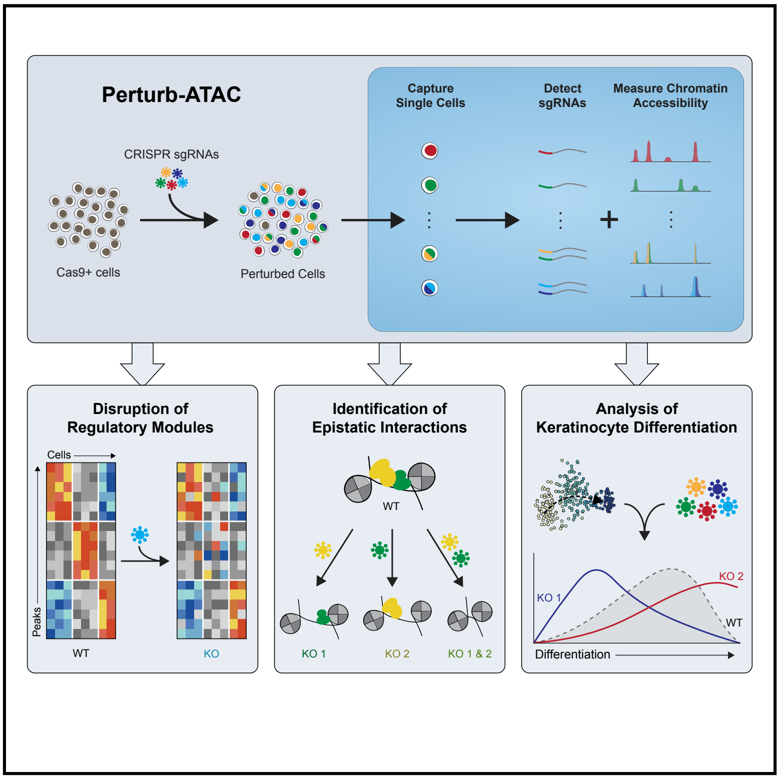
Resource

Cell

Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks

Graphical Abstract



Authors

Adam J. Rubin, Kevin R. Parker, Ansuman T. Satpathy, ..., William J. Greenleaf, Howard Y. Chang, Paul A. Khavari

Correspondence

howchang@stanford.edu (H.Y.C.), khavari@stanford.edu (P.A.K.)

In Brief

Perturb-ATAC combines CRISPR screening with chromatin accessibility profiling of single cells to uncover regulators of chromatin architecture and regulator occupancy and to determine epistatic relationships between regulatory factors in cell fate decisions.

Highlights

- A method to measure CRISPR perturbations and chromatin state in single cells
- Mapping of dynamic chromatin regulatory networks through intercellular variation
- Elucidating principles of epistatic interaction between *trans*-factors
- Perturb-ATAC screen of TF function in epidermal differentiation trajectories

Rubin et al., 2019, Cell *176*, 361–376
January 10, 2019
https://doi.org/10.1016/j.cell.2018.11.022

Cell

Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks

Adam J. Rubin,[1,9] Kevin R. Parker,[1,2,9] Ansuman T. Satpathy,[1,3,9] Yanyan Qi,[1] Beijing Wu,[1] Alvin J. Ong,[1,2] Maxwell R. Mumbach,[1,2,4] Andrew L. Ji,[1] Daniel S. Kim,[1] Seung Woo Cho,[1,2] Brian J. Zarnegar,[1] William J. Greenleaf,[4,5,6] Howard Y. Chang,[1,2,4,7,*] and Paul A. Khavari[1,8,10,*]

[1]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA
[3]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA
[6]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[7]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[8]Veterans Affairs Palo Alto Healthcare System, Palo Alto, CA 94304, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: howchang@stanford.edu (H.Y.C.), khavari@stanford.edu (P.A.K.)
https://doi.org/10.1016/j.cell.2018.11.022

SUMMARY

Here, we present Perturb-ATAC, a method that combines multiplexed CRISPR interference or knockout with genome-wide chromatin accessibility profiling in single cells based on the simultaneous detection of CRISPR guide RNAs and open chromatin sites by assay of transposase-accessible chromatin with sequencing (ATAC-seq). We applied Perturb-ATAC to transcription factors (TFs), chromatin-modifying factors, and noncoding RNAs (ncRNAs) in ~4,300 single cells, encompassing more than 63 genotype-phenotype relationships. Perturb-ATAC in human B lymphocytes uncovered regulators of chromatin accessibility, TF occupancy, and nucleosome positioning and identified a hierarchy of TFs that govern B cell state, variation, and disease-associated *cis*-regulatory elements. Perturb-ATAC in primary human epidermal cells revealed three sequential modules of *cis*-elements that specify keratinocyte fate. Combinatorial deletion of all pairs of these TFs uncovered their epistatic relationships and highlighted genomic co-localization as a basis for synergistic interactions. Thus, Perturb-ATAC is a powerful strategy to dissect gene regulatory networks in development and disease.

INTRODUCTION

Gene expression in eukaryotic organisms is regulated by the interplay of thousands of *trans*-acting regulatory factors and millions of *cis*-acting DNA elements (Kundaje et al., 2015). However, it is challenging to characterize each *trans*-factor or *cis*-element because epigenetic assays require large amounts of cellular material and are difficult to couple with genetic perturbations at scale. Therefore, most studies are limited to measuring how single genetic perturbations affect chromatin state in bulk cell populations.

We recently developed the assay for transposase-accessible chromatin with sequencing (ATAC-seq), which utilizes a hyperactive transposase (Tn5) to measure the activity of regulatory DNA elements (Buenrostro et al., 2013). This method informs the identification of enhancers, the positioning of nucleosomes, and the inference of transcription factor binding (Buenrostro et al., 2013; Schep et al., 2015, 2017). Importantly, ATAC-seq can be performed in single cells (scATAC-seq), uncovering cell-to-cell variability and rare epigenomic phenotypes (Cusanovich et al., 2018; Satpathy et al., 2018; Buenrostro et al., 2018). Additionally, single-cell ATAC-seq profiles can be paired with orthogonal measurements of RNA or protein expression in the same cell (Satpathy et al., 2018; Chen et al., 2018a).

Similar advances in the ability to measure transcriptomes in single cells have recently allowed high-throughput genetic screens coupled with simultaneous transcriptome phenotyping (Dixit et al., 2016; Adamson et al., 2016; Jaitin et al., 2016; Datlinger et al., 2017). We further develop this concept of high-content genetic screening by measuring the effects of CRISPR perturbations on the epigenome. This method, termed perturbation-indexed single-cell ATAC-seq (Perturb-ATAC), measures CRISPR single guide RNA (sgRNA) sequences and ATAC-seq profiles in single cells. We performed Perturb-ATAC in 2,936 immortalized B lymphoblasts and 1,356 primary human keratinocytes, encompassing 63 genotype-phenotype relationships. Analysis of a CRISPR interference (CRISPRi) screen in GM12878 lymphoblasts identified *trans*-factor control of several layers of epigenetic regulation. Single-cell ATAC-seq in primary human epidermal cells identified three regulatory modules in keratinocyte differentiation, and a Perturb-ATAC CRISPR-deletion screen revealed key transcription factor (TF) regulators of each module. We mapped epistatic relationships between TFs using multiplexed perturbations and suggest that genomic co-localization and co-expression of TFs may predict genetic interaction. Perturb-ATAC is a tool for studying relationships between factors that control chromatin states using high-throughput, high-complexity single-cell screens.

RESULTS

Perturb-ATAC: Simultaneous CRISPR Guide Detection and Epigenome Profiling in Single Cells

In the Perturb-ATAC protocol, cells are first captured on the integrated fluidics circuit (IFC; Fluidigm) in single-cell chambers and subjected to lysis and DNA transposition with the Tn5 enzyme (Figure 1A). After transposition, Tn5 is released from open chromatin fragments, and CRISPR sgRNAs or sgRNA-identifying barcodes from each cell are reverse-transcribed using target-specific primers. All chamber contents are then amplified by PCR. Single-cell libraries are then collected, and sgRNA or ATAC amplicons are further amplified separately with cell-identifying barcoded primers, pooled, and sequenced.

We first adapted a CRISPRi sgRNA vector to perform a simple mixing experiment with three populations of sgRNA-targeted cells (Adamson et al., 2016; Cho et al., 2018). We generated a sgRNA vector with unique 22- bp guide barcodes (GBCs) that corresponded to the identity of sgRNAs encoded by each vector (Figures 1A and S1A; Table S1). Using this vector, we targeted immortalized B lymphoblasts stably expressing dCas9-Krüppel-associated box (KRAB) with either non-human genome-targeting (NT) sgRNAs (NT1 or NT2) or sgRNAs targeting the promoter of the TF SPI1 (also known as PU.1), which is required for B cell development (Scott et al., 1994, 1997; McKercher et al., 1996). We then pooled cells and performed Perturb-ATAC on 309 single cells to assess the fidelity of pairing GBC detection with measurement of the epigenome. To eliminate the likelihood of recombination of sgRNAs and GBCs between vectors, as described recently(Adamson et al., 2018; Feldman et al., 2018; Hill et al., 2018; Xie et al., 2018), we performed packaging steps in separate cultures. Transduced cells were enriched by fluorescence-activated cell sorting (FACS) purification of mCherry+ cells and identified the sgRNA identity of each cell by amplifying the GBC (Figure 1A).

We used stringent cutoffs to assign the presence or absence of a GBC (Figure S1B). First, we counted reads for each possible GBC in every cell and adjusted for sequencing depth (Figures S1B and S1C). Next, we set a minimum cutoff of 1,000 GBC reads per cell and removed cells with a high percentage of background reads (STAR Methods). As expected, reads were almost exclusively assigned to one GBC (NT1, NT2, or SPI1); multiple GBCs were only detected in 9 of 309 cells. These results demonstrate that Perturb-ATAC consistently detects GBCs with high confidence and a low false positive rate (Figures 1B, S1B, and S1C).

Next, we evaluated the quality of ATAC-seq reads; high-quality single-cell ATAC-seq profiles were obtained in 79.2% of cells in which GBCs were also detected (Figure 1C). Cells passing filtering yielded an average of 11.33×10^3 fragments mapping to the nuclear genome, and approximately 43.05% of reads were within bulk ATAC-seq peaks, similar to previously published single-cell ATAC-seq as well as high-quality bulk ATAC-seq datasets (Buenrostro et al., 2015; Corces et al., 2017; Figure 1C). Single-cell ATAC-seq reads recapitulated the characteristics of bulk ATAC-seq data, including insert size periodicity and fragment enrichment at transcription start sites (Figure 1D). These results indicate that GBC detection does not interfere with the generation of ATAC-seq libraries in single cells.

Finally, we asked whether GBC-expressing cells exhibited the expected ATAC-seq phenotype. We first examined the promoter of *SPI1*, where *SPI1*-targeted sgRNAs were expected to recruit dCas9-KRAB. In cells expressing *SPI1* sgRNAs, we observed a loss of accessibility at the promoter compared with cells expressing NT sgRNAs (Figure 1E). More broadly, *SPI1*-targeted cells exhibited similar changes in accessibility across all peaks that were changed in bulk ATAC-seq experiments (Figure 1F). We then measured global TF activity by determining the relative accessibility of all SPI1 motif-containing sites (51,862 sites). We found that SPI1 sites exhibited a significant loss of accessibility in *SPI1*-targeted cells compared with NT cells (false discovery rate [FDR] < 1e−3, permutation test; STAR Methods), similar to bulk *SPI1*-targeting experiments (Figure 1G). Altogether, these findings demonstrate that Perturb-ATAC simultaneously measures GBC sequences and ATAC-seq data in single cells with high fidelity.

Perturb-ATAC Identifies Epigenomic Functions of Chromatin Regulators, Transcription Factors, and Noncoding RNAs in B Cells

We next performed an expanded Perturb-ATAC screen to compare how broadly expressed and lineage-specific *trans*-factors shape the chromatin landscape of B lymphoblasts. We generated 40 sgRNA genotypes in 2,627 single cells, derived from single or dual targeting of the promoters of 12 *trans*-factors and 2 NT control sgRNAs (Table S2). The 12 targeted *trans*-factors included TFs (EBF1, IRF8, NFKB1, RELA, and SPI1), chromatin modifiers (BRG1, DNMT3A, EZH2, and TET2), and noncoding RNAs (7SK, EBER1, and EBER2) that have been shown previously to affect normal and neoplastic B cell development and function (Nutt and Kee, 2007; Lunning and Green, 2015). CRISPRi guide RNAs were designed to maximize knockdown efficiency and minimize off-target effects (Figures S2A–S2C; Table S1). We distinguished cells receiving either one or two GBCs, as described above (Figures 2A and 2B). High-quality ATAC-seq profiles were obtained in 85.3% of cells in which GBCs were also detected. Cells passing filtering yielded an average of 10.68×10^3 fragments mapping to the nuclear genome, and approximately 43.85% of reads were within peaks present in bulk profiles (Figure 2C). Rare ATAC-seq reads mapping to the viral construct did not appear to influence ATAC-seq profiles, and observed accessibility at potential sgRNA mismatch loci indicated no evidence of off-target repression (STAR Methods; Figure S2D). Cells processed across several microfluidic chips exhibited no noticeable chip bias in ATAC-seq signal (Figure S2E).

We aggregated cells based on GBC identity and analyzed three levels of epigenetic regulation: accessibility of DNA elements, *trans*-factor activity genome-wide, and nucleosome positioning (Figure 2D). Depletion of EZH2, a catalytic subunit of Polycomb

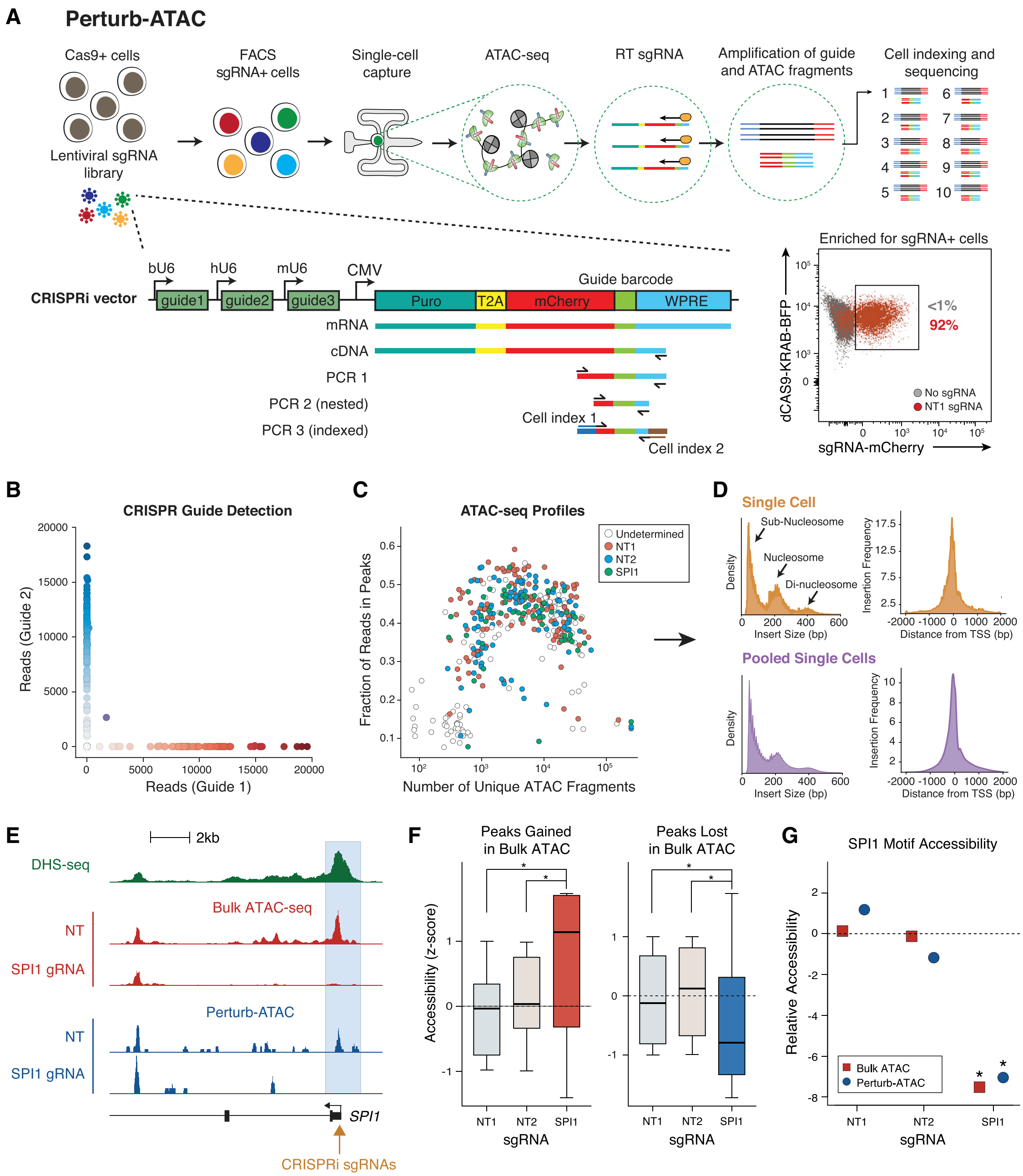


Figure 1. Perturb-ATAC Identifies sgRNA Barcodes and Expected Chromatin Phenotypes in Single Cells

(A) Schematic of the Perturb-ATAC protocol, lentiviral construct, and sequencing library generation for sgRNA detection.

(B) Scatterplot of guide barcode (GBC) reads from a pool of cells transduced with one of two constructs.

(C) Scatterplot of ATAC fragments and the fraction of ATAC fragments in peak regions for each cell. Colors indicate GBC detection in each cell.

(D) Histograms of ATAC fragment size distribution, indicating the expected nucleosome phasing (left) and relative frequency of ATAC insertions at transcription start sites (right).

(E) Genomic locus of the *SPI1* gene, indicating DNase I hypersensitivity sequencing, bulk ATAC-seq, and Perturb-ATAC-seq. The SPI1 promoter region exhibits selective loss of accessibility in cells expressing SPI1 sgRNA.

(F) Accessibility in merged single cells of individual genomic regions altered in bulk ATAC-seq. *$p < 1e-3$ by Kolmogorov-Smirnov (KS) test.

(G) Relative accessibility of SPI1 motif-containing regions (*Z* score of the SPI1 motif versus all other genomic features). *false discovery rate < 1e−3 by permutation test.

