MW ~ 25,000 (PEI)	Polysciences, Inc	Warrington, PA USA
205548	Caffeine	Millipore	Burlington, MA, USA
TR1003G	Polybrene	MilliporeSigma	Burlington, MA, USA
N1610	Nano Glo Dual Luciferase Assay	Promega	Madison, WI, USA
3912	Solid White 96-Well Plate	Corning	Glendale, AZ, USA
MPK10025	Neon Electroporation Kit, 100 µL Tips	Invitrogen	Waltham, MA, USA

Biologic Resources

Catalogue Number	Reagent	Manufacturer	Location
63586	psPAX2-D64V	Addgene	Watertown, MA, USA
133996	pLAI-Env	Addgene	Watertown, MA, USA
45968	pCMV-Enh-nLUC	Addgene	Watertown, MA, USA
6251	pHRL-TK	Promega	Madison, WI, USA
8441	pGL4.26	Promega	Madison, WI, USA
NA	HEK-293 T	ATCC	Manassas, VA, USA
NA	Leukoreduction Filters	Hoxworth Blood Center	Cincinnati, OH, USA

Human CD4+ Isolation

Peripheral blood mononuclear cells (PBMC) were isolated from Lymphocyte Reduction Filters obtained from deidentified male donors at Hoxworth Blood Center at the University of Cincinnati. CD4+ T cells were isolated by magnetic negative selection using Untouched Human CD4+ Isolation Kit (Invitrogen 11346D) per the manufacturer's recommendations. CD4+ T cells were cultured in Optimizer CTS (Gibco #A1048501) with 2 mM L-Glutamine supplement (Sigma # G7513) and 1X Penicillin–Streptomycin (Sigma # P4333) in incubators maintained at 37 °C with 5% CO2.

pLenti-STARR Vector Construction

pLenti-STARR was produced by the following steps: A lentivirus backbone plasmid was a gift from Kazuhiro Oka (Addgene plasmid #72263, RRID:Addgene_72263). The vector was digested with BspDI and KpnI and gel purified. A custom double-stranded (ds)DNA gBlock containing the origin of replication (ORI) and PolyA sequence was obtained from IDT (Supplemental Table 1), modeling the hORI-STARR screening vector generated by the Stark laboratory (Addgene plasmid #99296) [25]. Immediately following cloning, the original ORI was excised using a Site Directed Mutagenesis kit (NEB

E0554S). The resulting pLenti-STARR vector was transformed (NEB C3040H) and isolated using Endotoxin Free Maxi-Prep (Zymo D4203).

Lenti-STARR Library Preparation

The STARR-Seq inset library was created according to the ATAC-STARR-Seq approach [15]. Briefly, OMNI ATAC-Seq was performed as previously described with 1,600,000 human CD4+ T cells (in total) in batches of 100,000 cells per reaction (NEB #M0544S) [26]. The ATAC-Seq reaction products were cleaned up using the Qiagen Reaction Cleanup Kit (Qiagen #28206) and eluted in 12 µL per 100,000 cells transposed. Each resultant elution was amplified in 50 µL with custom primers (Supplemental Table 1) for 10 total cycles, as described previously [15]. The amplified library was purified using a PCR Cleanup Kit (Qiagen #28104) and size selected for 150–500 bp fragments on a 2% agarose gel, using 300 mg per column (NEB #T1020L). The eluted ATAC-Seq library was cloned into AgeI- and SalI-digested pLenti-STARR using a 3:1 molar ratio (insert:backbone) in a total reaction volume of 100 µL using the NEBuilder HIFI DNA Assembly Kit (NEB #E5520S). The reaction was concentrated to a total of 20 µL in a Reaction Cleanup Kit (Qiagen #28206). The cloned ATAC-Lenti-STARR was electroporated into MegaX DH10B cells (Invitrogen #C6400003) using 10 µL library per 100 µL cells [2.0 kV, 200 Ω, 25 µF] in 20 µL electroporation reactions in 0.1-cm cuvettes (BioRad #1652089) on a Harvard Apparatus (ECM Model 630). Immediately following electroporation, 950 µL of prewarmed recovery media was added, and cells were cultured for one hour at 37 °C. All electroporation reactions were pooled to 1 L sterile Luria Broth (Fisher #BP1426) and grown in 0.2 L per 2 L flask for 8 h at 37 °C, 300 RPM. Each 200 mL of culture was purified in Endotoxin Free MaxiPrep Reactions (Zymo #D4203) and eluted in 0.4 mL elution buffer per column.

