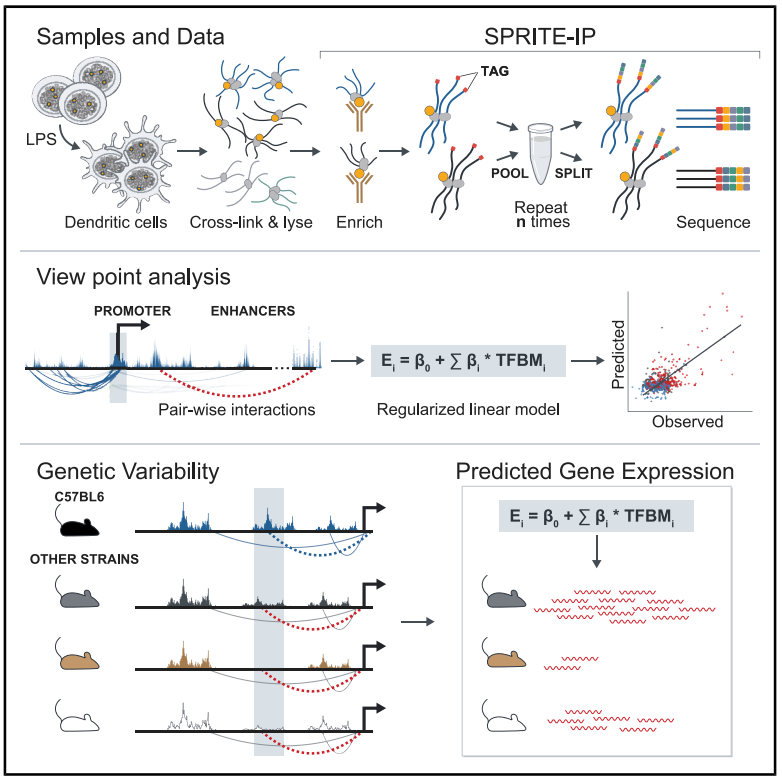
Molecular Cell

Resource

High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection

Graphical Abstract



Highlights

- A regression model predicts gene expression from DNA and chromosome conformation
- The effects of population variability in enhancer activity are predicted by the model
- Genes with heterogeneous regulatory interactions have heterogeneous expression
- AP1 transcription factors mediate LPS-inducible enhancer-promoter interactions

Authors

Pranitha Vangala, Rachel Murphy, Sofia A. Quinodoz, Kyle Gellatly, Patrick McDonel, Mitchell Guttman, Manuel Garber

Correspondence

manuel.garber@umassmed.edu

In Brief

Vangala et al. built a framework to quantify enhancer function by using SPRITE to define complex regulatory interactions at a single-molecule resolution. SPRITE-defined interactions vastly improve model performance. They then use the model to assess the effect of variability in enhancer activity across mouse strains on gene expression.

Vangala et al., 2020, Molecular Cell *80*, 359–373
October 15, 2020
https://doi.org/10.1016/j.molcel.2020.09.005

CellPress

Molecular Cell

CellPress

Resource

High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection

Pranitha Vangala,[1] Rachel Murphy,[1] Sofia A. Quinodoz,[2] Kyle Gellatly,[1] Patrick McDonel,[1,5] Mitchell Guttman,[2] and Manuel Garber[1,3,4,6,*]

[1]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA
[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[3]Department of Dermatology, Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA
[4]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA
[5]Present address: Dyno Therapeutics, Cambridge, MA, USA
[6]Lead Contact
*Correspondence: manuel.garber@umassmed.edu
https://doi.org/10.1016/j.molcel.2020.09.005

SUMMARY

Eukaryotic gene expression regulation involves thousands of distal regulatory elements. Understanding the quantitative contribution of individual enhancers to gene expression is critical for assessing the role of disease-associated genetic risk variants. Yet, we lack the ability to accurately link genes with their distal regulatory elements. To address this, we used 3D enhancer-promoter (E-P) associations identified using split-pool recognition of interactions by tag extension (SPRITE) to build a predictive model of gene expression. Our model dramatically outperforms models using genomic proximity and can be used to determine the quantitative impact of enhancer loss on gene expression in different genetic backgrounds. We show that genes that form stable E-P hubs have less cell-to-cell variability in gene expression. Finally, we identified transcription factors that regulate stimulation-dependent E-P interactions. Together, our results provide a framework for understanding quantitative contributions of E-P interactions and associated genetic variants to gene expression.