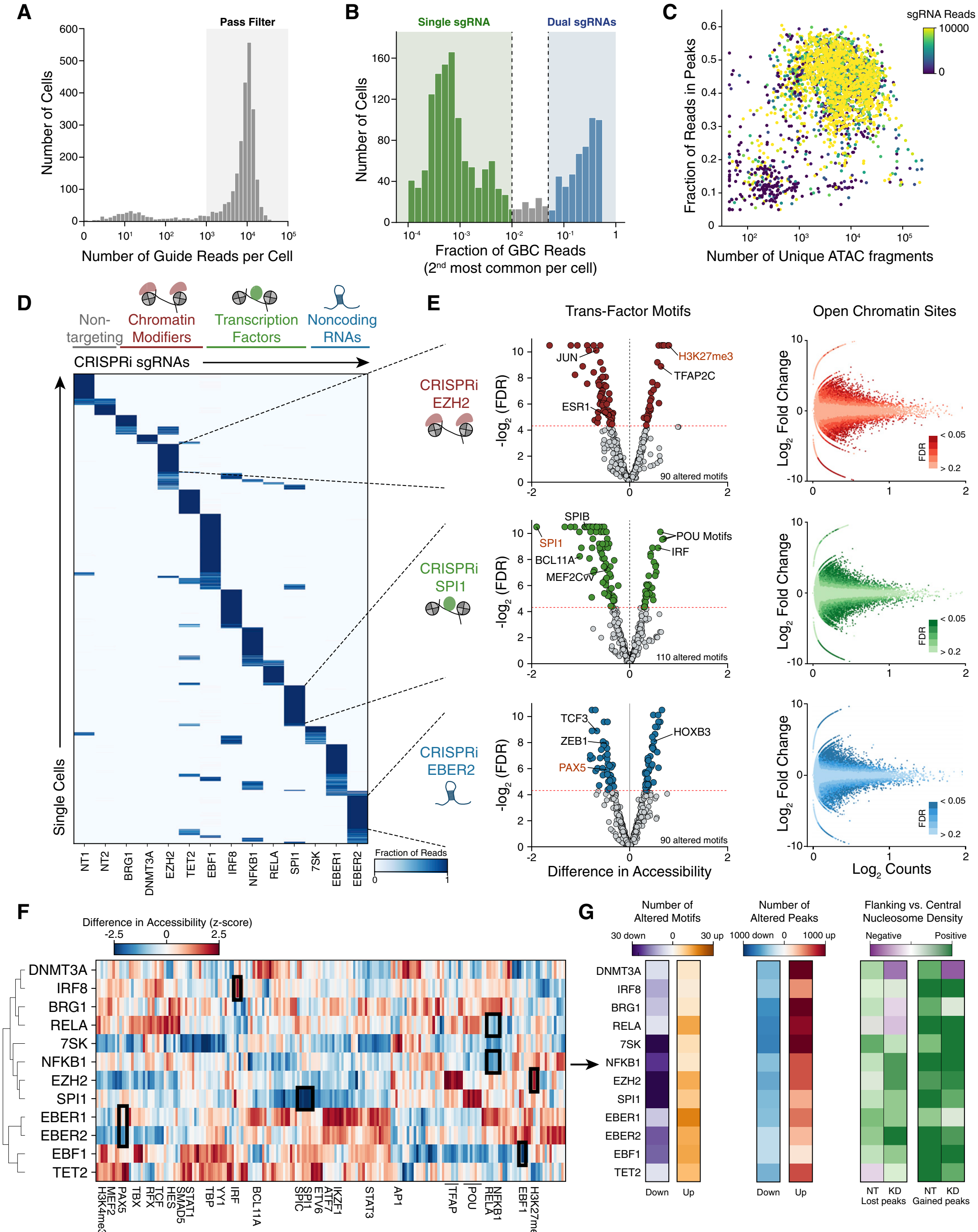


Figure 2. Perturb-ATAC Screen for Control of Accessibility Landscape by Transcription Factors, Long Noncoding RNAs, and Chromatin Regulators

(A) Histogram of total GBC reads per cell.

(B) Histogram of the second most common GBC identified in each cell. Cells on the low end of the distribution express a single guide RNA, whereas cells on the high end express two guide RNAs.

(C) Scatterplot of ATAC fragments and fraction of fragments in peak regions. Cells are colored by GBC read count.

(D) Heatmap of cells (rows) versus GBCs (columns), indicating the proportion of reads associated with each barcode.

(E) Left: volcano plots showing significantly altered genomic features between cells carrying non-targeting (NT) guides and guides targeting EZH2, SPI1, and EBER2 (FDR $\leq$ 0.025). Right: scatterplots of mean accessibility versus accessibility fold change of individual genomic peaks.

(legend continued on next page)

repressive complex 2 (PRC2), which deposits repressive Histone 3 lysine 27 trimethylation (H3K27me3) chromatin marks, significantly increased accessibility at regions marked with H3K27me3 (chromVAR accessibility gain, 0.78; FDR < 0.0001; Figure 2E). Similarly, the most significant change in TF motif accessibility in *SPI1*-targeted cells was at SPI1-motif-containing regions (chromVAR accessibility loss, 2.17; FDR < 0.0001; Figure 2E). In contrast, SPI depletion increased the accessibility of IRF motif regions, demonstrating that SPI1 exhibits activating and repressive functions, as reported previously (van Riel and Rosenbauer, 2014). Other altered TF motifs in SPI1-targeted cells included BCL11A, SPIB, and MEF2C factors, consistent with the function of SPI1 in regulating factors required for B cell lineage commitment (Figure 2E; Table S3; Su et al., 1996; Liu et al., 2003; Stehling-Sun et al., 2009). Finally, targeting of the noncoding RNA (ncRNA) EBER2 identified 90 significantly altered features, including regions containing the PAX5 motif, a factor that physically interacts with EBER2 to control gene expression (Lee et al., 2015). These results demonstrate that Perturb-ATAC identifies epigenomic phenotypes associated with perturbations of diverse categories of *trans*-factors.

An analysis of ATAC-seq profiles derived from 40 single- and double-GBC genotypes (including NT controls) revealed accessibility changes in 10,103 open chromatin sites (mean, 404; range, 0–2,250 sites; per genotype) and 833 features (mean, 23; range, 0–110 features; per genotype; Figures 2F and S3A–S3C; Table S3). We clustered feature accessibility across all perturbations and found three sub-clusters of *trans*-factors with correlated effects (Figure 2F). Cluster 1 included the B cell lineage-determining TFs IRF8 and RELA and the chromatin regulators BRG1 and DNMT3A, suggesting that these factors may cooperate to establish the B cell chromatin landscape (Corces et al., 2016; Lara-Astiaso et al., 2014). Interestingly, two nuclear factor κB (NF-κB) subunits, NFKB1 and RELA, showed overlapping and distinct effects, in line with prior studies demonstrating differences in the binding patterns of NF-κB subunits through the formation of distinct protein complexes (Figures 2F and S3B; Zhao et al., 2014). Consistent with reports of its repressor activity, IRF8 depletion increased accessibility at IRF sites (Tamura et al., 2008). Cluster 2 included the ncRNA 7SK, which represses enhancer transcription (Flynn et al., 2016), and the chromatin remodeler EZH2 (Figure 2F). Cluster 3 included the highly homologous ncRNAs EBER1 and EBER2 as well as TET2 and EBF1. Cells depleted of EBER1 or EBER2 exhibited similar ATAC-seq profiles, highlighting their functional similarity ($r = 0.714$, $p = 2.25e{-}85$; Figures 2F and S3B; Arrand et al., 1989; Samanta et al., 2006), and loss of either factor altered accessibility at EBF1 sites, consistent with these factors acting in the same pathway (Figure S3B).

Finally, we used the diversity in ATAC-seq fragment sizes to infer the occupancy and positioning of nucleosomes in each perturbation condition (Schep et al., 2015; Figure S3D; STAR Methods). To validate this approach, we examined the profiles of sub-nucleosome-sized and nucleosome-sized fragments surrounding CTCF binding sites (Figure S3E). As expected, sub-nucleosome fragments were enriched at the CTCF motif, indicating a central nucleosome-free region. In contrast, nucleosome-sized fragments were enriched upstream and downstream relative to the CTCF motif, representing the +1 and −1 nucleosomes (Vierstra et al., 2014).

We quantified a score representing the flanking accumulation and central depletion of nucleosome-sized ATAC-seq fragments in each perturbation (Figure 2G). At regions exhibiting an altered ATAC-seq signal, we compared this score between control and perturbed cells. Although some factors operated in the context of stable nucleosomes, others altered local nucleosome profiles. Depletion of either NFKB1 or RELA resulted in a stable nucleosome structure surrounding regions that gained accessibility, suggesting that these factors influence the binding of co-factors in an independently established nucleosome-free region (Figure S3F). However, regions that gained accessibility in DNMT3A-depleted cells exhibited a stronger central nucleosome, consistent with a model in which DMNT3A recruits negative regulatory factors to an open chromatin region. Reflecting another distinct mechanism, regions losing accessibility upon depletion of 7SK exhibited a stronger central nucleosome structure in 7SK-depleted cells, possibly indicating that 7SK interacts with nucleosome remodeling factors, as shown for the BAF complex (Figure S3F; Flynn et al., 2016). Overall, these results demonstrate that *trans*-factors regulate nucleosome positioning in distinct ways and that Perturb-ATAC can read out nucleosome structure changes associated with perturbations.

Discovery of Gene Regulatory Networks Controlled by *trans*-Factors

We next analyzed Perturb-ATAC profiles to identify relationships between *trans*-factors. We were inspired by studies analyzing intercellular heterogeneity to infer relationships between cellular features (Klein et al., 2015; Heath et al., 2016). For example, co-varying accessibility between sets of regions bound by two *trans*-factors may reflect common regulation. In contrast, inverse correlation may indicate that one factor controls the activity or expression of a negative regulator of the other set of regions. Using this framework, we measured the effects of perturbation on co-varying regulatory networks (Figure 3A).

We assessed feature correlations in NT cells and identified five modules (Figures 3B, S4A, and S4B; STAR Methods). Module 1 features included factors broadly involved in hematopoietic development, such as SPI1 and IRF8. Module 2 included the chromatin regulators CTCF and CHD1. ETS factors, which have roles in B cell development and immunological function, were highly correlated in module 3. Module 4 contained homeobox domain TFs, and module 5 features included specific regulators of B cell development, such as IKZF1 (also known as Ikaros), BCL11A, EBF1, NFKB, MEF2C, and AP-1 factors.

(F) Heatmap of perturbed factors (rows) versus genomic annotations (columns) indicating the difference in accessibility between perturbed and NT control cells. Only annotations significantly altered in at least one perturbation are shown.

(G) Heatmaps indicating the number of significantly altered features (left, absolute chromVAR deviation $Z \geq 0.75$, $FDR \leq 0.05$), the number of altered genomic regions (center, absolute log2 fold change [log2FC] ≥ 1.5, mean reads per ce 0.4), or quantification of the ratio of flanking to central nucleosome occupancy at altered peaks (right) for each single perturbation.

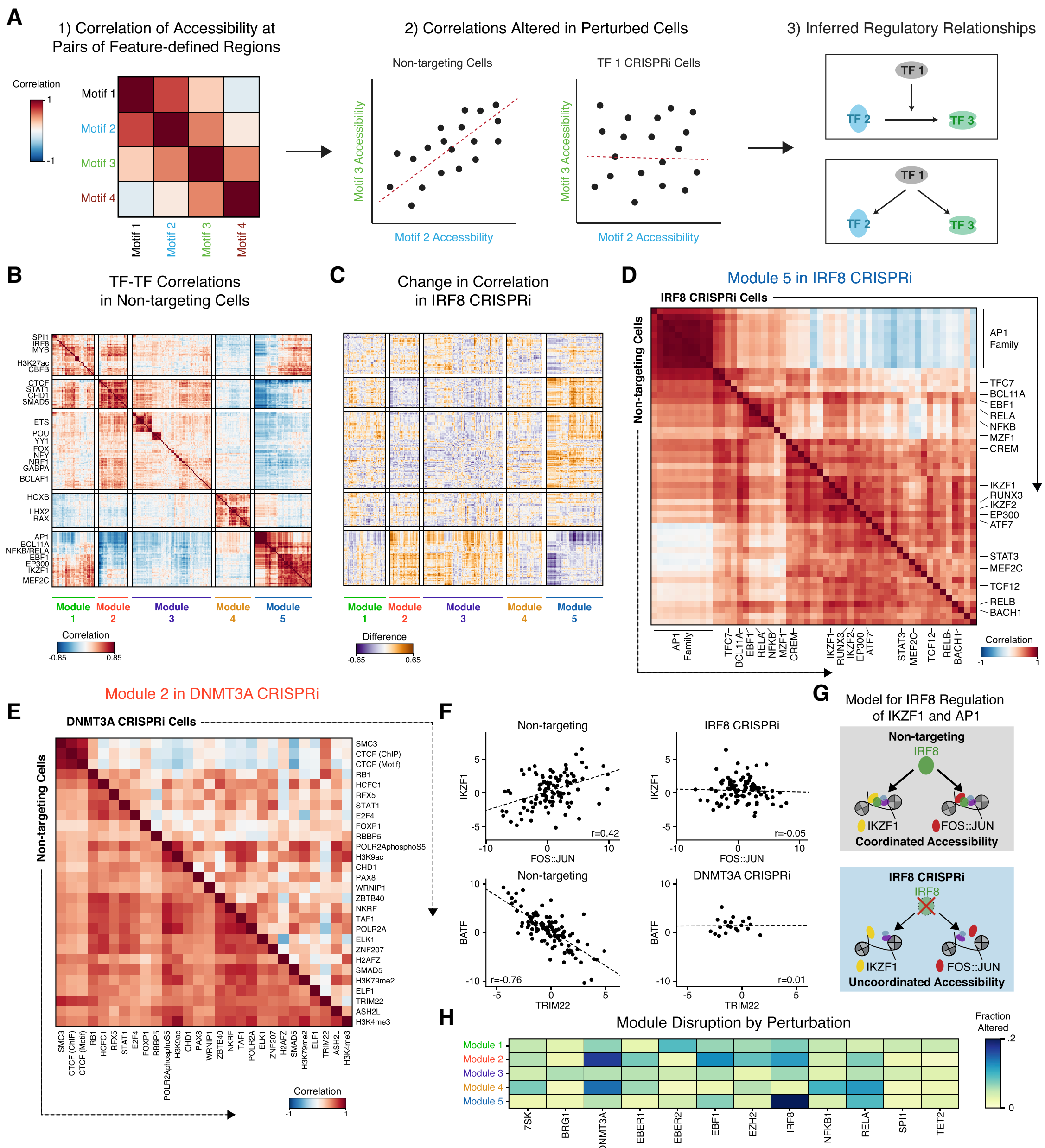


Figure 3. Perturbations Influence Inter-cellular Variability and Correlated Activity across Features

(A) Example workflow identifying genomic features with correlated activity across cells. Left: heatmap indicating correlation of motif activity across cells. Center: comparing NT control cells with perturbed cells identifies motif pairs that change in correlation as a result of perturbation. Right: functional relationships constrain hypothetical regulatory networks.

(B) Heatmap of Pearson correlations between features in NT cells.

(C) Heatmap displaying the difference in correlations between NT and IRF8 knockdown cells.

(D) Heatmap of module 5 feature correlations in NT (bottom half) and IRF8 (top half) knockdown cells.

(E) Heatmap displaying module 2 feature correlations in NT cells (bottom half) and DNMT3A (top half) knockdown cells.

(F) Scatterplots of accessibility for cells with line of linear best fit demonstrating correlation under specific conditions.

(G) Hypothetical model of IRF8 co-factor activity with AP-1 and IKZF1.

(H) Heatmap of the fraction of altered feature-feature correlations within modules by perturbation, showing specific effects on particular modules in different perturbations.

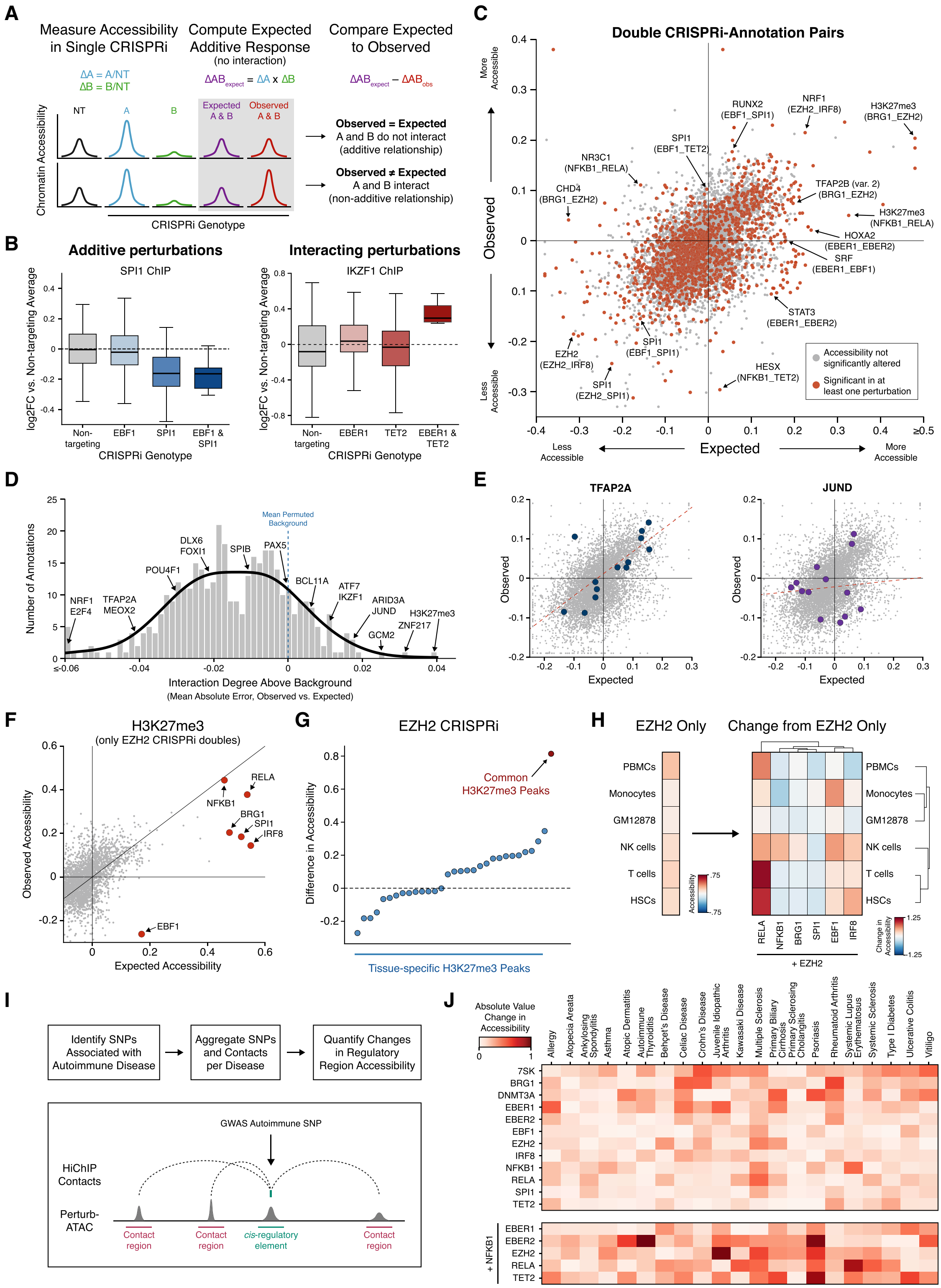


Figure 4. Epistasis Analysis Identifies the Functional Interaction between a Broadly Active Chromatin Regulator and Lineage-Specific Transcription Factors

(A) Schematic of calculation of expected accessibility in double knockdown based on the additive model of each single knockdown.