Lentivirus preparation

HEK-293 T cells were cultured in 15-cm^2 dishes in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco #11995065) with 10% fetal bovine serum (FBS) (R&D Systems #S11150). Cells were grown to 70% confluency, then DMEM was replaced two hours before transfection. Each 15-cm^2 plate was transfected with a total of 21 µg of plasmid DNA (1:2:1 ratio psPAX2-D64V:pLenti-STARR:pLAI-Env) [psPAX-D64V (Addgene #63586), and pLAI-HIV (Addgene #133996)]. DNA was mixed with 4.25X polyethylenimine (PEI) (Polysciences #23966) in at total of 2 mL Dulbecco's Phosphate-Buffered Saline (DPBS per plate), left to rest for complex formation for 10 min at 22 °C, and then added dropwise to the cell culture while swirling. Ten hours after transfection, the cells were washed once with DPBS and then cultured in DMEM with 10% FBS and 2 mM caffeine (Millipore #205548). Forty-eight hours after transfection, the media was collected and stored at 4 °C. Seventy-two hours after transfection, the media was collected and pooled. The viral supernatant was filtered through a 0.45-µM PES filter, loaded into ultracentrifuge tubes (Beckman #344058), and centrifuged for 2.5 h at 4 °C and 35,000 g. Each viral pellet was dissolved in 200 µL DPBS and left overnight at 4 °C to resuspend. Concentrated lentivirus was resuspended, and snap frozen in an ethanol/dry-ice bath at -80 °C until use.

Lenti-STARR transduction

Freshly isolated human CD4+ T cells were suspended to 6.6E6 cells per mL in CTS Optimizer and then combined with 0.33 mL concentrated lentivirus per mL of cells. Cells were spinfected in low-bind 6-well plates (Corning #3471) with 7 µg/mL polybrene (MilliporeSigma #TR1003G) at 950 g and 32 °C for 2 h. Immediately following spinfection, cells were gently resuspended and cultured for 2 h at 37 °C, 5% CO_2. Cells were then pooled, centrifuged, and resuspended in Optimizer CTS for 24 h of culture. Cells were treated with 100 units of DNaseI (NEB M0303S) for 30 min at 37 °C for dead cell removal.

Lenti-STARR library construction

Total RNA was purified from transduced cells (Qiagen #75144) and concentrated 7.5X using an RNA Cleanup Kit (NEB #T2040L). RNA was reverse transcribed according to manufacturer guidelines with STARR-Seq transcript–specific primers (Supplemental Table 1) with SuperScript IV (Invitrogen #18090010), using twice the recommended RNA input. cDNA was purified using a Reaction Cleanup Kit (Qiagen #28206) and amplified using custom STARR-Seq transcript–specific primers (Supplemental Table 1) as previously described [15]. Diluted Lenti-STARR-Seq plasmid 'input' control was amplified using an input concentration that yielded the same amplification cycle number as the STARR library for hCD4+ Donor I. PCR amplified library was purified using PCR cleanup columns (Qiagen #28206) and quantified for PE-150 bp sequencing (Novogene).

Alignment and next-generation sequencing processing

Samples were aligned to hg19 in Scientific Data Analysis Platform (SciDAP, https://SciDAP.com) with the Trim-ChIP-PE processing pipeline (https://github.com/datirium/workflows/blob/master/workflows/trim-chipseq-pe.cwl) [76]. Briefly, reads were trimmed and aligned to hg19, and bam files were produced. All mitochondrial reads were removed. Read count summary and alignment statistics are available in Supplemental Table 4.