INTRODUCTION

Gene expression regulation involves a combination of promoters and distal regulatory elements called enhancers (Dekker et al., 2013; Schoenfelder and Fraser, 2019; Furlong and Levine, 2018; Snetkova and Skok, 2018). Enhancers are thought to establish cell-type-specific gene expression programs during development and in response to environmental cues. While there are on average six enhancer elements per promoter (Donnard et al., 2018; González et al., 2015), it is still not understood how these enhancers coordinate gene expression.

Since the discovery that the chromatin marks H3K27 acetylation (H3K27ac) and H3K4 mono-methylation (H3K4me1) are enriched in enhancer regions (Tie et al., 2009; Creyghton et al., 2010; Calo and Wysocka, 2013), tremendous effort has been made to use these histone marks to annotate enhancers and associate them to their target genes. These efforts resulted in a comprehensive catalog of putative enhancers elements across many cell types (Koch et al., 2007; Roadmap Epigenomics Consortium et al., 2015). However, we have not solved the fundamental question of how and to what extent each enhancer contributes to the regulation of gene expression. For example, it is not currently possible to predict the spatiotemporal expression of a given gene, even with a known set of active enhancers.

There are several properties of enhancers that make predicting gene expression complex. First, while chromatin marks are correlated with enhancer activity, it remains unclear whether all H3K27ac regions represent functional enhancers (Pennacchio et al., 2013; Dickel et al., 2014). Second, enhancer-promoter (E-P) interactions can occur across tens of thousands of kilobases, sometimes skipping tens of intervening promoters (Claussnitzer et al., 2015). Third, many different enhancers may regulate a single gene (Donnard et al., 2018; González et al., 2015). Fourth, each enhancer can regulate several genes (Ghavi-Helm et al., 2014) and even simultaneously (Fukaya et al., 2016). Finally, recent reports have shown that some E-P interactions may not bring regulatory elements into close proximity to their targets but instead can form phase-separated condensates of activators, co-activators, and transcriptional machinery (Benabdallah et al., 2019).

Chromosome conformation capture methods have begun to reveal the complexity of eukaryotic gene regulation by capturing the interactions between genomic loci (Nott et al., 2019; Rubin et al., 2017; Mumbach et al., 2017; Li et al., 2012). However,