(B) Distribution of accessibility at SPI1 binding sites (left) and IKZF1 binding sites (right) for individual cells under single or double knockdown conditions.

(legend continued on next page)

Network reconstruction in perturbed cells highlighted the modules regulated by each perturbed factor (Figures 3C–3E). For example, depletion of IRF8 re-wired the developmental features of module 5 so that AP-1 factor activity was no longer coordinated with the activity of IKZF1, RUNX, and MEF2C (Figures 3D and S4C). In particular, the correlation between IKZF1 and FOS:JUN (AP-1) in NT cells ($r = 0.420$, $p = 5.60e-6$) was lost in IRF8-depleted cells ($r = -0.051$, $p = 0.627$), suggesting that IRF8 coordinates these two factors (Figures 3F and 3G). Indeed, IRF8 regulates the expression and activity of IKZF1 in pre-B cells (Ma et al., 2008; Pang et al., 2016). DNMT3A depletion altered the coordinated activity of CTCF with other module 2 factors, including SMAD5 (Figure 3E). Of note, the loss of co-variation is not dependent on altered accessibility of each factor, decoupling two mechanisms of TF activity (Figures 3E and S4D).

Similar interactions were observed for other factors (Figures 3H, S4E, and S4F). The effect of perturbation on a module was summarized as the number of significantly altered correlations (Figures S4G and S4H; STAR Methods). For example, depletion of EBF1 resulted in altered correlation of module 2 as well as its own module 5. The two NF-κB subunits, NFKB1 and RELA, exhibited distinct effects on module 5, which includes NFKB1 and RELA, while altering module 4 to a similar extent (Figures 3H and S4E). Finally, depletion of EBER2 altered interactions between AP-1, IRF, and NFKB, possibly through viral RNA binding and activation of viral sensors (Figure S4F; Samanta et al., 2006).

Mapping Epistatic Relationships Reveals Cooperative Functions of *trans*-Factors in Development and Disease

We generated dual-perturbed cells to analyze epigenetic epistasis. For each pair of perturbations, we computed an expected additive change in accessibility at each genomic feature from single-perturbed cells and compared that value with the observed accessibility in dual-perturbed cells (Dixit et al., 2016; Jaitin et al., 2016; Figure 4A; STAR Methods). For example, SPI1 motif sites were unchanged in EBF1-depleted cells, whereas depletion of SPI1 strongly decreased the accessibility of these sites (Figure 4B). In cells depleted of both EBF1 and SPI1, the observed accessibility matched the expected accessibility based on the combined effect of each factor alone, suggesting that these factors do not interact. Conversely, depletion of either EBER1 or TET2 alone did not affect accessibility at IKZF1 binding sites, whereas dual depletion resulted in an unexpected increase in accessibility (Figure 4B).

We measured the degree of interaction across all genomic features and categorized features as additive or non-additive (Figures 4C and 4D). For example, regions containing TFAP2A motifs were generally regulated by additive interactions (bottom 5% of all features), whereas regions containing JUND motifs were generally regulated by non-additive interactions (top 5% of all features) (Figures 4D and 4E).

Similarly, regions marked by the repressive histone modification H3K27me3 exhibited a high degree of interaction, particularly involving EZH2, suggesting that other factors may guide EZH2 recruitment or activity (Figure 4F). Because many EZH2-interacting factors were B cell lineage-determining TFs (EBF1, IRF8, and RELA), we reasoned that these interactions may repress alternate lineages. Indeed, depletion of EZH2 alone primarily de-repressed regions commonly marked by H3K27me3 across cell types, which may not require additional targeting specificity, whereas dual depletion of EZH2 and other factors de-repressed regions specifically active in alternate hematopoietic lineages (Figures 4G and 4H). For example, EBF1 cooperated with EZH2 to repress alternate monocyte and natural killer cell fates, whereas IRF8 and RELA cooperated with EZH2 to repress progenitor fates (Figure 4H).

Finally, we used Perturb-ATAC to inform regulators of noncoding genetic variants associated with human disease. We examined 21 autoimmune diseases, a subset of which demonstrate an enrichment of causal variants in B cell-specific enhancers (Farh et al., 2015). We measured the effect of each perturbation on the accessibility of regions proximal to or engaging in chromatin contacts with causal variants (Farh et al., 2015; Mumbach et al., 2017; Figure 4I; STAR Methods). NF-κB binding sites have been shown to be enriched near causal variants associated with multiple sclerosis and systemic lupus erythematosus (Farh et al., 2015), and these sites demonstrated altered accessibility in cells where *NFKB1* or *RELA* were depleted (Figure 4J). Although perturbation of NFKB1 did not alter many other variant enhancers in isolation, dual perturbation of NFKB1 and other factors modulated accessibility associated with several diseases, demonstrating that Perturb-ATAC uncovers epistatic interactions relevant to development and disease (Figure 4J).

The Regulatory Landscape of Human Epidermal Differentiation

Dynamic systems such as tissue differentiation present an opportunity for high-content screens to assess *trans*-factors while internally controlling for experimental variation. The human

(C) Scatterplot of observed versus expected accessibility for epistatic interactions. Each dot represents a single annotation in the pairing of two perturbed factors. Dots highlighted in red indicate significantly altered activity in either single or double perturbation.
(D) Histogram of background-corrected interaction degree for each feature. Background distribution was calculated by permuting single and double knockdown associations.
(E) Scatterplots of observed versus expected interactions, highlighting TFAP2A (relatively low interaction degree) and JUND (relatively high interaction degree).
(F) Scatterplot of observed versus expected change in accessibility at H3K27me3-marked regions in cells depleted of EZH2 and one other factor.
(G) Scatterplot of change in accessibility in EZH2 knockdown cells for subsets of H3K27me3 peaks. Common peaks have H3K27me3 marks across a majority of cell types.
(H) Left: heatmap indicating change in accessibility because of EZH2 depletion at regions marked by H3K27me3 in GM12878 and exhibiting the H3K27ac mark in each specific other cell type. Right: heatmap indicating change in accessibility of the same regions for cells simultaneously depleted of EZH2 and a transcription factor (TF).
(I) Workflow to aggregate SNPs associated with autoimmune diseases with 3D chromatin contact regions.
(J) Heatmap of the absolute change in accessibility for the SNP-contact feature set of each autoimmune disease and perturbation.

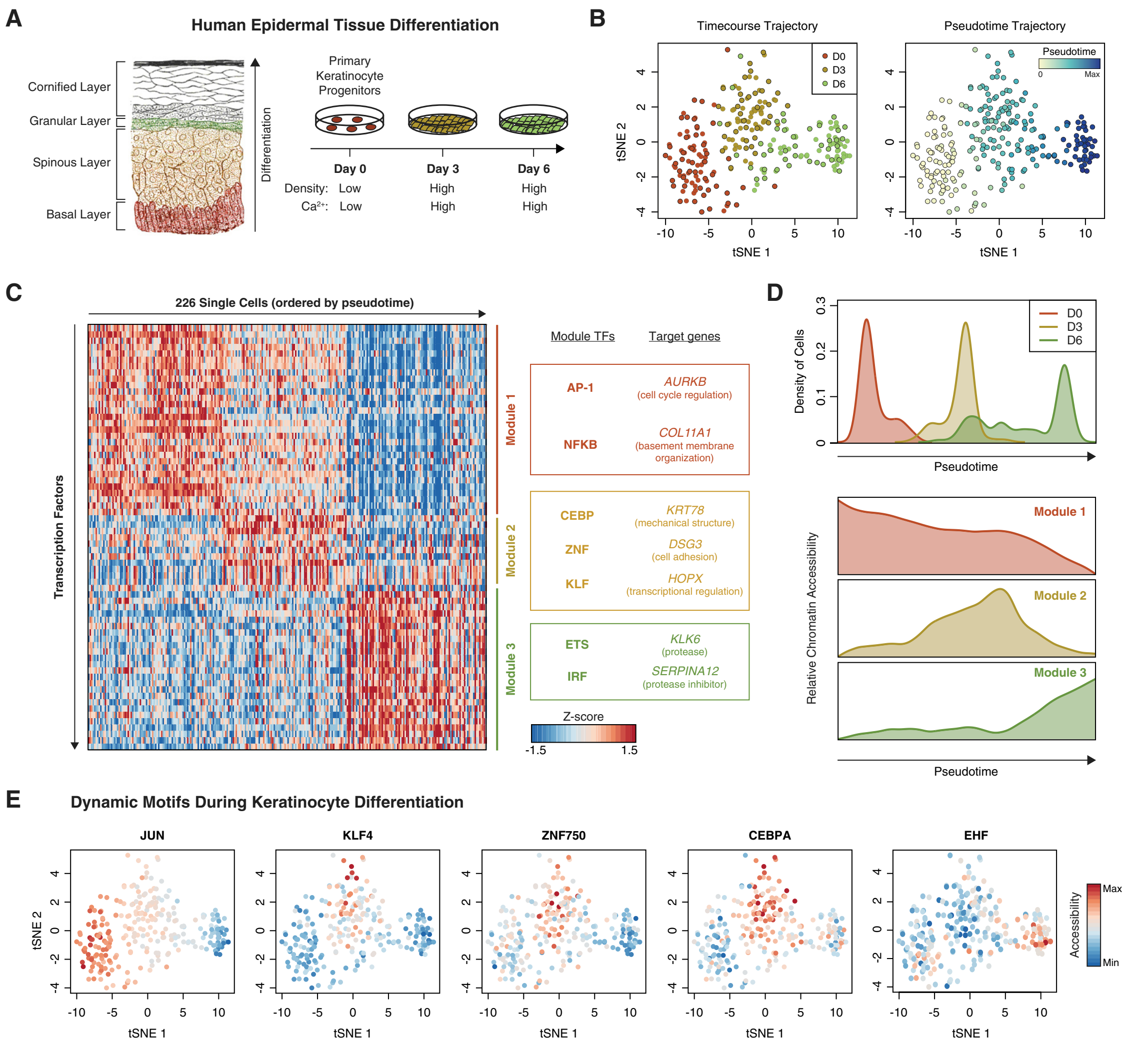


Figure 5. Modules of Transcription Factors Exhibit Distinct Temporal Activity in Epidermal Differentiation

(A) Schematic of human epidermis and cell culture model of epidermal differentiation (Gray, 1918).

(B) t-distributed stochastic neighbor embedding (t-SNE) projection of TF feature activity for epidermal cells labeled by differentiation day (left) or pseudotime (right).

(C) Heatmap of cells ordered by pseudotime (columns) versus TF feature activity (filtered for motifs with dynamic activity). Modules represent collections of TF features with similar temporal profiles. Target genes are proximal (<50 kb) to genomic regions associated with that module.

(D) Top: histogram of pseudotime values for cells from each day of differentiation. Bottom: average accessibility of each module identified in (C).

(E) t-SNE projections showing TF activity dynamics during differentiation.

epidermis is a constantly renewing stratified epithelial tissue, and thousands of genes change in expression during terminal epidermal differentiation as progenitor cells migrate from the basal layer, begin keratinization, and undergo cornification to form the outer layer of the skin (Lopez-Pajares et al., 2015). We sought to understand the role of dynamic chromatin in differentiation by perturbing key TFs.

We first assessed the landscape of chromatin accessibility in keratinocyte differentiation to identify candidate regulators for Perturb-ATAC. We cultured human primary progenitor keratinocytes under differentiation conditions and performed single-cell ATAC-seq on cells from each of three time points to capture undifferentiated, mid-differentiation, or late differentiation cells (Kretz et al., 2013; Figure 5A; STAR Methods). Although cells largely separated by their differentiation time point, a few "precocious" differentiating cells were observed in the day 0 population, and some day 6 cells remained in a mid-differentiation state (Figure 5B). Placement of each cell along a pseudotime trajectory highlighted the continuous nature of differentiation (Figure 5B; Qiu et al., 2017).

We next determined a set of TFs to perturb by identifying highly dynamic *cis*-regulatory features. We analyzed the accessibility of 94,633 *cis*-elements at the level of 411 TF motifs and chromatin immunoprecipitation sequencing (ChIP-seq) peaks (Schep et al., 2017). We identified 67 TFs with dynamic accessibility in differentiation that clustered into three modules (Figures 5C and 5D). The first module (30 features, including AP-1 and NF-κB) had high accessibility in progenitor cells that decreased in differentiation. Genes proximal to module 1 regions included *AURKB*, a regulator of mitosis, and *COL11A1*, a component of the basement membrane. Module 2 (12 features, including CEBP, KLF, and ZNF750), exhibited high accessibility specifically during mid-differentiation. (Figure 5B). The genes *KRT78*, *DSG3*, and *HOPX* were proximal to module 2 regions, representing programs of keratinization, cell-cell adhesion, and transcriptional regulation, respectively. Finally, module 3 (25 features) included target regions of ETS and IRF family factors and gained accessibility in late differentiation. Associated genes included *KLK6* and *SERPINA12*, a protease and protease inhibitor, respectively, which regulate coordinated desquamation of cells from the epidermis.

We used this analysis to prioritize candidate TFs for characterization with Perturb-ATAC. For a candidate module 1 regulator, we selected JUNB, an AP-1 family member that controls epidermal homeostasis and tumorigenesis (Eckert et al., 2013). We chose three module 2 regulators: KLF4, which induces epidermal differentiation genes; ZNF750, a factor engaged in positive and negative gene regulation in epidermal differentiation; and CEBPA, which has a known role in murine epidermis (Lopez et al., 2009; Sen et al., 2012). As a module 3 regulator, we chose EHF, an ETS factor implicated in controlling epidermal differentiation (Rubin et al., 2017). Closer analysis of these TFs confirmed patterns of dynamic accessibility matching their module (Figure 5E).

Perturb-ATAC Screen for TF Control of Cellular Differentiation Trajectories

We developed a second Perturb-ATAC protocol to achieve two goals: to directly detect the sgRNA rather than a GBC and to assess CRISPR gene knockouts rather than CRISPRi. In this workflow, keratinocytes were first transduced with a lentivirus encoding Cas9 (STAR Methods). Subsequently, Cas9-expressing keratinocytes were transduced with either one or two sgRNAs corresponding to the five TFs (JUNB, KLF4, ZNF750, CEBPA, and EHF) or two NT control sgRNAs producing both single- and double-targeted cells (Figures S5A and S5B; Table S4). Cells were maintained under undifferentiated culture conditions for 7–10 days to allow Cas9 disruption of target genes. Pooled cells were then transferred to differentiation conditions and harvested on day 3.

To detect sgRNAs, we performed reverse transcription using a reverse primer that matched the common 3′ end of the sgRNA, followed by PCR amplification using a pool of forward primers matching the variable 5′ ends of the sgRNAs (Figures 6A and S6A; Table S4). To analyze sgRNA sequencing reads and assign Perturb-ATAC genotypes to cells, we first set a plate-specific depth cutoff to remove failed sequencing reactions (Figure S6B). An average of 87.9% of reads mapped to an sgRNA (Figure S6C). Cells with high background reads (greater than 1%) were excluded, and the remaining cells clearly separated into single or double sgRNA-expressing populations (Figures S6D–S6F). Altogether, we identified 279 single sgRNA-expressing cells and 235 double sgRNA-expressing cells encompassing 23 distinct genotypes of all expected individual and double CRISPR perturbations (Figure 6B; Table S5).

We identified 399 features altered across TF knockout cells (FDR < 0.1). These features included multiple instances of TFs autoregulating their target sites (Figure 6C; Table S6). Notably, ZNF750 target regions were not uniformly more or less accessible in *ZNF750*-targeted cells, consistent with the role of this factor as a positive and negative gene regulator (Boxer et al., 2014). For example, the ZNF750-bound enhancer of the epidermal cornified envelope gene *SPRR2E* lost accessibility upon *ZNF750* disruption, whereas other ZNF750 binding sites gained accessibility (Figures 6D and S7A). Globally, *ZNF750* knockout resulted in a nearly even total of 649 gained and 620 lost peaks (FDR < 0.01, FC > 1). Module 2 factors, which had highest accessibility mid-differentiation, displayed diverging roles in differentiation. For example, perturbation of *ZNF750* increased accessibility at TP63 and NF-κB motif sites, two factors known to regulate epidermal identity (Yang et al., 2011). In contrast, perturbing *CEBPA* decreased accessibility of TP63 and NF-κB features. An analysis of the perturbed factors and their target regions uncovered an inter-connected network of regulation (Figure 6E).

To assess how TF perturbation changed cellular differentiation trajectories, we derived a differentiation pseudotime from wild-type cells and projected perturbed cells onto this pseudotime axis (Figures 6F, S7B, and S7C). We also determined how feature accessibility changed for each of the three regulatory modules identified in unperturbed cells (Figure 6G). This analysis revealed a role for CEBPA in initiating differentiation because *CEBPA* knockout cells were biased toward an early differentiation state. Correspondingly, *CEBPA* knockout cells lost accessibility of module 2 features but gained module 1 accessibility, indicating that *CEBPA* knockout cells were unable to fully engage the mid-differentiation program. In contrast, *KLF4* knockout cells shifted toward a later differentiation state. Together, we identified both the regulatory factors responsible for distinct epidermal states as well as rerouted differentiation trajectories in response to TF perturbation.

Comprehensive Epistatic Mapping Reveals the Logic of Regulatory Synergy

We extended our previous findings on epistasis by analyzing dual knockout cells. Each perturbation-feature pair was scored as additive (no interaction), synergistic (positive interaction), or buffering (negative interaction) (Figures 7A, 7B, and S7D). This workflow identified 344 additive relationships, 102 synergistic interactions, and 101 buffering interactions. Unlike in other systems, we observed a weak correlation between the magnitude of single perturbation and the degree of genetic interaction for that feature (Costanzo et al., 2010; Figure S7E).