The per duplication count of each unique read for each biologic donor was fitted to the negative binomial distribution for each biologic replicate, and for the input plasmid library. Duplicate reads exceeding the 75th percentile were removed from each sample by MACS2 to reduce false discovery of enhancers resulting from PCR duplication. Enhancers were called using MACS2 in which the plasmid reads were used as a control, [macs2 callpeak -f BEDPE –keep-dup all -g 2.7E9], and with -log10(*P*-value) greater than 75. NRE were also detected with MACS2, using the plasmid reads as treatment and the STARR RNA as control [macs2 callpeak -f BEDPE –keep-dup all -g 2.7E9], with -log10(*P*-value) greater than 30. Largely similar results can be obtained using FAST-NR [77]. Both enhancers and NRE are retained only if intersecting an ATAC-Seq narrowPeak called using MACS2 from the plasmid reads alone.

All final values and CRE locations are listed in Supplemental Table 2. Detailed data processing script can be found at https://github.com/Barski-lab/Lenti-STARR

Heatmaps

All heatmaps were constructed using deeptools [78]. [*ComputeMatrix reference-point* -referencePoint center] where each row is centered by the CRE peak center. Figures labelled 'C' due to space constraints indicate the reference point is centered by the CRE peak center.

Motif enrichment

De Novo Motif enrichment performed using HOMER findMotifsGenome.pl was run with default parameters [54].

Nucleosome position imputation

Reads mapped to the input plasmid to ATAC-STARR were sequenced as described above. Bam files were processed with NucleoATAC to produce nucleosome occupancy scores and converted to a .bigWig for heatmap visualization [43].

Luciferase assays

Putative regulatory elements were randomly selected from all MACS2 identified peaks of STARR-Seq functional enhancers and NREs. The sequences for functional testing were selected on the basis of fragment distribution in the control ATAC library and sought to include the input ATAC-Seq peak center. gBlocks or eBlocks were synthesized (IDT) ranging from 150–400 bp in length and cloned into pGL4.26 (Promega #8441) for enhancer screening, or CMV-Enh-Luc (Addgene #45968) for NREs screening (All tested sequences included in Supplemental Table 1). Total human CD4+T cells were purified as described above. Each transfection reaction consists of 20 µg of testing vector and 500 ng of Renilla-expressing phRL-TK (Promega #6251) mixed with 3.5E6 cells in buffer T using 100 µL Neon Tips (Invitrogen #MPK10025) [2200 V, 20-ms pulse width, 1X pulse]. Cells were cultured for 24 h, collected, and suspended in 60 µL DPBS for Nano-Glo Dual Luciferase Assay, according to the manufacturer's instructions (Promega #N1610). Absorbances were read using the Dual-Nano-Glo protocol with a GloMax plate reader (Promega #GM3000) in flat-bottom, opaque, white, 96-well plates (Corning #3912). Luciferase signal was averaged by donor across the technical transfection replicates. Control transfection (empty vector) was performed for each CD4+T cell donor and plasmid condition ($N=3$ hCD4+Donors per CRE performed on separate days). [one sided Mann–Whitney test]. The maximum likelihood is estimated from the binomial distribution given the number of significant luciferase CREs tested. Plotted are the maximum likelihood estimates of y sucesses (y=number of significant CRE by luciferase testing) from n trials ($n=20$ CRE tested) across a range of probabilities from 0 to 1 ($p=[0,1]$). Modeling was performed in R using the following script: [Define likeli.plot=function(y,n) where L=function(p) dbinom(y,n,p); mle=optimize(L, interval=c(0,1), maximum=TRUE)$max and $p=(1:100)/100$; plot(p, L(p))].

RNA expression analysis

RNA expression values [transcripts per million (TPM)] were downloaded from PolyA RNA-Seq, performed in total CD4+T cells [35]. The Nearest TSS of STARR-Seq CRE were annotated by Homer and retained if within 10 kb of the CRE [54]. Genes were assigned to the following bins: targets of enhancers alone, NREs alone, both enhancers and NREs, or ATAC-Seq peak association alone. CRE and target promoter assignments were annotated using Promoter-Capture Hi-C performed in total CD4+T cells [37]. CRE were intersected (bedtools intersect), and only high-confidence chromatin interactions were retained (CHiCAGO score greater than 5) [79]. Statistical differences in target gene expression between CRE groups were assessed with a non-parametric, Kruskal–Wallis one-sided test with Holm adjustment for multiple comparisons. A linear model explaining TPM by the number of CRE:Promoter contacts was used to test for significance between gene expression and the number of enhancers, NREs, or only ATAC-Seq peak contacts for genes only contacted by the indicated CRE type. The Beta slope coefficients of each independent variable in the linear model $Y_{TPM}=ß_0+ß^*X_{NREs}$, and $Y_{TPM}=ß_0+ß^*X_{Enhancers}$ (where X is the number of elements contacting target gene promoter) are presented along with standard error. Modeling was performed in R using the following script: [glm(TPM ~ NRE) and glm(TPM ~ ENH)].