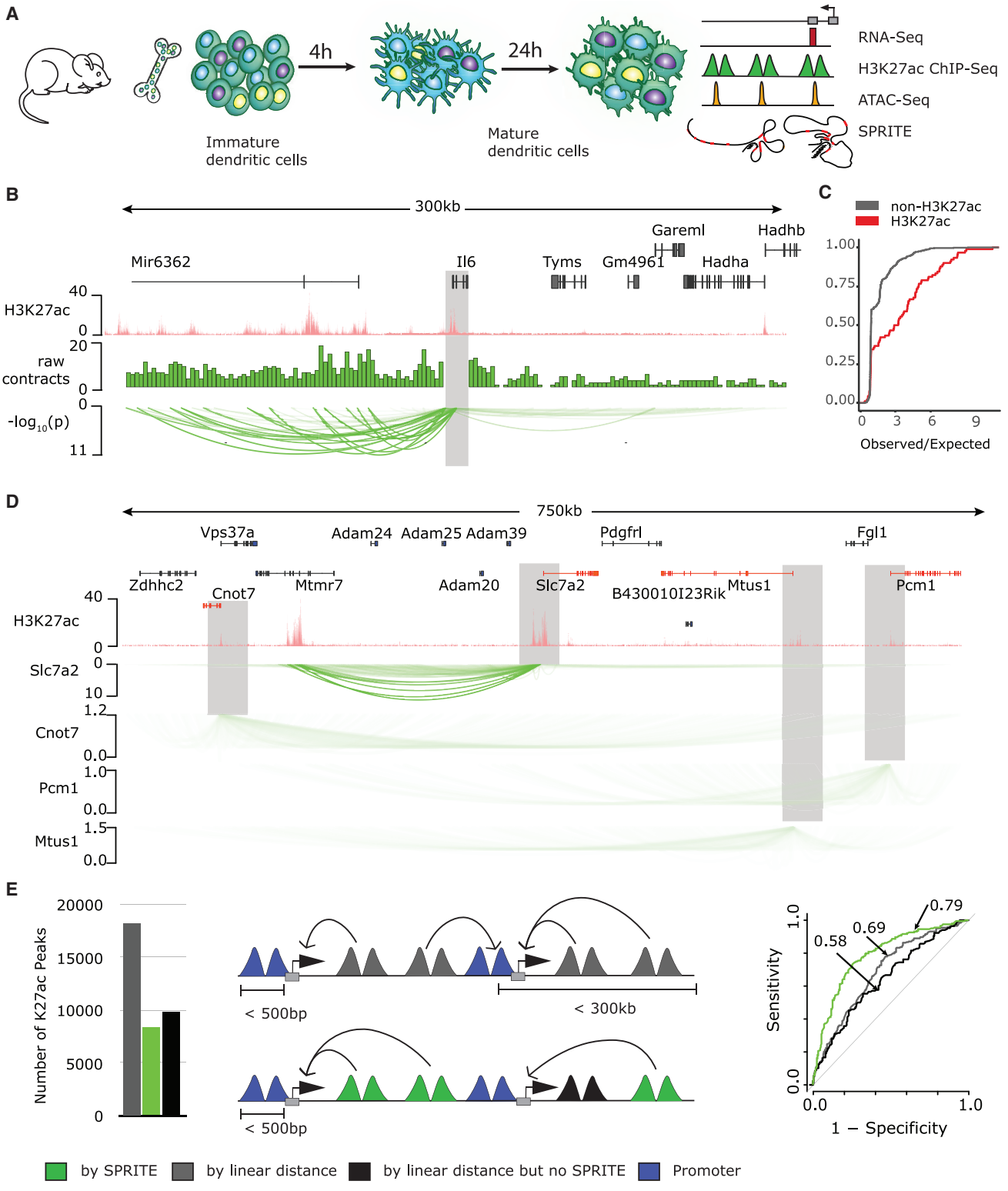


Figure 1. High-Resolution Identification of Enhancer-Promoter (E-P) Interactions Using SPRITE

(A) Mouse bone-marrow-derived dendritic cells (BMDCs) were treated with LPS for 0, 4, or 24 h. SPRITE was performed at each time point and integrated with RNA-seq, chromatin immunoprecipitation sequencing (ChIP-seq), and ATAC sequencing (ATAC-seq) data.

(legend continued on next page)

these methods can only measure pairwise interactions across cell populations, complicating the discernment of multiway interactions that occur simultaneously between multiple enhancers and promoters. It has been recently reported that some genomic interactions occur across distances that are beyond the cross-linking distance of the current chromosome conformation methods and thus may be depleted in maps built using these protocols (Benabdallah et al., 2019; Dekker, 2016; Giorgetti and Heard, 2016).

We recently described split-pool recognition of interactions by tag extension (SPRITE), a new method for mapping higher-order spatial interactions in the nucleus. We showed that SPRITE can map long-range DNA contacts that cannot be observed by proximity ligation methods (Quinodoz et al., 2018). Moreover, we showed that SPRITE can map multiway contacts that occur simultaneously within a single cell. In this study, we extended these observations to dissect regulatory contacts. We used SPRITE and developed an adaptation called SPRITE coupled with immunoprecipitation (SIP) to comprehensively define multiway E-P interactions in mouse bone-marrow-derived dendritic cells (BMDCs) stimulated with liposaccharide (LPS) with the goal of defining a predictive model of gene induction after stimulation.

We found that SIP provides evidence for strong E-P interactions for less than half of all active enhancers in BMDCs. Using sequence motifs within enhancers with SIP E-P interactions in a linear regression model dramatically improved its ability to computationally predict gene induction after LPS stimulation. We then show that sequence-motif-based models that predict gene expression can be used to quantify the impact of genetic variability on gene expression. Our analysis of multiway E-P interactions reveal that E-P interactions involve multiple genes that form transcriptional hubs and that the DNA elements within these hubs synergize to enhance transcription. Interestingly the stability of these E-P hubs predicts the stability of expression across a cell population. These hubs can be dynamically organized during LPS response by key transcriptional regulators, such as Jun, Fos, and other members of the activator protein 1 (AP1) family transcription factors (TFs).