This investigation revealed surprising interactions in keratinocyte differentiation. For instance, *JUNB* knockout increased accessibility of CEBP motif sites, whereas dual *EHF* and *JUNB* knockout had a modest effect, reflecting a negative interaction between EHF and JUNB (Figure 7C). However, JUNB and EHF did not interact at their respective target sites, suggesting an indirect effect. When comparing all pairs of targeted TFs, we noticed disparities in the prevalence of each interaction category, with a more than 3-fold variation in the proportion of synergistic interactions (Figure 7D). EHF and JUNB had relatively low synergy, in contrast to the high synergy observed for EHF and

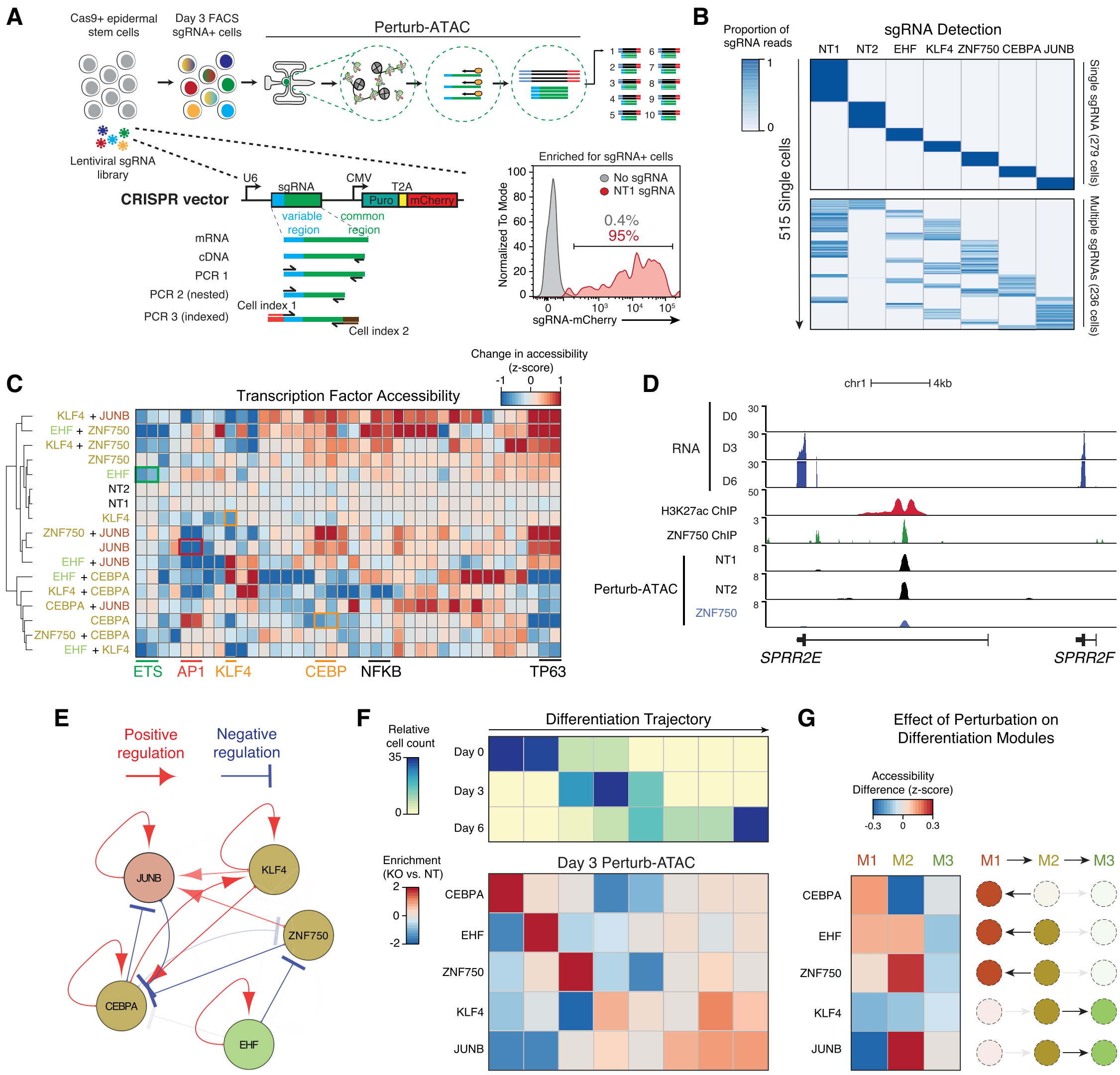


Figure 6. Multiplex Knockout Screen of Transcription Factors in Differentiation

(A) Schematic of sgRNA expression vector and library amplification for direct sequencing readout of sgRNA identity.

(B) Heatmap of sgRNAs (columns) versus single cells (rows), indicating the proportion of all reads associated with each sgRNA.

(C) Heatmap of genetic perturbations versus TF features, indicating activity of TF features in perturbed relative to NT cells. Similar motifs from AP-1, FOX, and ETS families were merged.

(D) Genomic locus of SPRR2E gene. Perturb-ATAC tracks show the signal from merged single cells receiving each sgRNA. Shown are H3K27ac and ZNF750 ChIP-seq tracks in day 3 differentiating keratinocytes (from Rubin et al., 2017 and Boxer et al., 2014).

(E) Representation of positive and negative regulation between targeted genes and sets of genomic regions. Arrows are shown with FDR < 0.25, and decreasing transparency is associated with a lower FDR.

(F) Top: heatmap displaying the frequency of cells in eight bins representing progression along the differentiation trajectory. Bottom: heatmap indicating the enrichment or depletion of cells in each differentiation bin compared with NT cells. For each perturbation, a custom-reduced dimensionality space was created to highlight altered features.

(G) Heatmap of perturbations (rows) versus modules (columns). For each module, the mean change in feature activity is shown.

ZNF750. This synergy was particularly apparent at motif sites for TP63, a master regulator of epidermal homeostasis and differentiation, with dual EHF and JUNB knockout cells increasing TP63 accessibility far more than would be predicted from the single knockouts (Bao et al., 2015; Figure 7C).

We reasoned that a mechanism to explain synergy between two factors' regulation of chromatin state could be genomic co-localization, analogous to the genetic interactions because of physical protein interactions observed in yeast (Costanzo et al., 2010). To test this hypothesis, we computed the degree of genomic overlap for the target regions of each pair of perturbed factors (Figure 7E). Indeed, the factor pairs with the highest prevalence of synergistic interactions exhibited greater overlap of target sites, consistent with a co-binding model

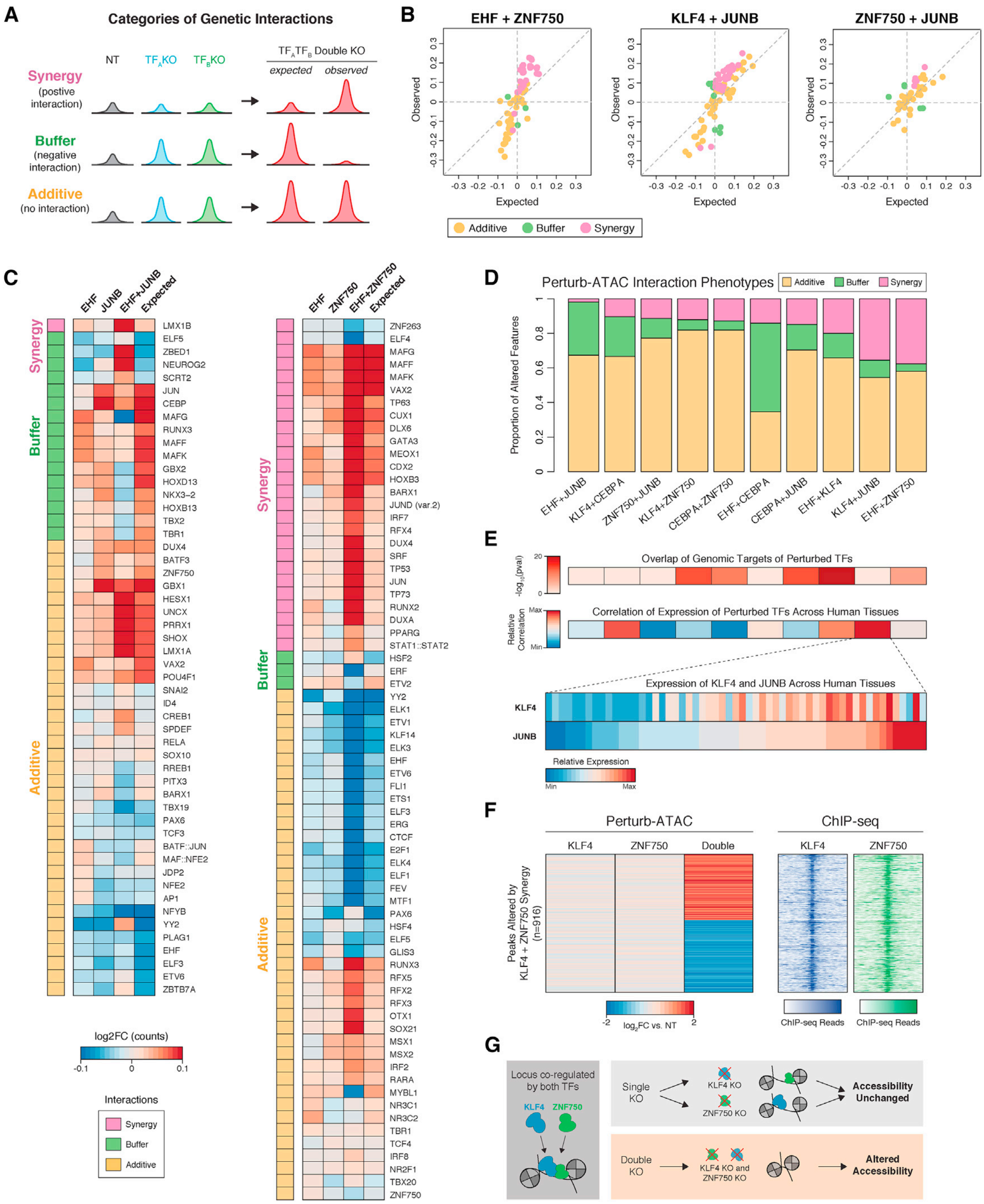


Figure 7. Pairs of Perturbations Exhibit Distinct Patterns of Epistatic Interactions

(A) Example representative peak signal for each category of interaction.

(B) Scatterplots of observed versus expected (based on additive model) accessibility in double knockout cells. Only features significantly altered under either single or double knockout conditions are plotted. Colors indicate the category of interaction.

(C) Left: heatmap of altered activity of features (rows) in EHF, JUNB, or simultaneous EHF and JUNB knockouts along with their expected activity. Right: similar to left, for EHF and ZNF750 knockouts.

(D) Proportion of interacting features in each category. Each column represents a pair of targeted genes. Only features altered in either the single or double perturbation are considered.

(legend continued on next page)

(Spearman's correlation = 0.4). Beyond this, we reasoned that coordinated patterns of gene expression across cell states could indicate that two factors act in similar pathways and, thus, may interact and calculated the correlation of expression between all pairs of TFs studied across a diverse set of tissues. Interestingly, correlated gene expression was associated with synergy (Spearman's correlation = 0.358), exemplified by KLF4 and JUNB, which exhibit the second-greatest degree of synergy (Figure 7E). This result may reflect the need for strict regulation of cooperative TF expression across tissues to achieve distinct cell states.

To further explore the relationship between genomic co-binding and synergy, we closely examined KLF4 and ZNF750, which exhibited a high degree of target region overlap. We asked whether the specific genomic regions at which synergy was observed were bound by both factors. Using ChIP-seq data (Boxer et al., 2014), we observed that these 916 regions were, in fact, commonly bound by both factors, providing strong support for the model that co-binding underlies epistatic interaction (Figures 7F and 7G).

DISCUSSION

Previous high-throughput genetic screens have largely relied on the introduction of genetic variants in a population of cells (for example, using CRISPR/Cas9), followed by selection of cells containing variants that confer a limited set of cellular phenotypes, such as increased cell growth or viability. However, identifying genetic variants that regulate genome-wide chromatin states, which are challenging to isolate with standard selection protocols, requires focused experiments, where candidate factors are perturbed one by one. Here we address this challenge by achieving simultaneous measurements of perturbations and ATAC-seq profiles in single cells. We applied this method to determine the roles of a diverse set of *trans*-regulatory factors, including TFs, chromatin modifiers, and human and viral ncRNAs. Analysis of Perturb-ATAC data enabled the study of several layers of chromatin regulation: individual *cis*-regulatory elements, inferred TF activity from *cis*-regulatory modules, and nucleosome positioning and occupancy. Therefore, Perturb-ATAC may be particularly well suited for examination of the precise molecular switches in the noncoding genome that underlie cell state compared with existing methods that pair whole-transcriptome profiles with perturbations in single cells.

We analyzed 63 genotypes using ~4,300 single cells (~70 cells/genotype) to over-shoot the number of cells required to reliably measure the effects of perturbations. We suggest that future screens target fewer cells per genotype (20) while expanding the number of genotypes assayed. Using our microfluidic platform, this strategy should enable screening of 150 genotypes in 3,000 cells. However, we foresee technological improvements that will enable higher throughput and more complex experimental designs. For example, this methodology should be adaptable to additional high-throughput methods for single-cell ATAC-seq, such as single-cell FACS (Chen et al., 2018b, 2018a), the split pool approach (Cusanovich et al., 2015, 2018), or nano-well or droplet-based methods (Mezger et al., 2018).

We designed the Perturb-ATAC platform to be compatible with widely used CRISPR constructs, and therefore, we hope that it will be easily adaptable to existing screens in diverse systems. The ability to detect targeted RNA sequences could be easily adapted to detect alternative CRISPR constructs and classes of targets, such as orthogonal CRISPR guide RNAs, which could simultaneously regulate target gene activation and repression in the same cell (Boettcher et al., 2018; Najm et al., 2018). In addition to perturbation of *trans*-regulatory factors, this system could also be used to target individual *cis*-regulatory elements; for example, to study the local effects of enhancer targeting on the expression of local genes. This approach may be useful to dissect loci where both *cis*-regulatory elements and ncRNA transcripts have been shown to have effects on gene expression (Engreitz et al., 2016; Cho et al., 2018). Perturb-ATAC provides a high-throughput platform to link genotypes with epigenetic phenotypes and ultimately reveal molecular mechanisms that govern cell fate and function through modulation of the chromatin state.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - GM12878 cell line culture
  - Human keratinocyte isolation and culture
- METHOD DETAILS
  - CRISPRi targeting in GM12878
  - Culture, differentiation, and CRISPR knockout in primary keratinocytes
  - Bulk ATAC-seq
  - Single-cell ATAC-seq
  - Perturb ATAC-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single cell and bulk ATAC primary processing and chromVAR analysis
  - Guide barcode sequencing analysis for GM12878 experiments
  - Direct sgRNA sequencing and analysis for keratinocyte experiments
  - Identification of differentially accessible genomic features and regions
  - Statistical analysis of SPI1 motif-containing region accessibility in SPI1-depleted cells

(E) Top: heatmaps indicating significance of genomic overlap (hypergeometric test) or correlation of gene expression for pairs of TFs corresponding to pairs displayed in (D). Bottom: heatmap displaying relative RNA expression of KLF4 and JUNB across tissues from the Roadmap Epigenomics Project.

(F) Left: heatmap indicating relative accessibility of genomic regions (rows) with synergistic behavior in KLF4 and ZNF750 double knockout cells. Right: heatmap of the ChIP-seq signal for KLF4 and ZNF750 at regions displayed on the left.

(G) Model of KLF4 and ZNF750 redundancy in maintaining accessibility at co-occupied loci.

- ○ Inferred nucleosome and sub-nucleosome profiles and score calculation
- ○ Analysis of inferred regulatory networks
- ○ Analysis of epistasis for accessibility of genomic features
- ○ Analysis of tissue H3K27me3 and autoimmune-associated SNPs
- ○ Pseudotime calculation and identification of feature modules
- ○ Altered differentiation trajectory and module activity analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.11.022.

ACKNOWLEDGMENTS

We thank the members of the Khavari, Chang, and Greenleaf laboratories for helpful discussions. This work was supported by the Veterans Affairs Office of Research and Development (to P.A.K.), the NIH (AR45192 and AR43799 to P.A.K., P50-HG007735 to H.Y.C. and W.J.G., and R35-CA209919 to H.Y.C.), the Parker Institute for Cancer Immunotherapy (to A.T.S. and H.Y.C.), and the Scleroderma Research Foundation (to H.Y.C.). A.J.R was supported by a Stanford Bio-X fellowship. K.R.P was supported by a Stanford graduate fellowship. A.T.S. was supported by a Parker Bridge Scholar Award from the Parker Institute for Cancer Immunotherapy and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. W.J.G is a Chan Zuckerberg Biohub investigator. H.Y.C. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

A.J.R., A.T.S., H.Y.C., and P.A.K. conceived the project. A.J.R., K.R.P., A.T.S., Y.Q., B.W., A.J.O., D.S.K., A.L.J., S.W.C., B.J.Z., and M.R.M. performed the experiments and analyzed data. P.A.K., H.Y.C., and W.J.G. guided the experiments and data analysis. A.J.R., K.R.P., A.T.S., H.Y.C., and P.A.K. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

H.Y.C. and W.J.G. are scientific co-founders of Epinomics. H.Y.C. is a co-founder of Accent Therapeutics and a consultant for 10X Genomics and Spring Discovery. Stanford has filed a provisional patent on ATAC-seq; H.Y.C. and W.J.G. are listed as inventors.

Received: June 27, 2018
Revised: September 12, 2018
Accepted: November 12, 2018
Published: December 20, 2018

REFERENCES

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21.

Adamson, B., Norman, T.M., Jost, M., and Weissman, J.S. (2018). Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. bioRxiv. https://doi.org/10.1101/298349.

Arrand, J.R., Young, L.S., and Tugwood, J.D. (1989). Two families of sequences in the small RNA-encoding region of Epstein-Barr virus (EBV) correlate with EBV types A and B. J. Virol. *63*, 983–986.

Bao, X., Rubin, A.J., Qu, K., Zhang, J., Giresi, P.G., Chang, H.Y., and Khavari, P.A. (2015). A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. Genome Biol. *16*, 284.

Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat. Biotechnol. *36*, 170–178.

Boxer, L.D., Barajas, B., Tao, S., Zhang, J., and Khavari, P.A. (2014). ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes. Genes Dev. *28*, 2013–2026.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535–1548.e16.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Chen, X., Litzenburger, U.M., Wei, Y., Schep, A.N., LaGory, E.L., Choudhry, H., Giaccia, A.J., Greenleaf, W.J., and Chang, H.Y. (2018a). Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. Nat. Commun. *9*, 4590.

Chen, X., Natarajan, K.N., and Teichmann, S.A. (2018b). A rapid and robust method for single cell chromatin accessibility profiling. bioRxiv. https://doi.org/10.1101/309831.

Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell *173*, 1398–1412.e22.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Cusanovich, D.A., Reddington, J.P., Garfield, D.A., Daza, R.M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H.A., Christiansen, L., Qiu, X., Steemers, F.J., et al. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature *555*, 538–542.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Eckert, R.L., Adhikary, G., Young, C.A., Jans, R., Crish, J.F., Xu, W., and Rorke, E.A. (2013). AP1 transcription factors in epidermal differentiation and skin cancer. J. Skin Cancer *2013*, 537028.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Feldman, D., Singh, A., Garrity, A.J., and Blainey, P.C. (2018). Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. bioRxiv. https://doi.org/10.1101/262121.