RELI analysis

We used the RELI method as a tool to estimate the significance of intersection between the identified CRE and published ChIP-Seq binding experiments. Briefly, the RELI algorithm, described in detail previously, was used to enrich 10,981 ChIP-Seq experiments of transcription factors and 2634 non-transcription factor datasets (histone modification ChIP-Seq, FANTOM5 enhancer annotations, and DNaseI experiments) [44]. RELI was run for 1000 iterations, using an open chromatin null model in hg19, without peak extension. [reli-batch-sub -rep 1000 -species hg19 -null OpenChrom]. This yields an adjusted *P*-value representing the significance of intersection between the identified CRE and each experiment [Bonferroni correction].

Utilized data sets

All utilized datasets are referenced in the appropriate section. We downloaded the following data from the ENCODE portal [35]: ENCFF214KEL, ENCFF112RXM, ENCFF787ATX, ENCFF889GEU, ENCFF625FSW, ENCFF256XXK, ENCFF655AQC, ENCFF112RXM, ENCFF225RZA, ENCFF364ULP, ENCFF211IJP, ENCFF418WZZ, ENCSR545MEZ, ENCFF681VRR. All ENCODE utilized data sets are compiled in Supplemental Table 3.

Statistical analysis

All statistical tests were performed in R and detailed in the appropriate methods sections and figure legends. (R version 4.2.2).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09288-3.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Acknowledgements

We thank Benjamin Wronowski for Tn5 preparation. We also thank Matt Weirauch and Xiaoting Chen for their assistance with RELI. We thank Michael Kotliar for his help producing the Dockerfile.

Authors' contributions

KS wrote the manuscript, designed and performed experiments, and performed analysis. AB conceptualized experiments and revised the manuscript. All authors have read and approve the final manuscript.

Funding

This work was supported by the National Institutes of Health [Ruth L. Kirschstein National Research Service Award F30 AI157421-01A1 to K.S., R21 GM135634-01A1 to A.B., and T32 GM063483 to the University of Cincinnati Medical Scientist Training Program (MSTP)].

Availability of data and materials

All generated datasets from Lenti-STARR have been deposited in GEO under GSE217535.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

K.S. has no conflict of interests to report.
A.B. is a co-founder of Datirium, LLC.

Received: 30 December 2022 Accepted: 31 March 2023
Published online: 11 May 2023

References

1. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. 2011;21(8):1273–83.
2. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013;49(5):825–37. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23473601.
3. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215–6. https://doi.org/10.1038/nmeth.1906.
4. Vu H, Ernst J. Universal annotation of the human genome through integration of over a thousand epigenomic datasets. Genome Biol. 2022;23(1):9. https://doi.org/10.1186/s13059-021-02572-z.
5. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30. https://doi.org/10.1038/nature14248.
6. Huang D, Petrykowska HM, Miller BF, Elnitski L, Ovcharenko I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. Genome Res. 2019;29(4):657–67.
7. Yukawa M, Jagannathan S, Vallabh S, Kartashov A V, Chen X, Weirauch MT, et al. AP-1 activity induced by co-stimulation is required for chromatin opening during T cell activation. J Exp Med. 2019;jem.20182009. Available from: http://jem.rupress.org/content/early/2019/10/24/jem.20182009.abstract.
8. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Ai R, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699–710. https://doi.org/10.1038/s41586-020-2493-4.
9. Anurag S, Mengting G, Emrah G, Chan L, Koon-Kiu Y, Joel R, et al. Supervised enhancer prediction with epigenetic pattern recognition and targeted validation. Nat Methods. 2020;17(8):807–14. Available from: https://uc.idm.oclc.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Fsupervised-enhancer-prediction-with-epigenetic%2Fdocview%2F2429348945%2Fse-2%3Faccountid%3D2909.
10. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol. 2016;34(2):192–8. https://doi.org/10.1038/nbt.3450.
11. Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD. Candidate silencer elements for the human and mouse genomes. Nat Commun. 2020;11(1):1061. https://doi.org/10.1038/s41467-020-14853-5.
12. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013;339(6123):1074–7.
13. Cinghu S, Yang P, Kosak JP, Conway AE, Kumar D, Oldfield AJ, et al. Intragenic enhancers attenuate host gene expression. Mol Cell. 2017;68(1):104-117.e6.

14. Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LTM, Fernandez N, et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. Nat Commun. 2015;6(1):6905. https://doi.org/10.1038/ncomms7905.
15. Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. Nat Commun. 2018;9(1):5380. https://doi.org/10.1038/s41467-018-07746-1.
16. Diao Y, Fang R, Li B, Meng Z, Yu J, Qiu Y, et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat Methods. 2017;14(6):629–35.
17. Chaudhri VK, Dienger-Stambaugh K, Wu Z, Shrestha M, Singh H. Charting the cis-regulome of activated B cells by coupling structural and functional genomics. Nat Immunol. 2020;21(2):210–20.
18. Sawada S, Scarborough JD, Killeen N, Littman DR. A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. Cell. 1994;77(6):917–29.
19. Pang B, Snyder MP. Systematic identification of silencers in human cells. Nat Genet. 2020;52(3):254–63. https://doi.org/10.1038/s41588-020-0578-5.
20. Hansen TJ, Hodges E. ATAC-STARR-seq reveals transcription factor-bound activators and silencers across the chromatin accessible human genome. Genome Res. 2022; Available from: http://genome.cshlp.org/content/early/2022/07/18/gr.276766.122.abstract.
21. Cano-Gamez E, Soskic B, Roumeliotis TI, So E, Smyth DJ, Baldrighi M, et al. Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. Nat Commun. 2020;11(1):1801. https://doi.org/10.1038/s41467-020-15543-y.
22. Bourges C, Groff AF, Burren OS, Gerhardinger C, Mattioli K, Hutchinson A, et al. Resolving mechanisms of immune-mediated disease in primary CD4 T cells. EMBO Mol Med. 2020;12(5):e12112.
23. Schmidl C, Hansmann L, Lassmann T, Balwierz PJ, Kawaji H, Itoh M, et al. The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. Blood. 2014;123(17):e68-78. https://doi.org/10.1182/blood-2013-02-486944.
24. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015;12:963. https://doi.org/10.1038/nmeth.3542.
25. Muerdter F, Boryń ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, et al. Resolving systematic errors in widely used enhancer activity assays in human cells. Nat Methods. 2018;15(2):141–9. Available from: https://pubmed.ncbi.nlm.nih.gov/29256496.
26. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14(10):959–62.
27. Joshi A, Sedano M, Beauchamp B, Punke EB, Mulla ZD, Meza A, et al. HIV-1 Env glycoprotein phenotype along with immune activation determines CD4 T cell loss in HIV patients. J Immunol. 2016;196(4):1768–79.
28. Yi G, Choi JG, Bharaj P, Abraham S, Dang Y, Kafri T, et al. CCR5 Gene Editing of Resting CD4(+) T Cells by Transient ZFN Expression From HIV Envelope Pseudotyped Nonintegrating Lentivirus Confers HIV-1 Resistance in Humanized Mice. Mol Ther Nucleic Acids. 2014;3(9):e198.
29. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.
30. Zentner GE, Scacheri PC. The chromatin fingerprint of gene enhancer elements*. J Biol Chem. 2012;287(37):30888–96. Available from: https://www.sciencedirect.com/science/article/pii/S0021925820630356.
31. Thibodeau A, Uyar A, Khetan S, Stitzel ML, Ucar D. A neural network based model effectively predicts enhancers from clinical ATAC-seq samples. Sci Rep. 2018;8(1):16048. https://doi.org/10.1038/s41598-018-34420-9.