RESULTS

SPRITE Identifies High-Resolution Promoter Interactions in Dendritic Cells (DCs)

Stimulating innate immune cells with TLR4 ligands results in highly reproducible temporal patterns of gene expression and epigenetic states. The temporal trajectories of genes and *cis*-regulatory elements (measured by H3K27ac signal and chromatin accessibility) follow four broad categories: early induction (<2 h post-LPS stimulation), late induction (>2 h post-LPS stimulation), downregulated, and nonchanging. These clearly defined modules of genes and their associated regulatory elements make this system well suited to understand the nuances of how enhancers regulate gene expression regulation (Garber et al., 2012; Link et al., 2018; Bornstein et al., 2014; Donnard et al., 2018).

To characterize regulatory interactions, we complemented previously generated expression and chromatin state maps (Garber et al., 2012; Donnard et al., 2018) with 3D chromatin interactions from SPRITE for BMDCs at 0, 4, and 24 h after LPS stimulation (Figure 1A). Each library yielded an average of 225 million valid reads (reads having all barcodes after the split and pool process; STAR Methods; Table S1) from 50 million unique clusters of interacting DNA molecules (i.e., having two or more unique genomic regions) (Figure S1A). To capture all 3D interactions across LPS stimulations, we merged the data from all three time points.

Manual inspection of interactions involving promoters of highly induced genes revealed that the unnormalized SPRITE signal was strongest on distal regions marked by H3K27ac (Figures 1B and 1C). To test the generality of this observation, we employed a viewpoint-centric analysis. We centered our analysis on transcription start sites (TSSs) marked by the active H3K27ac histone mark. For each active promoter, we identified interactions that were overrepresented relative to linear-distance-based signal decay and genomic coverage (STAR Methods).

This viewpoint analysis further confirmed the enrichment of active regulatory regions among all promoter-interacting loci. This is clear at several distinct genes, where distal regulatory elements display a strong association with their promoters. For example, an enhancer located within the *Mtmr7* shows spatial proximity only to the promoter of the *Slc7a2*, even though there are five intervening genes between *Mtmr7* and *Slc7a2*. Notably, no other expressed gene promoter in this locus interacts with this enhancer (Figure 1D). Globally, when focusing on active promoters, the SPRITE signal was enriched for active regulatory regions compared to nonactive regions (Figure S1B). Using this signal, we were able to connect 6,636 (76%) active promoters with ~9,000 (40% of distal H3K27ac regions) putative enhancers in DCs (Figure 1E, left).

For the majority of H3K27ac regions (60%), we did not detect interactions with an active promoter. This could be because (1) these regions are not involved in gene regulation in steady state or in response to LPS, (2) there was a lack of power to detect interactions, or (3) enhancers may be acting from a distance that is beyond what SPRITE can capture (Benabdallah et al., 2019). To test the accuracy of the associations we used a machine learning approach and evaluated the ability of transcription factor binding motifs (TFBMs) within associated putative enhancers to

(B) 3D interactions of the *Il6* promoter (gray box). H3K27ac signal is shown in pink. For tracks 2–4, the region around the *Il6* promoter was divided into 5-kb bins, and each bin was assigned the number of SPRITE clusters that have an overlapping read (raw contacts, green second track). Interactions with the *Il6* promoter are shown as arcs, with height proportional to the interaction score.

(C) Cumulative distribution plots of the interaction score of distal bins within H3K27ac regions (red) and non-H3K27ac regions (gray).

(D) Similar to (B), showing interaction profiles of expressed genes in the *Slc7a2* locus (colored in red). The H3K27ac profile is shown in pink. Viewpoints are highlighted in gray and for each viewpoint, and interactions are displayed as in (B).

(E) Left: number of enhancers associated with linear proximity (gray), SPRITE signal (green), or delta (black). Middle: schematic depicting distance-defined enhancers (purple) SPRITE-defined enhancers (green), and delta enhancers (black). Right: receiver operating characteristic (ROC) curve for each set of enhancers defined in the middle plot.