Flynn, R.A., Do, B.T., Rubin, A.J., Calo, E., Lee, B., Kuchelmeister, H., Rale, M., Chu, C., Kool, E.T., Wysocka, J., et al. (2016). 7SK-BAF axis controls pervasive transcription at enhancers. Nat. Struct. Mol. Biol. *23*, 231–238.

Gray, H. (1918). Anatomy of the Human Body (Lea and Febiger).

Heath, J.R., Ribas, A., and Mischel, P.S. (2016). Single-cell analysis tools for drug discovery and development. Nat. Rev. Drug Discov. *15*, 204–216.

Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J., Jackson, D., Shendure, J., and Trapnell, C. (2018). On the design of CRISPR-based single-cell molecular screens. Nat. Methods *15*, 271–274.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883–1896.e15.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Lee, N., Moss, W.N., Yario, T.A., and Steitz, J.A. (2015). EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell *160*, 607–618.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, P., Keller, J.R., Ortiz, M., Tessarollo, L., Rachel, R.A., Nakamura, T., Jenkins, N.A., and Copeland, N.G. (2003). Bcl11a is essential for normal lymphoid development. Nat. Immunol. *4*, 525–532.

Lopez, R.G., Garcia-Silva, S., Moore, S.J., Bereshchenko, O., Martinez-Cruz, A.B., Ermakova, O., Kurz, E., Paramio, J.M., and Nerlov, C. (2009). C/EBPalpha and beta couple interfollicular keratinocyte proliferation arrest to commitment and terminal differentiation. Nat. Cell Biol. *11*, 1181–1190.

Lopez-Pajares, V., Qu, K., Zhang, J., Webster, D.E., Barajas, B.C., Siprashvili, Z., Zarnegar, B.J., Boxer, L.D., Rios, E.J., Tao, S., et al. (2015). A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation. Dev. Cell *32*, 693–706.

Lunning, M.A., and Green, M.R. (2015). Mutation of chromatin modifiers; an emerging hallmark of germinal center B-cell lymphomas. Blood Cancer J. *5*, e361.

Ma, S., Pathak, S., Trinh, L., and Lu, R. (2008). Interferon regulatory factors 4 and 8 induce the expression of Ikaros and Aiolos to down-regulate pre-B-cell receptor and promote cell-cycle withdrawal in pre-B-cell development. Blood *111*, 1396–1403.

McKercher, S.R., Torbett, B.E., Anderson, K.L., Henkel, G.W., Vestal, D.J., Baribault, H., Klemsz, M., Feeney, A.J., Wu, G.E., Paige, C.J., and Maki, R.A. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J. *15*, 5647–5658.

Mezger, A., Klemm, S., Mann, I., Brower, K., Mir, A., Bostick, M., Farmer, A., Fordyce, P., Linnarsson, S., and Greenleaf, W. (2018). High-throughput chromatin accessibility profiling at single-cell resolution. Nat. Commun. *9*, 3647.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.

Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N., et al. (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat. Biotechnol. *36*, 179–189.

Nutt, S.L., and Kee, B.L. (2007). The transcriptional regulation of B cell lineage commitment. Immunity *26*, 715–725.

Pang, S.H.M., Minnich, M., Gangatirkar, P., Zheng, Z., Ebert, A., Song, G., Dickins, R.A., Corcoran, L.M., Mullighan, C.G., Busslinger, M., et al. (2016). PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. Leukemia *30*, 1375–1387.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Wiessman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rubin, A.J., Barajas, B.C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M.R., Howard, I., Kim, D.S., Boxer, L.D., Cairns, J., Spivakov, M., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat. Genet. *49*, 1522–1528.

Samanta, M., Iwakiri, D., Kanda, T., Imaizumi, T., and Takada, K. (2006). EB virus-encoded RNAs are recognized by RIG-I and activate signaling to induce type I IFN. EMBO J. *25*, 4207–4214.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495–502.

Satpathy, A.T., Saligrama, N., Buenrostro, J.D., Wei, Y., Wu, B., Rubin, A.J., Granja, J.M., Lareau, C.A., Li, R., Qi, Y., et al. (2018). Transcript-indexed ATAC-seq for precision immune profiling. Nat. Med. *24*, 580–590.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. *25*, 1757–1770.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573–1577.

Scott, E.W., Fisher, R.C., Olson, M.C., Kehrli, E.W., Simon, M.C., and Singh, H. (1997). PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. Immunity *6*, 437–447.

Sen, G.L., Boxer, L.D., Webster, D.E., Bussat, R.T., Qu, K., Zarnegar, B.J., Johnston, D., Siprashvili, Z., and Khavari, P.A. (2012). ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev. Cell *22*, 669–677.

Stehling-Sun, S., Dade, J., Nutt, S.L., DeKoter, R.P., and Camargo, F.D. (2009). Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor Mef2c. Nat. Immunol. *10*, 289–296.

Su, G.H., Ip, H.S., Cobb, B.S., Lu, M.M., Chen, H.M., and Simon, M.C. (1996). The Ets protein Spi-B is expressed exclusively in B cells and T cells during development. J. Exp. Med. *184*, 203–214.

Tamura, T., Yanai, H., Savitsky, D., and Taniguchi, T. (2008). The IRF family transcription factors in immunity and oncogenesis. Annu. Rev. Immunol. *26*, 535–584.

van Riel, B., and Rosenbauer, F. (2014). Epigenetic control of hematopoiesis: the PU.1 chromatin connection. Biol. Chem. *395*, 1265–1274.

Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

Xie, S., Cooley, A., Armendariz, D., Zhou, P., and Hon, G.C. (2018). Frequent sgRNA-barcode recombination in single-cell perturbation assays. PLoS ONE *13*, e0198635.

Yang, X., Lu, H., Yan, B., Romano, R.-A., Bian, Y., Friedman, J., Duggal, P., Allen, C., Chuang, R., Ehsanian, R., et al. (2011). ΔNp63 versatilely regulates a Broad NF-κB gene program and promotes squamous epithelial proliferation, migration, and inflammation. Cancer Res. *71*, 3688–3700.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhao, B., Barrera, L.A., Ersing, I., Willox, B., Schmidt, S.C.S., Greenfeld, H., Zhou, H., Mollo, S.B., Shi, T.T., Takasaki, K., et al. (2014). The NF-κB genomic landscape in lymphoblastoid B cells. Cell Rep. *8*, 1595–1606.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
One Shot Stbl3 Chemically Competent *E. coli*	Thermo Fisher	Cat#C737303
Biological Samples		
Human: Keratinocytes	Stanford Hospital	N/A
Chemicals, Peptides, and Recombinant Proteins		
BstXI	NEB	Cat#R0113L
BlpI	NEB	Cat#R0585L
NEBuilder Hifi DNA Assembly Master Mix	NEB	Cat#E2621L
Sensiscript Reverse Transcriptase	QIAGEN	Cat#205211
Q5 Polymerase	NEB	Cat#M0491S
NEBnext High-Fidelity PCR Master Mix	NEB	Cat#M0541L
Kapa Library Quantification Kit	Kapa Biosystems	Cat#KK4854
Critical Commercial Assays		
Nextera DNA Library Preparation Kit	Illumina	FC-121-1030
C1 Single-Cell Auto Prep IFC for Open App	Fluidigm	100-8133
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE116297
DHS-seq peak calls	ENCODE Project	http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeUwDnase
ENCODE accessible region blacklist	ENCODE Project	http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeMapability/
GM12878 ChIP-seq narrowPeaks	ENCODE Project	https://www.encodeproject.org/search/?type=Experiment&assay_title=ChIP-seq&biosample_term_name=GM12878&assembly=GRCh38&files.file_type=bed+narrowPeak
H3K27me3 ChIP-seq	Roadmap Epigenomics Project	https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/
H3K27ac ChIP-seq	Roadmap Epigenomics Project	https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/
GENCODE gene definitions	GENCODE Project	https://www.gencodegenes.org/; RRID:SCR_014966
GWAS autoimmune SNPs	Farh et al., 2015	Table S1
Keratinocyte ATAC-seq peak calls	ENCODE Project	https://www.encodeproject.org/treatment-time-series/ENCSR968JDE/
GM12878 H3K27ac HiChIP interactions	Mumbach et al., 2017	GEO: GSE101498
Experimental Models: Cell Lines		
Human: GM12878 dCas9-KRAB-BFP-2A-Blast	Mumbach et al., 2017	N/A
Human: HEK293T Lenti-X	Clontech	Cat#632180
Oligonucleotides		
Primers for sgRNA sequencing in CRISPRi system	This paper	Table S1
Primers for sgRNA sequencing in CRISPR KO system	This paper	Table S4
Primers for qPCR check of CRISPRi knockdown	This paper	Table S1
Primers for Sanger sequencing check of CRISPR KO	This paper	Table S4
Recombinant DNA		
pMJ114 (bovine U6 for CRISPRi sgRNA cloning)	Addgene	Cat#85995
pMJ117 (human U6 for CRISPRi sgRNA cloning)	Addgene	Cat#85997

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
pMJ179 (mouse U6 for CRISPRi sgRNA cloning)	Addgene	Cat#85996
lentiCRISPRv2 (Source for CRISPR KO Cas9 and sgRNA expression cassette)	Sanjana et al., 2014	Addgene: 52961
U6 sgRNA (F+E)	Chen et al., 2013	Addgene: 59986
pLentiguide	This paper	Addgene: 117986
pLex_Cas9	This paper	Addgene 117987
Software and Algorithms		
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml; RRID:SCR_016368
Picard	Broad Institute	RRID:SCR_006525
Samtools	Li et al., 2009	http://www.htslib.org/; RRID:SCR_002105
chromVAR	Schep et al., 2017	https://bioconductor.org/packages/release/bioc/html/chromVAR.html
Monocle 2	Qiu et al., 2017	http://cole-trapnell-lab.github.io/monocle-release/
Seurat	Satija et al., 2015	https://satijalab.org/seurat/; RRID:SCR_016341
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS; RRID:SCR_013291
BEDTools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/; RRID:SCR_006646
Levenshtein	https://pypi.org/project/python-Levenshtein/	https://pypi.org/project/python-Levenshtein/
NumPy	http://www.numpy.org/	RRID:SCR_008633
SciPy	https://www.scipy.org/	RRID:SCR_008058
MatPlotLib	http://matplotlib.org/	RRID:SCR_008624
scikit-learn	https://scikit-learn.org/	RRID:SCR_002577
Fluidigm Script for Open App IFC: Biomodal Single-Cell Omics	Fluidigm Script Hub	https://www.fluidigm.com/c1openapp/scripthub/script/2018-06/c1-t-atac-sequencing20-1529617256973-1

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Paul A. Khavari (khavari@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

GM12878 cell line culture

GM12878 (female) were maintained in RPMI 1640 (Thermo Fisher) with 10% FBS and 1% Penicillin-Streptomycin (Thermo Fisher) at 37°C with 5% CO_2. Cells stably expressing dCas9-KRAB-BFP were generated previously (Mumbach et al., 2017) and originally purchased from the Coriell Institute for Medical Research. Briefly, GM12878 cells were transduced with lentivirus containing a dCas9-BFP-KRAB-2A-Blast construct and subsequently selected for blasticidin resistance. Cells were maintained between 200,000 to 1 million cells/mL during routine culture.

Human keratinocyte isolation and culture

Primary human keratinocytes were isolated from fresh, surgically discarded neonatal foreskin. All human cells were collected and analyzed by protocols approved by the Stanford Human Subjects Institutional Review Board and in accordance with the NIH genomic data sharing policy Keratinocytes were maintained in a 1:1 mixture of Keratinocyte-SFM (Thermo Fisher) and Medium 154 (Thermo Fisher). Keratinocyte differentiation was induced by the addition of 1.2 mM calcium for 3 or 6 days at full confluence.

METHOD DETAILS

CRISPRi targeting in GM12878

To generate the Perturb-ATAC vector with guide barcodes used in the GM12878 experiments, we modified previously-described CRISPRi vectors (Adamson et al., 2016; Cho et al., 2018). Briefly, we designed three sgRNAs per target gene, each targeting a different region between the transcriptional start site and 200 nucleotides into the gene body. Guides were designed using the CRISPOR online tool (http://crispor.tefor.net/) and chosen to minimize the number of genomic loci with 0-4 mismatches, with the intention of minimizing the chance for any off-target CRISPRi activity. For the prediction of repressive activity at mismatch loci, empirical observations of how sgRNA mismatches at various positions in the sgRNA affect the activity of the dCas9-KRAB construct were used to calculate a predicted fraction of repressive activity relative to the on-target locus (Qi et al., 2013). As observed in Qi et al. (2013), a multiplicative model was used, where the total repressive activity of sgRNAs at loci with multiple mismatches is equal to the product of the predicted loss of activity from each mismatch (i.e., multiple sgRNA mismatches inhibit repressive activity to a greater extent than the sum of the degree to which each individual mismatch reduces activity).

One sgRNA each was cloned into pMJ114 (bovine U6, Addgene, Cat#85995), pMJ117 (human U6, Addgene, Cat#85997) or pMJ179 (mouse U6, Addgene, Cat#85996), digested with BstXI and BlpI, using NEBuilder Hifi DNA Assembly Master Mix. Then the respective U6 promoter and sgRNA sequences were amplified by PCR and assembled into the lentiviral vector (digested using XbaI and XhoI) using NEBuilder Hifi DNA Assembly Master Mix. Subsequently, individual colonies for each 3x sgRNA plasmid were digested using PciI and EcoRI, and a randomized 22 bp barcode (ordered from IDT as 5′-[overhang][NNN…][overhang]-3′) was assembled with NEBuilder Hifi DNA Assembly Master Mix. The sgRNA sequences and GBC sequences of all plasmids were confirmed by Sanger sequencing.

To generate CRISPRi virus, we used HEK293T cells (Clontech Lenti-X) maintained in DMEM (Thermo Fisher) with 10% FBS, 1% Pen-Strep. Cells were seeded at 4 million per 10cm dish, and the following day transfected with 4.5ug pMP.G, 1.5ug psPAX2, and 6ug sgRNA vector using OptiMEM and Lipofectamine 3000. Two days later, the supernatant was collected and filtered with a 0.44 μm filter, and virus was concentrated 1:10 using Lenti-X Concentrator (Clontech).

GM12878 maintained in RPMI 1640 (Thermo Fisher) with 10% FBS and 1% Penicillin-Streptomycin (Thermo Fisher) were then seeded at 300,000 cells per well of a 6-well plate and 40ul of concentrated virus was added to the media the following day. Two days later, we exchanged the media for media containing 1ug/ml puromycin to select for the sgRNA vector. Selection media was refreshed on day five, and on day seven cells selection media was exchanged for regular media (containing no puromycin) and cells were either assayed or frozen in viable conditions with BamBanker cryopreservation media. Cells were sorted by flow cytometry for viability and expression of mCherry before being assayed by Perturb ATAC-seq. Cells were maintained between 200,000 and 1 million per mL. RNA was extracted with Trizol and purified using QIAGEN RNeasy columns, and gene expression knockdown was confirmed using the Agilent Brilliant II qRT-PCR 1-Step kit. qRT-PCR was performed in duplicate, and expression values for each sample were normalized against 18S. Gene expression values for CRISPRi are reported as average fold change against both non-targeting control samples.

Culture, differentiation, and CRISPR knockout in primary keratinocytes

We generated custom Cas9 and sgRNA expression vectors for CRISPR knockout in keratinocytes. For Cas9 expression, we amplified the Cas9 gene from the lentiCRISPRv2 vector (Sanjana et al., 2014) and cloned this fragment into pLex-MCS (Thermo Fisher) along with a fusion P2A-blasticidin resistance cassette in exchange for the IRES-puromycin resistance cassette in pLex-MCS. For sgRNA expression, we modified the sgRNA F+E scaffold (Chen et al., 2013; https://www.addgene.org/59986/) in two ways. First, we exchanged the murine U6 promoter and telomerase-targeting sgRNA with the human U6 promoter, stuffer region, and associated BsmBI cloning sites from lentiCRISPRv2. Additionally, we removed a BsmBI restriction site in the puromycin resistance gene by introducing a non-synonymous mutation. These vectors are available from Addgene as pLex_Cas9 (Plasmid #117987) and pLentiGuide (Plasmid #117986).

To generate lentivirus, we seeded 400,000 HEK293T cells into a single well of a 6-well dish, and the following day we transfected either our Cas9 vector or sgRNA vector (1.3 ug) along with pMDG (0.3 ug) and p8.91 (1 ug) using Lipofectamine 3000 (Thermo Fisher). Supernatant was collected at 48hrs and 72 hr, filtered through a 0.45um PES membrane, and concentrated to a pellet with Lenti-X Concentrator. One unit of Cas9 virus corresponded to the concentrated supernatant from one 6-well of HEK293T. One unit of sgRNA virus corresponded to one eighth of the concentrated supernatant from one 6-well of HEK293T.

Primary keratinocytes were seeded at 300,000 cells per well of a 6-well dish along with one unit of Cas9 virus and polybrene (0.1 ug/ml). After one day, two wells were harvested, mixed, and expanded into a 15cm dish containing normal culture media with 2ug/ml blasticidin. After four to six days of selection, cells were again seeded at 300,000 cells per well of a 6-well dish along with one unit of sgRNA virus and polybrene (0.1 ug/ml). After one day, one well was harvested and transferred to a 15cm dish containing normal culture media, puromycin (1 ug/ml) and blasticidin (2 ug/ml). After six days of selection, cells were seeded at high confluence with 1.2 mM calcium for differentiation. Cells were harvested after three days of differentiation and viably frozen in culture media with 10% DMSO.

Cas9 nuclease activity was assessed by PCR amplifying ~800bp fragments of cDNA surrounding sgRNA binding sites and analyzing the resulting fragments by Sanger sequencing (oligo sequences in Table S4). Images depicted in Figure S6 were generated using Geneious 7.1.4. cDNA was generated by extracting RNA from cells with the RNeasy Mini Kit (QIAGEN) and performing reverse transcription with the iScript cDNA Synthesis Kit (Bio-Rad).

Bulk ATAC-seq

Cells were isolated and subjected to ATAC-seq as previously described (Corces et al., 2017). Briefly, 50,000 cells were pelleted after sorting and resuspended in 50ul of ATAC resuspension buffer (RSB) with 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin. After three minutes, 1ml of ATAC RSB with 0.1% Tween-20 was added, tubes were inverted, and nuclei were centrifuged at 500 *rcf.* for 10 min. Supernatant was carefully removed and nuclei were resuspended in 50ul transposition mix (25ul TD buffer, 2.5ul transposase, 16.5ul PBS, 0.5ul 0.1% digitonin, 0.5ul 10% Tween-20, and 5ul water). Transposition was performed for 30 minutes at 37 C with shaking in a thermomixer at 1000 RPM. Reactions were purified with a Zymo DNA Clean & Concentrator 5 kit and library generation was performed as described previously (Corces et al., 2017).