32. Sartorelli V, Lauberth SM. Enhancer RNAs are an important regulatory layer of the epigenome. Nat Struct Mol Biol. 2020;27(6):521–8. https://doi.org/10.1038/s41594-020-0446-0.
33. Han Z, Li W. Enhancer RNA: What we know and what we can achieve. Cell Prolif. 2022;55(4):e13202. https://doi.org/10.1111/cpr.13202.
34. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc. 2016;11(8):1455–76. https://doi.org/10.1038/nprot.2016.086.
35. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48(D1):D882–9.
36. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37.
37. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016;167(5):1369-1384.e19.
38. Ngan CY, Wong CH, Tjong H, Wang W, Goldfeder RL, Choi C, et al. Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. Nat Genet. 2020;52(3):264–72. https://doi.org/10.1038/s41588-020-0581-x.
39. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, et al. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev. 2007;21(4):385–90.
40. Kalay G, Lachowiec J, Rosas U, Dome MR, Wittkopp P. Redundant and Cryptic Enhancer Activities of the Drosophilayellow Gene. Genetics. 2019;212(1):343–60.
41. Laugesen A, Højfeldt JW, Helin K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. Cold Spring Harb Perspect Med. 2016;6(9):a026575.
42. Gisselbrecht SS, Palagi A, Kurland JV, Rogers JM, Ozadam H, Zhan Y, et al. Transcriptional silencers in drosophila serve a dual role as transcriptional enhancers in alternate cellular contexts. Mol Cell. 2020;77(2):324-337.e8. https://doi.org/10.1016/j.molcel.2019.10.004.
43. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 2015;25(11):1757–70.
44. Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet. 2018;50(5):699–707. https://doi.org/10.1038/s41588-018-0102-3.
45. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455–61.
46. Xing S, Gai K, Li X, Shao P, Zeng Z, Zhao X, et al. Tcf1 and Lef1 are required for the immunosuppressive function of regulatory T cells. J Exp Med. 2019;216(4):847–66. https://doi.org/10.1084/jem.20182010.
47. Labbé E, Letamendia A, Attisano L. Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-β and Wnt pathways. Proc Natl Acad Sci. 2000;97(15):8358–63. https://doi.org/10.1073/pnas.150152697.
48. Wei G, Abraham BJ, Yagi R, Jothi R, Cui K, Sharma S, et al. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity. 2011;35(2):299–311.
49. Kruse EA, Loughran SJ, Baldwin TM, Josefsson EC, Ellis S, Watson DK, et al. Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. Proc Natl Acad Sci. 2009;106(33):13814–9. https://doi.org/10.1073/pnas.0906556106.
50. Smeets MFMA, Chan AC, Dagger S, Bradley CK, Wei A, Izon DJ. Fli-1 overexpression in hematopoietic progenitors deregulates T cell development and induces pre-T cell lymphoblastic leukaemia/lymphoma. PLoS ONE. 2013;8(5):e62346.
51. Riggi N, Knoechel B, Gillespie SM, Rheinbay E, Boulay G, Suvà ML, et al. EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell. 2014;26(5):668–81.
52. Gil-Yarom N, Radomir L, Sever L, Kramer MP, Lewinsky H, Bornstein C, et al. CD74 is a novel transcription regulator. Proc Natl Acad Sci. 2017;114(3):562–7. https://doi.org/10.1073/pnas.1612195114.
53. Becker-Herman S, Rozenberg M, Hillel-Karniel C, Gil-Yarom N, Kramer MP, Barak A, et al. CD74 is a regulator of hematopoietic stem cell maintenance. PLoS Biol. 2021;19(3):e3001121.

54. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–89.
55. Crona F, Dahlberg O, Lundberg LE, Larsson J, Mannervik M. Gene regulation by the lysine demethylase KDM4A in Drosophila. Dev Biol. 2013;373(2):453–63.
56. Young NL, Dere R. Mechanistic insights into KDM4A driven genomic instability. Biochem Soc Trans. 2021;49(1):93–105.
57. Xi W, Beer MA. Loop competition and extrusion model predicts CTCF interaction specificity. Nat Commun. 2021;12(1):1046. https://doi.org/10.1038/s41467-021-21368-0.
58. Nishana M, Ha C, Rodriguez-Hernaez J, Ranjbaran A, Chio E, Nora EP, et al. Defining the relative and combined contribution of CTCF and CTCFL to genomic regulation. Genome Biol. 2020;21(1):108. https://doi.org/10.1186/s13059-020-02024-0.
59. Korinfskaya S, Parameswaran S, Weirauch MT, Barski A. Runx Transcription Factors in T Cells-What Is Beyond Thymic Development? Front Immunol. 2021;12:701924.
60. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem. 2011;80:437–71.
61. Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. Genome Res. 2013;23(8):1195–209.
62. Zhang Y, See YX, Tergaonkar V, Fullwood MJ. Long-distance repression by human silencers: chromatin interactions and phase separation in silencers. Cells. 2022;11(9):1560.
63. Pang B, van Weerd JH, Hamoen FL, Snyder MP. Identification of non-coding silencer elements and their regulation of gene expression. Nat Rev Mol Cell Biol. 2022. Available from: doi: https://doi.org/10.1038/s41580-022-00549-9.
64. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. Neuropsychopharmacology. 2013;38(1):23–38. https://doi.org/10.1038/npp.2012.112.
65. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer. 2011;2(6):607–17.
66. Chahrour M, Jung SY, Shaw C, Zhou X, Wong STC, Qin J, et al. MeCP2, a Key Contributor to Neurological Disease Activates and Represses Transcription. Science (80-). 2008;320(5880):1224–9. https://doi.org/10.1126/science.1153252.
67. Pachano T, Sánchez-Gaya V, Ealo T, Mariner-Faulí M, Bleckwehl T, Asenjo HG, et al. Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. Nat Genet. 2021;53(7):1036–49. https://doi.org/10.1038/s41588-021-00888-x.
68. Alajem A, Roth H, Ratgauzer S, Bavli D, Motzik A, Lahav S, et al. DNA methylation patterns expose variations in enhancer-chromatin modifications during embryonic stem cell differentiation. PLOS Genet. 2021;17(4):e1009498. https://doi.org/10.1371/journal.pgen.1009498.
69. Bal E, Kumar R, Hadigol M, Holmes AB, Hilton LK, Loh JW, et al. Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. Nature. 2022;607(7920):808–15. https://doi.org/10.1038/s41586-022-04906-8.
70. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, et al. The interaction landscape between transcription factors and the nucleosome. Nature. 2018;562(7725):76–81.
71. Whitehouse I, Tsukiyama T. Antagonistic forces that position nucleosomes in vivo. Nat Struct Mol Biol. 2006;13(7):633–40.
72. Saxton DS, Rine J. Nucleosome positioning regulates the establishment, stability, and inheritance of heterochromatin in saccharomyces cerevisiae. Proc Natl Acad Sci. 2020;117(44):27493–501. https://doi.org/10.1073/pnas.2004111117.
73. Creamer KM, Job G, Shanker S, Neale GA, Lin Y, Bartholomew B, et al. The Mi-2 homolog Mit1 actively positions nucleosomes within heterochromatin to suppress transcription. Mol Cell Biol. 2014;34(11):2046–61.
74. Yu Q, Zhang X, Bi X. Roles of chromatin remodeling factors in the formation and maintenance of heterochromatin structure. J Biol Chem. 2011;286(16):14659–69.
75. Zeng W, Chen S, Cui X, Chen X, Gao Z, Jiang R. SilencerDB: a comprehensive database of silencers. Nucleic Acids Res. 2021;49(D1):D221-8. https://doi.org/10.1093/nar/gkaa839.
76. Kartashov AV, Barski A. BioWardrobe: an integrated platform for analysis of epigenomics and transcriptomics data. Genome Biol. 2015;16(1):158.
77. He N, Wang W, Fang C, Tan Y, Li L, Hou C. Integration of Count Difference and Curve Similarity in Negative Regulatory Element Detection. Front Genet. 2022;13: 818344.
78. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160–5.
79. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033.

Publisher's Note