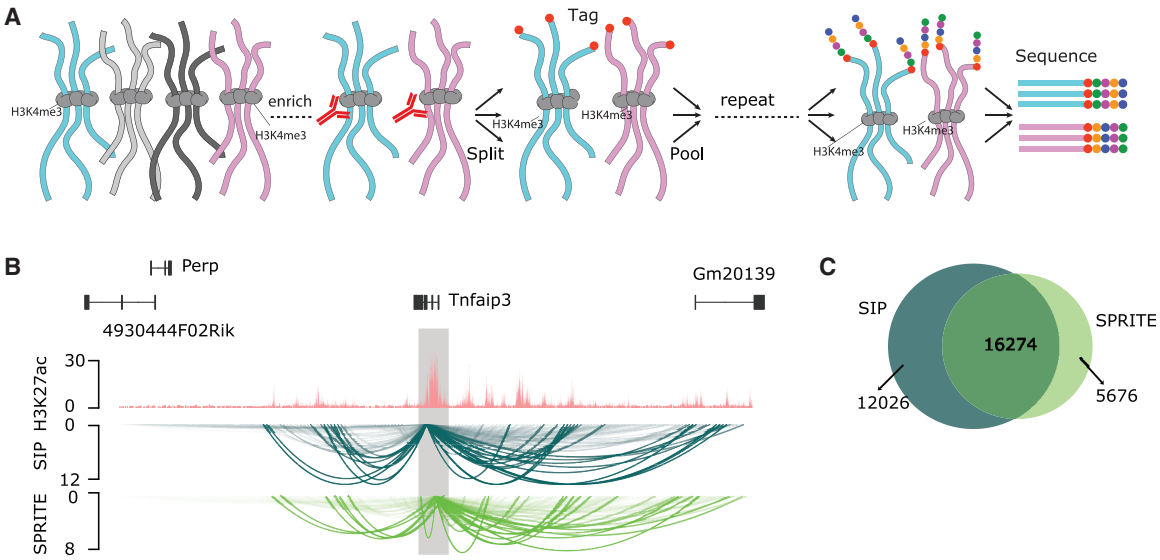


Figure 2. SPRITE Coupled with Immunoprecipitation (SIP) to Enrich in Promoter Interactions

(A) Schematic of the SIP workflow. Four different DNA clusters are shown. Proteins that are crosslinked with DNA are shown in gray. Clusters containing the protein of interest are pulled down using an antibody. The resulting material is barcoded using the SPRITE protocol.

(B) SIP-H3K4me3 and SPRITE capture similar interactions of the *Tnfaip3* gene promoter. Top track shows the H3K27ac signal (pink), and the middle and bottom tracks show significant interactions identified by SIP-H3K4me3 (dark green) and SPRITE (light green).

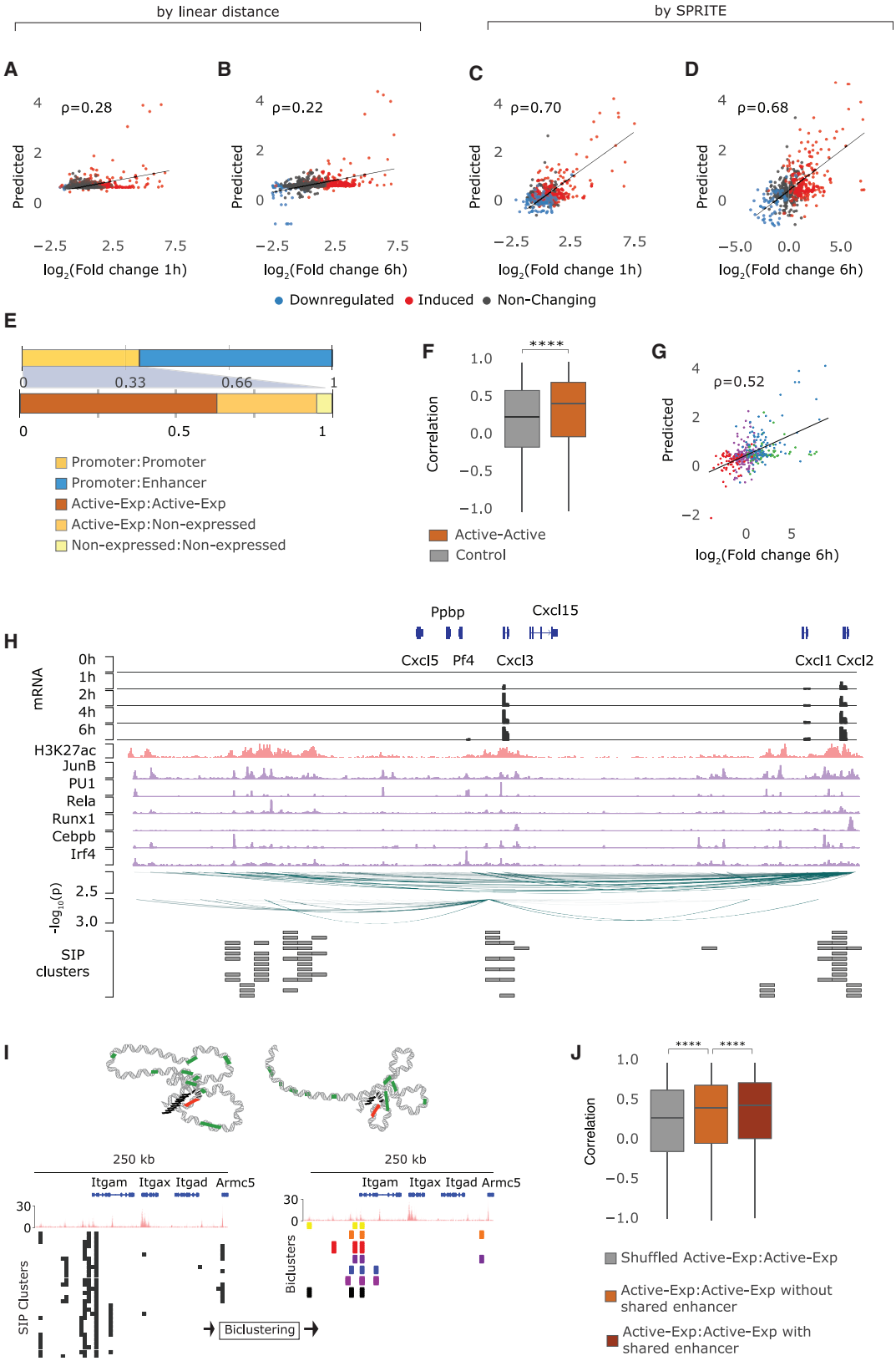
(C) Overlap of E-P interactions predicted using SIP-H3K4me3 and SPRITE.