Single-cell ATAC-seq

Single-cell ATAC-seq was performed as previously described (Buenrostro et al., 2015). In brief, cells were sorted by flow cytometry for viability and to remove cell aggregates. The C1 Single-Cell Auto Prep System was used with the Open App program (Fluidigm). The Open App scripts from the "ATAC Seq" collection from Fluidigm were used to prime the C1 IFC microfluidic chip, load cells, and run the ATAC sample prep protocol. Fluidigm scripts are available from Fluidigm Script Hub, https://www.fluidigm.com/c1openapp/scripthub.

Perturb ATAC-seq

Cell isolation and microfluidic reactions on the IFC

We adapted the C1 Single-Cell Auto Prep System with its Open App program (Fluidigm) to perform Perturb-ATAC-seq. C1 IFC microfluidic chips were first primed by following the Open App script "Biomodal Single-Cell Genomics: Prime." Single cells were then captured using the Fluidigm Open App script "Biomodal Single-Cell Genomics: Cell Load." GM12878 or keratinocyte cells were first isolated by FACS sorting and then washed three times in C1 DNA Seq Cell Wash Buffer (Fluidigm). Cells were resuspended in DNA Seq Cell Wash Buffer at a concentration of 300 cells/μL and mixed with C1 Cell Suspension Reagent at a ratio of 3:2 (cells:reagent). 15 μl of this cell mix was loaded onto the IFC. After cell loading, all wells were visualized by imaging on a Leica CTR 6000 microscope to identify captured cells.

Cells were then subjected sequentially to lysis and transposition, transposase release, quenching with $MgCl_2$, reverse transcription, and PCR, using the custom Open App IFC script "Biomodal Single-Cell Omics: Sample Prep." For lysis and transposition, 30 μL of Tn5 transposition mix was prepared (22.5 μL 2x TD buffer, 2.25 μL transposase (Nextera DNA Sample Prep Kit, Illumina), 2.25 μL C1 Loading Reagent without salt (Fluidigm), 0.45 μL 10% NP40, 2.25 μL SuperaseIN RNase inhibitor, and 0.3 μL water). For transposase release, 20 μL of Tn5 release buffer mix was prepared (2 μL 500 mM EDTA, 1 μL C1 Loading Reagent without salt, and 17 μL 10 mM Tris-HCl Buffer, pH 8). For $MgCl_2$ quenching, 20 μL of $MgCl_2$ quenching buffer mix was prepared (18 μL 50 mM MgCl2, 1 μL C1 Loading Reagent without salt, and 1 μL 10 mM Tris-HCl Buffer, pH 8). For reverse transcription, 30 μL of RT mix was prepared (15.55 μL H_20, 3.7 μL 10x Sensiscript RT buffer (QIAGEN), 3.7 μL 5 mM dNTPs, 1.5 μL C1 Loading Reagent without salt (Fluidigm), 1.85 μL Sensiscript RT (QIAGEN), and 3.7 μL 6 μM RT primer mix (6uM each of V1 GBC sequencing oligos or 6uM each of V1 sgRNA sequencing oligos, see Tables S1 and S4 for oligo sequences). Finally, for ATAC and GBC/sgRNA PCR, 30uL of PCR mix was prepared (8.62 μL H_20, 13.4 μL 5x Q5 polymerase buffer (NEB), 1.2 μL 5 mM dNTPs, 1.5 μL C1 Loading Reagent without salt, 0.67 μL Q5 polymerase (2U/μL; NEB), 0.8 μL 25 μM non-indexed custom Nextera ATAC-seq PCR primer 1, 0.8 μL 25 μM non-indexed custom Nextera ATAC-seq primer 2, and 3 μL 6 μM GBC or sgRNA primer mix.

7 μL lysis and transposition mix, 7 μL transposase release buffer, 7 μL $MgCl_2$ quenching buffer, 24 μL RT mix, and 24 μL PCR mix were added to the IFC inlets. On the IFC, Tn5 lysis and transposition reaction was carried out for 30 minutes at 37°. Next, transposase release was carried out for 30 min at 50°C. $MgCl_2$ quenching buffer was immediately added and chamber contents were immediately incubated with RT mix for 30 minutes at 50°C. Finally, gap filling and 8 cycles of PCR were performed using the following conditions: 72°C for 5 min and then thermocycling at 94°C for 30 s, 62°C for 60 s, and 72°C for 60 s. The amplified transposed DNA was harvested in a total of 13.5 μL C1 Harvest Reagent. Following completion of the on-chip protocol (~4-5hrs), chamber contents were transferred to 96-well PCR plates, mixed, and divided for further amplification of ATAC-seq fragments (6-7 μl) or GBC/sgRNA fragments (6.5 μl).

For method development and RT primer troubleshooting, the Perturb-ATAC-seq protocol can be exactly scaled 1000x and performed on 1000 cells in Eppendorf tubes. Following lysis, transposition, and transposase release, RNA can be reverse-transcribed and subjected to PCR amplification to check the amplification efficiency and specificity of a chosen primer set.

Amplification of ATAC-seq libraries

~7 μL of harvested libraries were amplified in 50 μL PCR for an additional 15 cycles with 1.25 μM Nextera dual-index PCR primers in 1x NEBnext High-Fidelity PCR Master Mix using the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and thermocycling at 98°C for 10 s, 72°C for 30 s, and 72°C for 1 min. The PCR products were pooled and purified on a single MinElute PCR purification column (QIAGEN). Libraries were quantified using qPCR (Kapa Library Quantification Kit for Illumina, Roche) prior to sequencing using 2x76bp paired-end reads on an Illumina NextSeq 550 or 2x75bp reads on an Illumina MiSeq.

Amplification of guide barcode and guide RNA sequencing libraries

Three rounds of off-C1 PCR were performed to generate GBC and sgRNA sequencing libraries (See Tables S1 and S3 for V1,V2, and V3 oligo sequences). First (V1 PCR), 6.5ul of harvested libraries were amplified in a 20 ul PCR (harvested DNA with 10ul NEBNext Master Mix, 0.1 ul of each V1 primer at 200uM, and remaining volume of water). Reactions amplified for 17 cycles with the following parameters: 98 C for 30 s, then cycling of 98 C for 10 s, 63 C for 30 s, and 72 C for 45 s, followed by 72 C for 5 min. Second, 2ul of the V1 PCR product (without purification) was transferred to a subsequent 20ul reaction with 10ul NEBNext Master Mix, 0.1 ul of each V2 primer at 200uM, and remaining volume of water. Reactions were amplified for 15 cycles using the same parameters used for V1 reactions. A final 20ul V3 cell indexing PCR was performed using 2ul of the V2 reaction product, 2ul each of Illumina Indexing primers at 10 uM, 10ul NEBNext Master Mix, and the remaining volume of water. Reactions were amplified for 15 cycles using the same parameters used for V1 and V2 reactions.

Finally, V3 reactions were pooled and purified using the QIAGEN MinElute kit. Libraries were further purified by size selection on polyacrylamide gel electrophoresis (6% TBE Novex gel, Thermo Fisher). Libraries were mixed with BlueJuice loading dye (Thermo Fisher), run for 35 min at 160 V and visualized using SybrSafe stain (Thermo Fisher), using 5ul of stain in 30ml of TBE running buffer for 10 min. Gels were visualized on a blue-light transilluminator and slices in size range for GBC library fragments (289 bp) or sgRNA library fragments (232 bp) were cut using a scalpel. Gel slices were placed in a 0.75ml tube with a hole punctured in the bottom using a syringe, and this tube was placed in a 1.5ml DNA LoBind tube (Eppendorf). These tubes were centrifuged for 3 min at 13k RPM to crush the gel slice, then 300ul Salt Crush Buffer (500 mM NaCl, 1 mM EDTA, 0.05% SDS) was added and this mix was incubated at 55 C overnight in a thermomixer with 1000 RPM shaking. The next day, samples were cooled to RT, centrifuged through a Spin-X column (one minute, 13k RPM), and purified with a Zymo DNA Clean & Concentrator 5 kit. Libraries were quantified by qPCR (Kapa Library Quantification Kit for Illumina, Roche) before sequencing on an Illumina MiSeq at 10-14 pM final concentration with 15%–40% PhiX.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single cell and bulk ATAC primary processing and chromVAR analysis

Single cell and bulk ATAC read alignment, quality filtering, and duplicate removal were performed as previously described (Buenrostro et al., 2015). Briefly, adaptor sequences were trimmed, sequences were mapped to the hg19 reference genome using Bowtie2 (Langmead and Salzberg, 2012; and the parameter -X2000), and PCR duplicates were removed using Picard Tools. Reads mapping to the mitochondria were discarded for further analysis. We observed an extremely low rate of ATAC reads matching the CRISPR viral construct (median 0.0049%) and found no evidence of the abundance of CRISPR construct matching reads influencing epigenomic profiles.

Single cell ATAC-seq calculation of TF deviation was performed using chromVAR (in R, version 1.1.1; Schep et al., 2017). Briefly, for each TF, ‘raw accessibility deviations’ were computed by subtracting the expected number of ATAC-seq reads in peaks for a given motif from the observed number of ATAC-seq reads in peaks for each single cell. Expected reads were calculated from the population average of all cells for the GM12878 experiment and unperturbed cells only for the keratinocyte experiment. This value is subtracted by the mean deviation calculated for sets of ATAC-seq peaks with similar accessibility and GC content to obtain a bias-corrected deviation value, and additionally divided by standard deviation of the deviation calculated for the background sets to obtain a Z-score.

For the GM12878 experiments, we used a set of peaks derived from DNase I hypersensitivity data (downloaded from http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeUwDnase) from a broad variety of hemopoietic cell lines (all GM lines, HL-60, Th1, Jurkat, K562) plus additional lines (HepG2, HUVEC, NHEK), to account for the possibility of opening peaks outside the blood lineage. These peaks were each filtered against the wgEncodeDacMapabilityConsensusExcludable.bed blacklist, sorted by intensity, and the top 75,000 peaks for each sample were merged. These peaks were then centered and resized to 1kb uniform peaks (238,349 final peaks).

For the keratinocyte experiment, we merged peaks called on bulk ATAC-seq from undifferentiated cells and cells differentiated for three or six days (processed bulk ATAC-seq reads available from the ENCODE project portal: https://www.encodeproject.org/treatment-time-series/ENCSR968JDE/). Peaks were called using the MACS2 command *macs2 callpeak–nomodel –nolambda –-call-summits–shift −75–extsize 150* (Zhang et al., 2008). First, peaks with q-value < 0.01 from each day were merged. In the case of overlapping peaks, the summit associated with the lowest q-value was selected as the merged peak summit, and the 1kb window centered on that summit was used as the uniform peak for chromVAR (94,633 final peaks).

For GM1878 analysis, narrowPeak ChIP-seq files (optimal IDR thresholded peaks) were downloaded from ENCODE and imported as supplementary annotations in chromVAR. Prior to use, these files were filtered against the wgEncodeDacMapabilityConsensusExcludable.bed blacklist. H3K27me3 and H3K27ac narrowPeak files for different tissues were downloaded from the Roadmap Epigenomics website (http://www.roadmapepigenomics.org/data/).

Guide barcode sequencing analysis for GM12878 experiments

For GM12878 experiments, raw reads for GBC libraries were matched to a list of GBC sequences to generate a table of counts for each cell and each GBC analyzed in the experiment (see Figure S1, custom scripts written in Python available upon request). First, any read not containing the expected 27 nt sequence prior to the GBC was discarded, allowing for a maximum Levenshtein distance of 2 to account for sequencing errors. The subsequent 22 nt sequence was then compared to a list of GBC sequences, allowing for a maximum Levenshtein distance of 3 to be considered a match. Note that the minimum Levenshtein distance between any two of our GBC sequences was 10. This generated a counts-per-cell table for each GBC sequence and cell.

This table was normalized for read depth by plate by assessing the maximum density of log-transformed counts using the scipy.stats.gaussian_kde function (see Figure S1C). This distribution exhibits a bimodal distribution corresponding to wells with productive and unproductive GBC detection. A normalized GBC read cutoff of 1000 reads/cell was set (Figure 2A, this was empirically determined based off the separation between wells with and without a cell capture). Cells displaying high background reads, as determined by having greater than 0.005 proportion reads not aligning to the top two GBC sequences, were further filtered (this cutoff was set from empirical observations of "background" in doublet wells, which are expected to contain up to four GBC sequences; Figure S1C). We distinguished cells expressing a single or double sgRNAs based off the percent of reads aligning to the second-most common GBC (single, < 1% double, > 5%). This workflow resulted in far more double-targeted cells than would be observed solely from the observed doublet rate calculated from the appearance of double GBC-expressing cells in our initial single-targeting experiment (~2.9%). t-SNE plots shown in Figure S2 were generated using the manifold. t-SNE function in the Python package scikit-learn.

We empirically determined a target minimum cell number required for analysis by down-sampling cells from a larger pool and comparing accessibility profiles. This analysis indicated that the vast majority of samples of five cells were highly correlated ($r > 0.8$) with a bulk ATAC-seq profile. Additionally, previous reports have shown that aggregation of five or more cells is sufficient to accurately reproduce chromatin accessibility profiles (Satpathy et al., 2018; Schep et al., 2017). In line with these findings, we designed Perturb-ATAC experiments to yield the maximal number of genotypes supported by at least five cells; indeed 38/40 genotypes for GM12878 cells and 23/23 genotypes for keratinocytes consist of greater than five cells.

Direct sgRNA sequencing and analysis for keratinocyte experiments

For keratinocyte experiments, raw reads for sgRNA sequencing were matched to a list of sgRNA sequences used in the experiment. We required strict matching of the 20bp variable sequence along with 18bp of the standard sgRNA backbone. Matching was performed with custom scripts (available upon request) and resulted in the counts-per-cell table for each sgRNA.

We then normalized this table for read depth by assessing the plate-specific distribution of log-transformed total counts per cell (Figure S6). The collection of counts per cell exhibited a bimodal distribution likely corresponding to productive and failed sgRNA detection. We drew a cutoff in between the two modes as a first filter, and further required cells to exhibit low background (reads associated with the third most common sgRNA in each cell). Cells with greater than 1% of reads associated with background were excluded from analysis. Finally, we distinguished cells expressing one or two sgRNAs based on the distribution of proportions of reads associated with the second most common sgRNA in each cell. Cells with fewer than 1% of reads associated with the second most common sgRNA formed a clear mode in this distribution and were considered to express only the most common sgRNA, while cells with greater than 10% of reads associated with the second most common sgRNA were considered to express both the first and second most common sgRNAs.

Identification of differentially accessible genomic features and regions

We generated an empirical null distribution of accessibility values for each feature in order to assess the significance of any observed difference between mean accessibility in a set of perturbed cells compared to cells expressing non-targeting control sgRNAs. For each genomic feature (peak or chromVAR motif/annotation), we first calculated the median deviation z-score (for chromVAR features) or fragment counts (for peaks) in cells expressing each sgRNA or combination of sgRNAs. Cells expressing a targeting sgRNA in combination with a non-targeting sgRNA were analyzed with targeting sgRNA-only cells. With the goal of assessing the null hypothesis that targeting and non-targeting cells exhibit the same accessibility, we pooled equal numbers of cells from targeting and non-targeting cells. This population was then randomly divided into two sets by permuting the cell-genotype labels, and the permuted median accessibility difference of these two populations were compared to the observed median accessibility difference. This process was repeated 5000 times to generate a null distribution, and the rate of detecting a median accessibility difference as extreme or greater in the null distribution compared to the observed targeting cells was reported as the false discovery rate (FDR). The network representation of altered features in Figure 6E was generated using Cytoscape v3.1.0.

Differentially accessible regions were found using a similar approach with the exception that we limited the set of total regions under consideration to those exhibiting at least one read per five cells in one of the conditions under consideration for each comparison. Genome browser tracks of differentially accessible regions were generated by pooling cells associated with a particular sgRNA genotype. We first generated bedGraph files scaled to 500,000 reads using the genomeCoverageBed tool (BedTools v2.17.0) then generated bigWig files using the bedGraphToBigWig tool from UCSC (http://hgdownload.soe.ucsc.edu/admin/exe/). Tracks were finally displayed in the WashU Epigenome Browser.

Statistical analysis of SPI1 motif-containing region accessibility in SPI1-depleted cells

For Figure 1G, we determined an empirical false discovery rate for the observed changes in SPI1 motif region accessibility. For bulk-ATAC and Perturb-ATAC samples separately, we calculated the z-score of the SPI1 motif accessibility change in perturbed cells compared to all other features. Then to generate a null distribution, we permuted the sample labels between Non-targeting #1, Non-targeting #2, and SPI1-targeting 1000 times and in each trial recorded the z-score of SPI1 motif change in accessibility compared to the non-targeting controls. In this analysis, for both bulk-ATAC and Perturb-ATAC, no trial yielded a result as extreme as the result observed in the unpermuted sample.

Inferred nucleosome and sub-nucleosome profiles and score calculation

The aggregate profiles of nucleosomal signals at differentially accessible regions were derived from total ATAC fragments as described previously (Bao et al., 2015). Briefly, ATAC fragments sized 180-247bp were considered nucleosome-spanning and used to infer positions of nucleosomes in aggregate locus profiles (metaplots). Differentially accessible regions were centered based on the signal summit as identified by Macs2 (using the flags –-call-summits–shift −75–extsize 150) and filtered for an FDR < 0.1 and log2 fold change > 1. We then calculated the fragment count in 10bp windows spanning 1000 bp upstream and downstream of the region summit. These profiles were normalized to the average signal in the 25 downstream windows to account for sequencing depth and the resulting enrichment values were smoothed in R using the smooth.spline() function with parameter spar = 0.5.

To quantify the presence of peak central versus flanking nucleosome in each metaplot, we calculated the ratio of flanking nucleosome signal density (−180 to −80bp relative to peak summit and +80 to +180bp relative to peak summit) to central nucleosome signal density (−20 to +20bp relative to peak summit). We report this ratio as the central nucleosome score.

Analysis of inferred regulatory networks

To identify sets of genomic features whose activities were correlated across single cells, suggestive of shared regulatory relationships, we computed the Pearson correlation of each feature with each other feature across all single cells of a given genotype. Only features that were significantly altered in at least one genotype were considered, and redundant annotations were removed, resulting in 390 motif/ChIP feature annotations for analysis. Ward's hierarchical clustering was performed and features displaying low intra-cluster correlation were excluded from further analysis (Figure S4A). The modules shown in subsequent analysis were defined based off Ward's hierarchical clustering of the remaining features in non-targeting cells. Clustering was performed using the Seaborn clustermap function using Ward's method for clustering.

For each Perturb-ATAC genotype, the feature-feature correlation across single cells was computed. The difference in correlation between a given genotype and non-targeting cells was computed by subtracting the Pearson correlation in the respective genotype from non-targeting cells. A permutation test was used to assess the significance of the observed change in correlation for any pair of features. For each genotype, the same number of cells was randomly sampled from all perturbed cells 10,000 times, and the changes in correlation in the randomly sampled cells relative to non-targeting cells were used to create a null distribution for each feature-feature pair (in each genotype). A 5% cutoff was used to call significantly altered correlations. To quantify module-level changes in regulatory relationships, we quantified the percent of all feature-feature pairs in a given module whose correlations were significantly altered.

Analysis of epistasis for accessibility of genomic features

We assessed the degree of epistasis in double perturbation conditions by comparing observed phenotypes in double perturbation conditions to phenotypes expected based on a model of non-interaction. For this analysis, we scored the accessibility of genomic features based on the sum of raw reads accumulating in peaks associated with that feature in each cell. Feature counts were normalized by the total number of reads for features in each cell and log2-transformed with the addition of a pseudocount. For each collection of cells sharing a genotype, the mean value of log2 counts was compared to the mean value of log2 counts for a mix of cells expressing non-targeting sgRNAs, resulting in a log2 (fold change of perturbation versus non-targeting). The additive expectation was based on a multiplicative model of non-interaction, (i.e., CRISPR AB = CRISPR A x CRISPR B), which we calculated by adding the single perturbation fold changes in log2-space. For each genomic feature, the degree of interaction (difference between observed accessibility change and that expected under the non-interaction model) was calculated.

To identify generally additive versus non-additive features (Figures 4D and 4E), the interaction degree was averaged across perturbations. To compute the permuted background, we permuted the single-double pairings by randomly choosing a double sgRNA genotype and two random single sgRNA genotypes. The difference between the "expected" change (based on the two random sgRNA genotypes) and the "observed" changed (based on the random double sgRNA genotype) was then computed. This process was repeated once for each double sgRNA genotype observed in our dataset.

We further categorized features as additive, synergizing, and buffering for a particular interaction (Figure 7) by comparing the observed degree of interaction to a null distribution generated by permuting cell identities. This procedure was performed separately for each feature to account for differences in scale and variability across features. The null distribution was generated by randomly sampling three pools of cells from all perturbed cells: a null double perturbation set, and two null single perturbation sets. The difference between observed double perturbation phenotype and the expected value from the non-interaction model was calculated, and this procedure was repeated 1000 times. Genotypes exhibiting interaction degrees beyond 95% of the null values were considered interacting. Interactions in which the double phenotype had a more extreme magnitude than expected were labeled synergistic, while others were labeled buffering.

Analysis of tissue H3K27me3 and autoimmune-associated SNPs

128 consolidated narrowPeak files for H3K27me3 peaks (corresponding to different tissues/cell-types) were downloaded from the Roadmap Epigenomics Consortium website. Peaks that were found across at least 30 samples were considered common H3K27me3 peaks. Individual narrowPeak files were then filtered against this set of common H3K27me3 peaks, as well as the wgEncodeDacMapabilityConsensusExcludable blacklist. The resulting files were subsequently centered and resized to create uniform 1kb peaks, and imported into chromVAR as an annotation set. To identify peaks repressed in the GM12878 lineage but active in other tissues, H3K27ac narrowPeaks from blood tissues present in the Roadmap Epigenomics Consortium dataset were downloaded and intersected with the GM12878 H3K27me3 narrowPeak set using the bedtools intersect command. These

were similarly filtered aginst the same blacklist, centered, and resized to create uniform 1kb peaks, and imported as a chromVAR annotation set.

SNPs associated with autoimmune diseases were downloaded from (Farh et al., 2015). These were aggregated by each autoimmune disease, and intersected with FitHiC calls (processed using 10kb genomic windows) from GM12878 H3K27ac HiChIP data (Mumbach et al., 2017). For each disease, the SNP (ultimately resized to a 10kb genomic window), as well as any windows in contact with that SNP, were aggregated to create a disease-specific chromVAR annotation set. As it is difficult to determine *a priori* whether a disease state would result from increased or decreased accessibility at a given site, we reported the absolute value change chromVAR deviation z-score for each genotype.

Pseudotime calculation and identification of feature modules

For the keratinocyte experiment, the normal differentiation pseudotime trajectory was calculated using Monocle 2 (Qiu et al., 2017). The feature deviation matrix including unperturbed and CRISPR knockout cells was first processed using Seruat 2.0.1 (Butler et al., 2018) to regress out plate and experiment batch effects. The Seurat function ScaleData was used (with parameters do.scale = F and do.center = F) to perform batch regression. To identify modules of dynamic features across differentiation, we first filtered for features that exhibited standard deviation greater than 1.3 in any comparison of normal differentiation conditions (Day 0, 3, or 6). Similar features associated with the AP-1 motif were merged into a single feature. The matrix of these features versus cells (arranged by increasing pseudotime) was hierarchically clustered using the heatmap.2 function in the gplots R package, resulting in three major clusters (referred to as modules).

Individual peaks approximately matching the kinetics of modules were identified in order to find associated genes (Figure 6C). Peaks exhibiting a log2 fold change less than 0.5 between conditions were considered stable and a fold change greater than 2 was considered dynamic. Peaks exhibiting decreased accessibility on both Day 3 and Day 6 (relative to Day 0) were considered Module 1 peaks. Peaks exhibiting increased accessibility on Day 3 versus Day 0 but stable accessibility between Day 6 and Day 0 were considered Module 2 peaks. Peaks exhibiting stable accessibility between Day 3 and Day 0 but gained accessibility on Day 6 versus Day 0 were considered Module 3 peaks. Genes (GENCODE definition) were considered potential regulatory targets of a peak if the gene transcription start site fell within 50kb of the peak.

Altered differentiation trajectory and module activity analyses

For each single perturbation in the keratinocyte experiment, a custom pseudotime was calculated in order assess the enrichment or depletion of cell occupancy along the differentiation trajectory (Figure 6F). ChromVAR deviations regressed for experimental batch effects and merged AP-1 features were used for this analysis. Cells from each perturbation were pooled with non-targeting cells and a custom principal component analysis (PCA) space was generated. Features altered in each perturbation (FDR < 0.1, change in z-score > 0.25) were selected in order to achieve maximum separation of control and perturbed cells, and a PCA was generated with the R prcomp function (center = T, scale = T). Next, non-perturbed cells from all stages of differentiation were analyzed and a trajectory was calculated progressing from undifferentiated cells (Day 0) to mid-differentiation (Day 3) and finally late-differentiation (Day 6). The trajectory was determined by plotting a linear path between centroids of the three cell populations representing each stage of differentiation. Finally, the distribution of non-targeting cells and targeted cells was calculated along eight equally sized bins in this trajectory, and the log2 fold change of the proportion of cells in each been was reported as an enrichment.

DATA AND SOFTWARE AVAILABILITY

The sequencing data (FASTQ files) and processed data files have been deposited in GEO under accession code GEO: GSE116297.

Cell

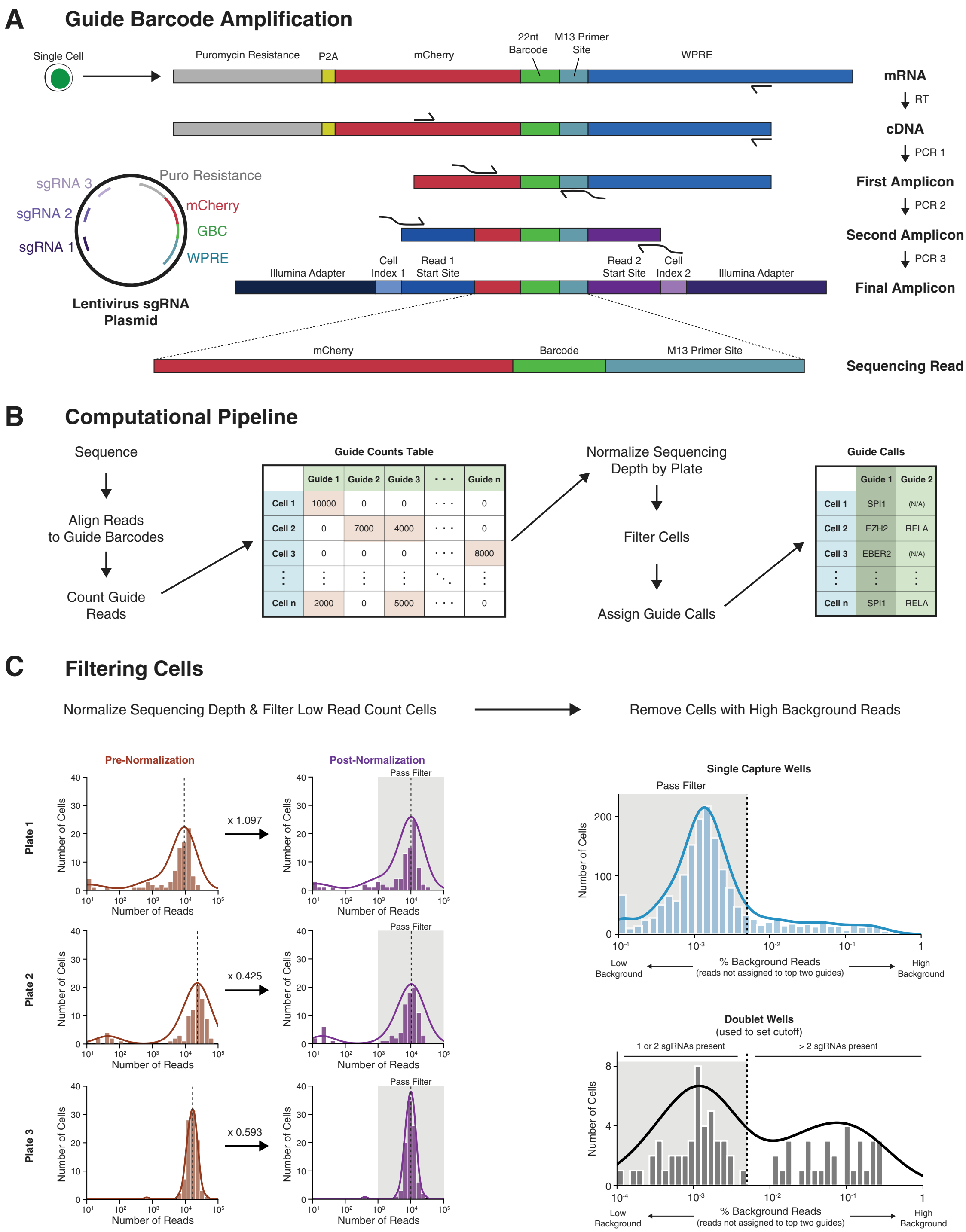


Figure S1. Perturb-ATAC CRISPRi Construct and Guide Bar Code Detection Scheme, Related to Figure 1

(a) Schematic of lentiviral plasmid encoding sgRNAs for CRISPRi as well as selection marker containing guide barcode. Stepwise targeted reverse transcription and PCR steps are displayed from top to bottom. (b) Overview of computational pipeline taking sequencing reads for GBC and producing final table of guide calls for each cell. (c) Detail on how filtering parameters for per-cell sequencing depth and background reads were derived. Left: distribution of reads aligning to any guide barcode are displayed for each of three representative plates. Middle: distribution of reads after plate-specific depth adjustment for sequencing depth, resulting in uniform median depth across plates and a uniform filter threshold of 1,000 normalized reads per cell. Right: Distribution of reads per cell not assigned to two most abundant guides, for cells annotated as single cell or doublet capture. Doublet wells separate into two modes, allowing determination of threshold separating unexpected high background in single capture wells.

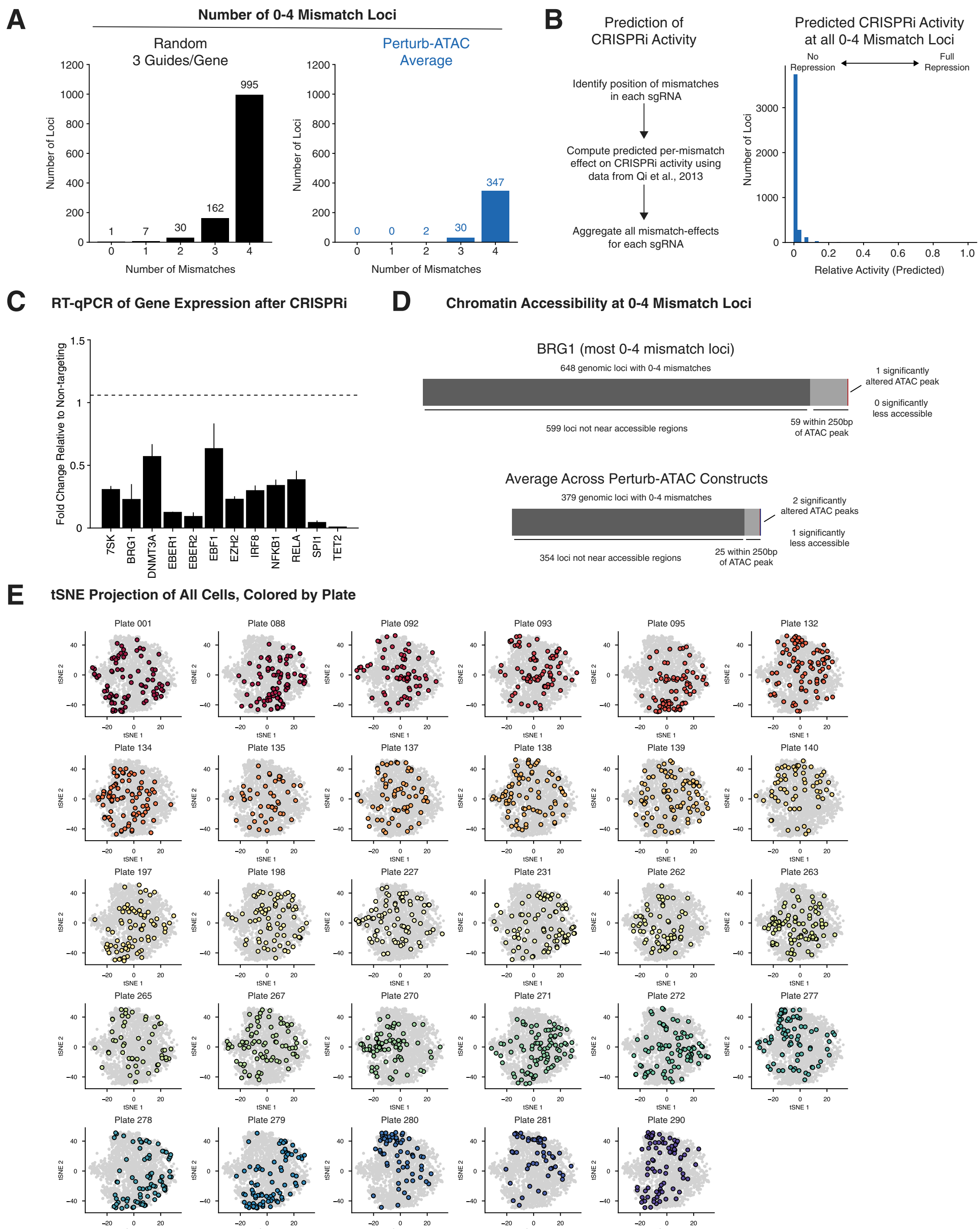


Figure S2. Analysis of CRISPR sgRNA Performance and Consistency of Perturb-ATAC Data across Separate C1 Chips, Related to Figure 2

(a) Bar plots indicating the count of sgRNA sequence mismatch for random guides or guides selected for Perturb-ATAC. (b) Left: description of workflow to calculate predicted off-target CRISPRi activity based on contribution of mismatches. Right: histogram of predicted relative off-target activity for all sgRNAs used in this study, including up to 4 mismatches. (c) qPCR validation of CRISPRi gene expression knockdown after transduction with sgRNAs targeting the specified genes. Error bars indicate one SD. (d) Bar plots indicating categories of sgRNA mismatch loci based on ATAC peak proximity and observed accessibility compared to non-targeting cells. (e) t-SNE plots of all cells assayed in GM12878 experiment based on chromVAR feature deviation z-scores. For each plot, the cells assayed on a particular plate are highlighted.