distinguish LPS responsive genes from those that are not (STAR Methods). We built a random forest classifier using TFBMs to predict whether a gene is induced using three different sets of putative enhancers (Figure 1E, middle): (1) putative enhancers within 300 kb of a promoter (distance defined), (2) putative enhancers associated with a promoter region by SPRITE (SPRITE defined), and (3) putative enhancers that are assigned to promoters based on a linear proximity of $\leq$300 kb but have no enriched signal in SPRITE (delta). We found that the prediction accuracy of the models (as measured by area under the curve [AUC]) improved to 0.79 for the classifier built with SPRITE-defined enhancers relative to an AUC of 0.69 for the classifier based on distance-defined enhancers (Figure 1E, right; and Figure S1C). The poorer performance of the classifier built with distance-defined enhancers is independent of the distance threshold used to assign distance-defined enhancers or the distance defined by SPRITE associations (STAR Methods; Figures S1D and S1E). The classifier built with enhancers not connected by SPRITE data has almost no predictive power (0.58). Notably, the most influential TFBMs for the classifier built on SPRITE-defined interactions are the DNA binding motifs of TFs that are key regulators of LPS-induced genes (Medzhitov and Horng, 2009) (Figure S1F). This suggests that the model built on our enhancer assignments captured the biologically relevant signal and that unconnected enhancers introduce noise that impacts the performance of the classifier.

SPRITE-IP (SIP): a Zoom-In Strategy to Enrich for Protein-Centric Interactions at High Specificity and Sensitivity

To increase the sensitivity to detect E-P interactions and their temporal dynamics, we coupled SPRITE with immunoprecipitation in a method we called SPRITE coupled with immunoprecipitation (SIP) (Figure 2A). To enrich for interactions with promoter regions, we used an antibody against H3K4me3, a chromatin modification associated with promoters (Lauberth et al., 2013; Koch et al., 2007; Heintzman et al., 2007). The resulting libraries were highly enriched for H3K4me3-marked TSSs (Figure S2A; Table S1). SIP libraries are 3-fold enriched for clusters containing promoters and putative enhancers compared to SPRITE data. We further confirmed that interactions captured by our combined analytical and experimental approach represented bona fide interactions by generating a low coverage mouse embryonic stem cell (mESC) SIP library. We found that 72% of SIP interactions were also identified using an assay where the Hi-C library preparation is followed by a chromatin immunoprecipitation ChIP (HiChIP) with a H3K27ac antibody (H3K27ac-HiChIP) data (Figure S2B; STAR Methods), and our analysis recapitulated most (4 of the 5) of the previously identified interactions from 3C/4C methods (Table S2, sheet1) and 9 out 11 (Table S2, sheet2) experimentally validated distal enhancer interactions, including the enhancer associated with *Sox2* promoter (Figure S2C) (Li et al., 2014; Moorthy et al., 2017; Juric et al., 2019). Manual inspection of several loci in BMDCs showed that similar interactions are enriched in both SIP and SPRITE libraries (Figure 2B). We conclude that H3K4me3 SIP enriches for regulatory interactions that involve promoters, thus reducing the effective sequencing depth required to capture them.

We generated SIP libraries for BMDCs at the same time points for which we generated SPRITE libraries (0, 4, and 24 h). Interactions defined by both SPRITE and SIP are highly reproducible across experiments: (1) SIP recovered the majority ($\sim$74%) of E-P interactions identified by SPRITE (Figure 2C; STAR



(legend on next page)

Methods) while also yielding a large number of new interactions; (2) the frequency of interactions involving promoters is well correlated between SPRITE and SIP libraries (Figure S2D), as well as between different time points in SIP libraries (STAR Methods); (3) interactions that are specific to either SPRITE or SIP have lower scores than interactions enriched in both the assays (Figure S2D).

Given the high concordance of signal and interactions called between SPRITE and SIP libraries at the three time-point libraries (Figures S2E–S2J), we merged our SIP and SPRITE data, which together represent ~1 billion unique reads with 78M medium-sized clusters (2–100 unique DNA molecules) across all three time points (0, 4, and 24 h). This combined dataset uncovered 34,187 interactions involving 7,130 active promoters (86% of all active promoters) (Table S2). The 1,122 genes with active promoters for which we still could not detect interactions had low expression or were not highly induced by LPS (Figure S2L).

SPRITE Contacts Dramatically Improved Gene Expression Prediction

TF binding is critical to gene regulation, and it has been shown that gene expression changes can be predicted from TFBMs using regression models and that the performance of these models was vastly improved by iteratively reassigning enhancers of inactive genes to active genes (González et al., 2015). We therefore reasoned that improvement in E-P associations would result in improved regression model performance and that the accuracy of E-P associations could be quantified by the prediction error of regression models that relied on TFBMs in associated enhancers.

We assigned a score to each TFBM in accessible regions of regulatory elements using SeqGL (Setty and Leslie, 2015), which reflects the importance of the TFBM instance to distinguish enhancer regions from non-enhancer regions. We previously found that accounting for the timing when TFBMs are accessible improved a TFBM-based gene classifier (Donnard et al., 2018); therefore, we computed SeqGL scores for TFBMs separately in early-induced, late-induced, downregulated, and nonchanging enhancers. To predict changes in gene expression after LPS stimulation, we used an elastic net regression model that is based on the sum of the scores across all regulatory elements with similar DNA accessibility kinetics for each gene (STAR Methods).

We first built regression models for predicting gene expression changes after LPS stimulation in which TFBM scores were combined for enhancers within 300 kb of a gene promoter. This model had poor prediction at both 1 h (Spearman rank correlation $\rho = 0.3$) and 6 h post-LPS stimulation ($\rho = 0.21$) (Figures 3A and 3B). We then built models using only TFBMs within enhancers for which there was 3D interaction evidence and observed a dramatic improvement, with higher correlation between the observed fold change and predicted fold change at 1 and 6 h after LPS stimulation ($\rho = 0.70$ and $\rho = 0.66$ respectively; Figures 3C and 3D; STAR Methods). Together, these results show that assigning enhancers based on SIP interactions instead of linear distance notably improves both the specificity and sensitivity of the predictive models of gene expression.

Upon examining the features with non-zero coefficients, we found that TFBMs such as REL, STAT, SPI, and CEBP have positive coefficients consistent with their role as activators after 1 h of LPS stimulation, while negative coefficients for PRDM1, RUNX1 TFBMs suggest a repressive function for these factors (Figure S3A). Similarly, the 6-h model points to STAT and SPI as the strongest activators while also highlighting PRDM1 and RUNX family members as transcriptional repressors (Figure S3B).