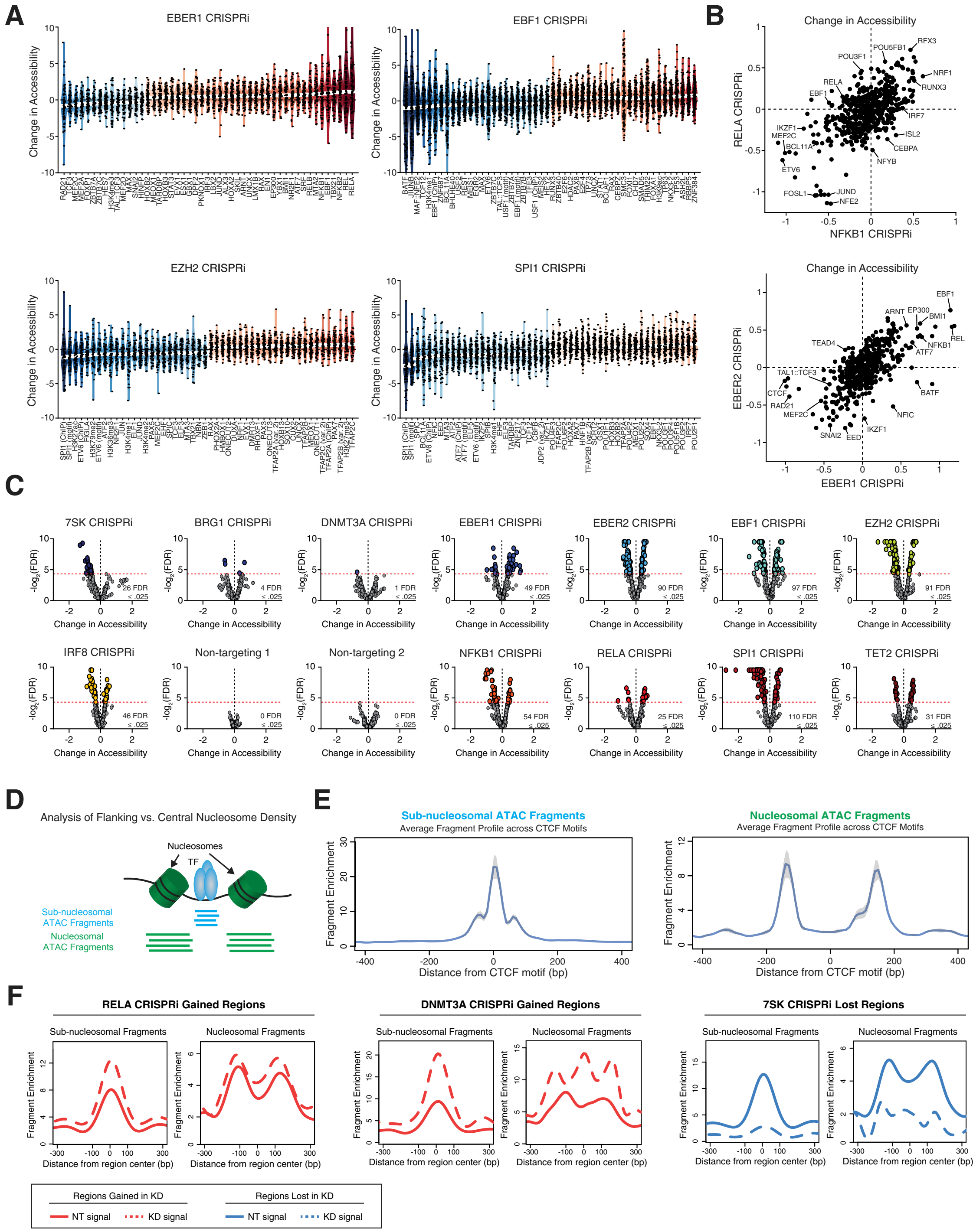
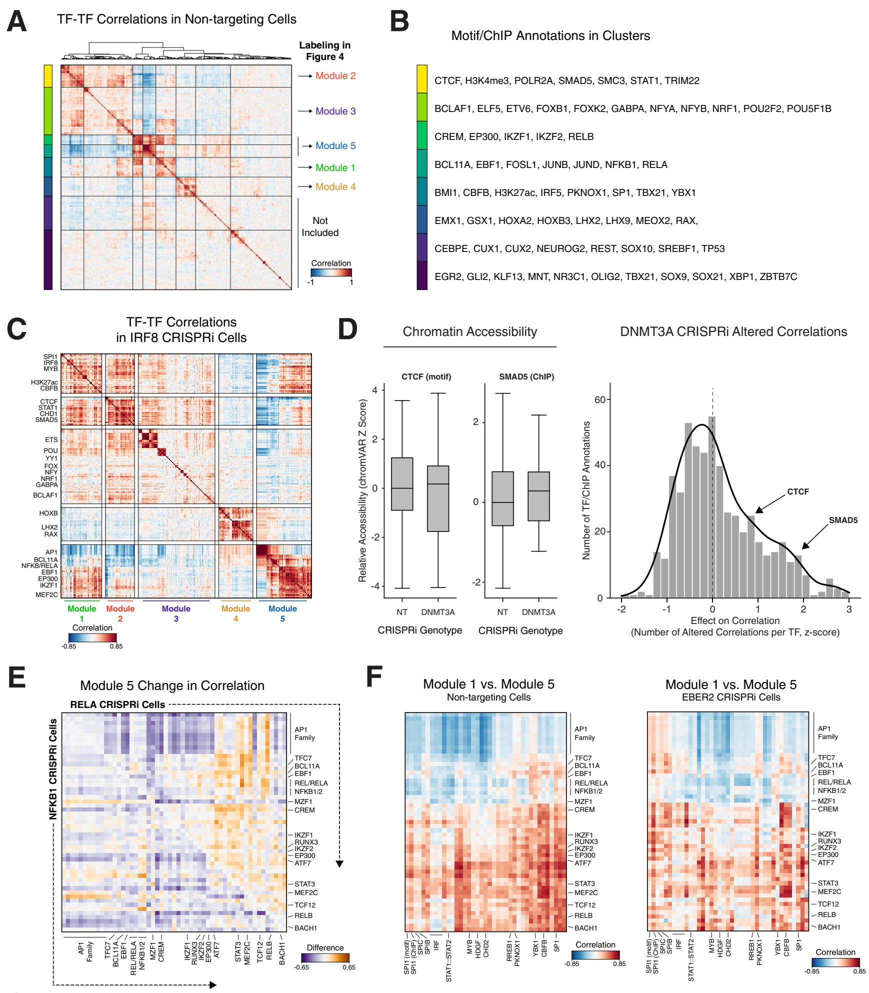


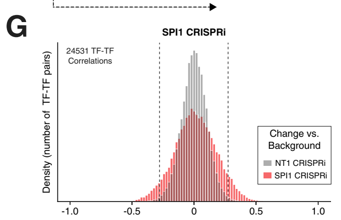
Figure S3. Identification of Differentially Accessible Genomic Features and Inferred Nucleosome Profiles in the GM12878 Screen, Related to Figure 2

(a) Violin plots of single cell accessibility relative to mean accessibility in non-targeting cells for significantly altered features in either EBER1, EBF1, EZH2, or SPI1 targeted cells. Each point represents an individual genomic feature (collection of genomic regions sharing an annotation such as a TF motif or ChIP-seq peak) in an individual cell. A maximum of 50 features are shown per genotype. (b) Scatterplots of accessibility in knockdown conditions, NFKB1 versus RELA (top) or EBER1 versus EBER1 (bottom). (c) Volcano plots for each single perturbation condition comparing perturbed cells to non-targeting control cells. Each point represents a genomic feature; significance threshold of FDR $\leq$ 0.025. (d) Schematic depicting generation of short (< 100bp) ATAC

(legend continued on next page)

fragments from sub-nucleosome regions and large fragments (180-247bp) spanning nucleosome-protected regions. (e) Metaplots of sub-nucleosome and nucleosome fragment signal at CTCF motif regions overlapping with CTCF ChIP-seq peaks in GM12878. Signal represents average of two non-targeting cell populations, gray range represents standard deviation between samples. (f) Metaplots of sub-nucleosome and nucleosome signal at differentially accessible regions.





H

Gene	Number Past 5% FDR	Most Frequently Affected TF Annotations
7SK	1253	RELA, NFKB2, TFAP2A (var. 3), VSX2, NFKB1
BRG1	1214	PAX5, E2F2, NR2C1, FOXA1, SOX10
DNMT3A	1277	LARP7, FOXC2, NFIC, EMX1, LBX1
EBER1	1006	MZF1 (var. 2), NXI1, ZNF384, MEF2A, ZBTB40
EBER2	1265	ELF4, RAD51, ELF5, ZIC3, SPIB
EBF1	1283	CHD2, PLAG1, RCOR1, RELB, TFAP2A
EZH2	791	PRDM1, EWSR1-FLI1, NOTO, POU3F3, E2F6
IRF8	2117	MEF2, EP300, SP8, FOS::JUN, JUN (var. 2)
NFKB1	1149	FOXD1, FOXO3, FOXI1, SMAD2::SMAD3::SMAD4, ZIC4
RELA	1234	ZEB1, ETV5, ELK1, IRF1, E2F4
SPI1	1025	TCF7, LHX6, LHX2, POU1F1, EMX2
TET2	1281	BMI1, IKZF2, DPF2, MLLT1, EED

Figure S4. Expanded Analysis of Perturbed Intercellular Genomic Feature Correlation Networks, Related to Figure 3

(a) Heatmap of correlation matrices for genomic features. Values indicate Pearson correlation across non-targeting cells for accessibility of two genomic features. Ward's hierarchical clustering was used to identify five modules with substantial intra-cluster correlation. (b) Listing of key features in each module. (c) Heatmap of correlation matrix for genomic features in IRF8 knockdown cells. (d) Left: boxplots of single cell accessibility for CTCF and SMAD5 features in non-targeting and DNMT3A knockdown cells. Right: histogram of z-score of number of altered correlations for each feature in DNMT3A knockdown cells. (e) Heatmap of difference in feature correlations between NFKB1 knockdown cells (bottom) and RELA knockdown cells (top). (f) Heatmaps of feature correlations for Module 1 versus Module 5 in non-targeting cells or EBER2 knockdown cells. (g) Histogram of change in feature correlations for SPI1 knockdown versus non-targeting 1 cells, used to inform thresholds for designation of altered correlation. (h) Table of counts and highlighted top altered-correlation features based on 5% FDR threshold.

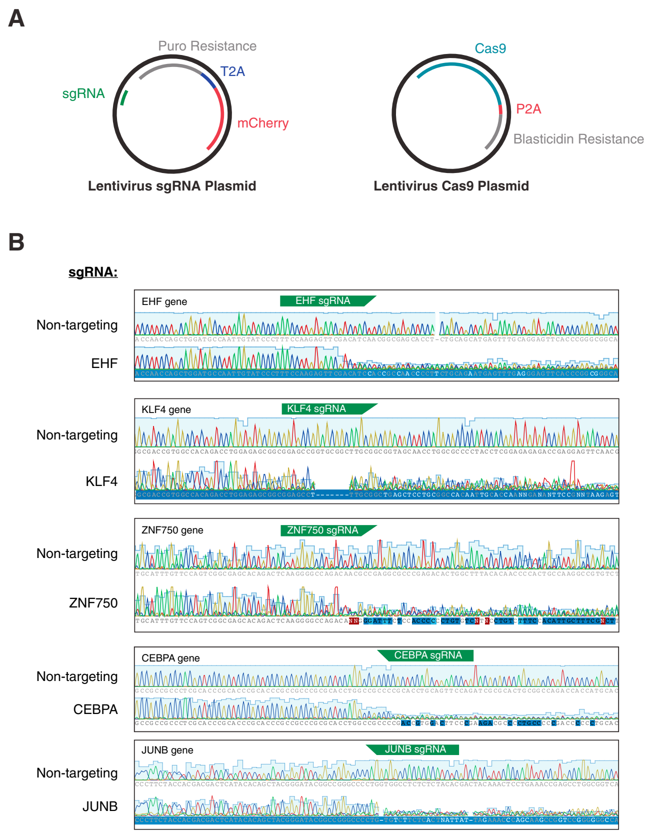


Figure S5. Perturb-ATAC CRISPR Knockout Constructs and Activity, Related to Figure 6

(a) Schematic of lentiviral plasmids for sgRNA and Cas9 expression. (b) Sanger sequencing traces of the 100bp surrounding sgRNA 3′ end for each target gene. Sequencing proceeded in forward direction (left to right), resulting in abrupt drop in sequencing alignment after sgRNA due to mixture of indels.

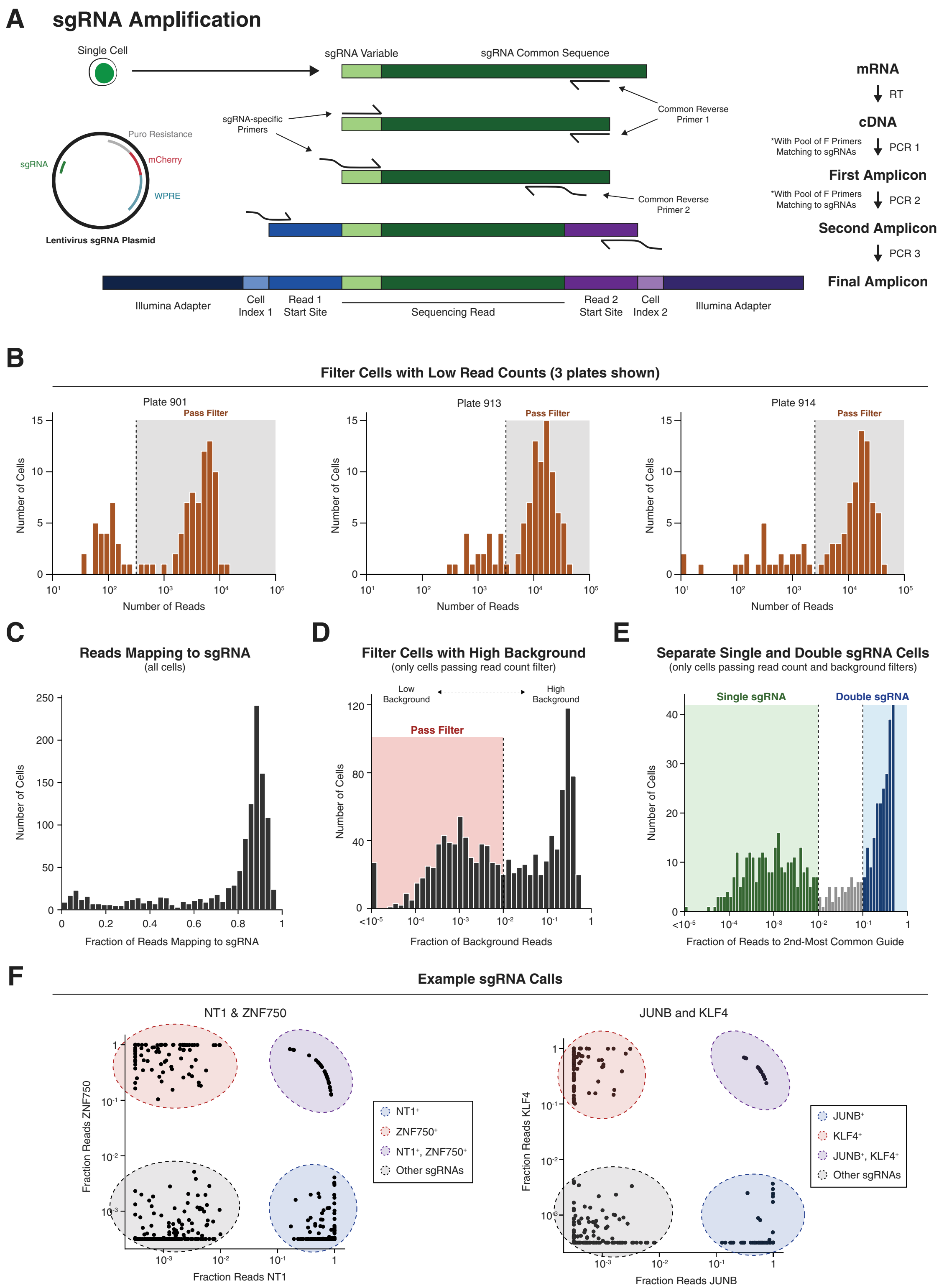


Figure S6. Perturb-ATAC CRISPR Knockout Direct Guide Detection Scheme, Related to Figure 6

(a) Schematic of lentiviral plasmid encoding sgRNA for CRISPR knockout. Stepwise targeted reverse transcription and PCR steps are displayed from top to bottom. (b) Distributions of reads per cell mapping to a sgRNA variable sequence. For each plate, a clear high mode of reads was identified and used to determine a depth cutoff. (c) Distribution of proportion of all reads per cell mapping to known sgRNA sequence. (d) Distribution of proportion of reads per

(legend continued on next page)

cell associated with background (third most common) guide sequence. Cells in low mode passed filter. (e) For cells passing previous filters, distribution of proportion of reads associated with second most common guide. Cells in the low mode of this distribution were considered to express a single guide, while cells in the high mode were considered to express two guides. (f) Scatterplots of proportion of reads associated with two guide sequences for all cells passing final filters.

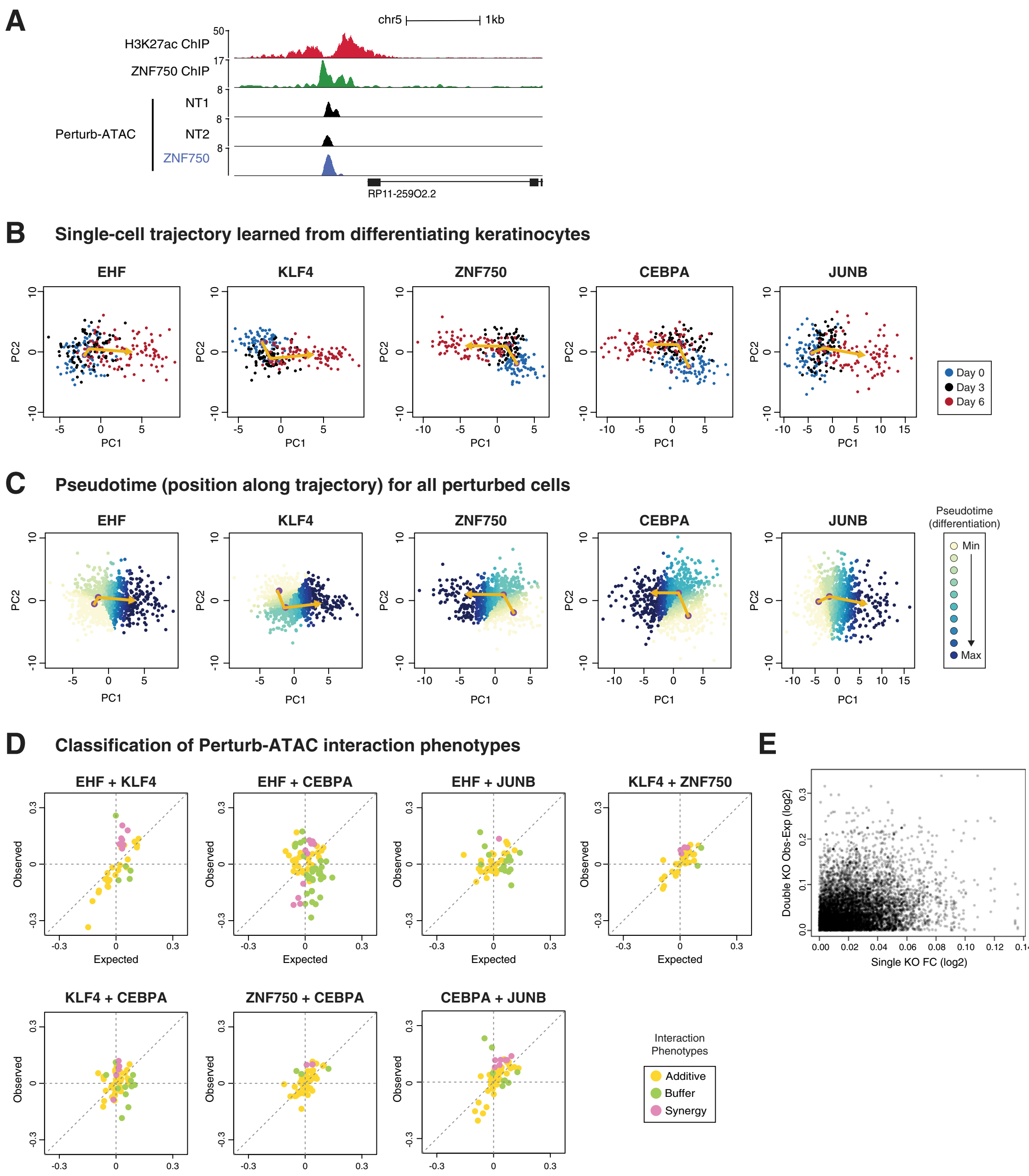


Figure S7. Altered Features in Keratinocyte Differentiation Induced by Genetic Perturbations, Related to Figures 6 and 7

(a) Signal track indicating a ZNF750 binding site that gains accessibility in targeted cells, indicating repressive activity of ZNF750. (b) Scatterplot of principal component (PC) values for unperturbed keratinocytes. PC space was generated using altered features from specific single TF knockout cells. Yellow line represents pseudotime trajectory connecting centroids of cells from each differentiation day. (c) Scatterplot of PC values for all perturbed and non-targeting cells embedded in PC space generated in (a). Cells are scored and colored by progression along pseudotime trajectory. These pseudotime values were used to assess the enrichment or depletion of knockout versus non-targeting cells in Figure 7F. (d) As in Figure 7B, scatterplots of observed versus expected (based on additive model) accessibility in double knockout cells. (e) Scatterplot of absolute log2 fold changes of features in single knockout cells versus double knockouts (r ∼0.18).