It is important to note that models built on all accessible DNA regions without considering the temporal dynamics reduced the model predictive power (Figure S3C). However, scoring schemes that used the intensity of either the assay for transposase-accessible chromatin (ATAC) or H3K27ac signal did not improve the model performance (Figure S3D).

Induced, Coregulated Genes Form Transcriptional Hubs

It has been previously observed that physically interacting promoters form transcriptional hubs that result in coregulated gene expression (Li et al., 2012) due to synergistic binding of transcriptional machinery between coregulated genes (Furlong and Levine, 2018; Rieder et al., 2012; Edelman and Fraser, 2012). We sought to probe these observations by exploiting our data on gene expression changes in response to LPS and the unique ability of SIP to capture complex genomic

Figure 3. SIP-Supported E-P Interactions Improve Models that Predict LPS Responsiveness from Transcription Factor Binding Sequence Motifs

(A–D) Scatterplots of observed versus predicted $\log_2$ fold change of gene expression for linear models trained using TFBMs within distal regulatory elements that were associated to gene promoters using linear distance at 1 h (A) and 6 h (B) after LPS stimulation or SIP signal at 1 h (C) and 6 h (D) after LPS stimulation.

(E) Fraction of regulatory interactions that are promoter-promoter and promoter-enhancer (top). Fraction of promoter-promoter pairs involving promoters of two expressed genes (active-expressed: active-expressed), a promoter of an expressed gene and a promoter of a gene that is not expressed (active-expressed: nonexpressed) and promoters of two genes that are not expressed (nonexpressed: nonexpressed) (bottom).

(F) Boxplots of correlation coefficient of gene expression across time after LPS stimulation for noninteracting promoter pairs sampled after controlling for expression level and genomic context (gray) and for active-expressed: active-expressed promoter pairs (orange). Significance was calculated using Wilcoxon rank-sum test.

(G) Same model as in (D), but the predictions excluded TFBMs scores from active-expressed: active-expressed promoter pairs.

(H) Integrative Genomics Viewer (IGV) plot of the *Cxcl3* locus. RNA-seq signals at 0, 1, 2, 4, and 6 h post-LPS in black tracks. H3K27ac signal is shown in pink. Signals for the TFs Junb, Pu1, Rela, Runx1, Cebpa, and Irf4 are shown in purple. Pairwise interactions of *Cxcl3* and *Cxcl2* are shown as arcs. SIP clusters with more than two interactions are shown as gray boxes in the bottom track. Each row represents a unique cluster.

(I) Two possible regulatory conformations of a promoter (red) with enhancers (green) and transcriptional output (black). The *Itgam* gene exemplifies the approach used to identify multiway interactions from SIP data. Clusters containing three or more loci are biclustered, and a permutation approach is used to assess the significance of each bicluster.

(J) Boxplot of expression correlation coefficients across time for noninteracting promoter pairs (same as F) (gray), genes with interacting promoters and no shared enhancer (orange), and genes with interacting promoters and shared enhancer(s) (brown). Significance was calculated using Wilcoxon rank-sum test.

interactions involving multiple DNA elements at a single-molecule resolution.

A large fraction of pairwise interactions in our dataset (35%, 11,953) represent direct contacts between two gene promoters (P-P interactions). In the vast majority (95%) of P-P interactions, at least one gene is expressed (>10 transcripts per million [TPMs]), and for the majority of pairs (66%, 7,530), both genes are expressed (Figure 3E). We observed that LPS-induced genes with interacting promoters tend to have higher temporal correlation than gene pairs with no interacting promoters, even after controlling for expression level, the number of enhancers interacting with individual promoters, and linear distance between the promoter pairs ($p < 10^{-15}$, Wilcoxon rank-sum test; Figure 3F; STAR Methods). Furthermore, our model performance worsens when TFBMs within interacting promoters are excluded ($\rho =$ 0.51versus $\rho = 0.65$, Figure 3G).

Multiway interactions involve key cytokines that are highly induced after LPS stimulation. For example, the chemoattractants *Cxcl1*, *Cxcl2*, and *Cxcl3* on chromosome 5 are all early induced and have very similar temporal dynamics. Multiway interactions in this region show an intertwined network in which promoters of these three chemokines simultaneously interact with one or more enhancers in the same cell. This interaction network is specific to these three cytokines, as interleaved genes in the region that are not induced are excluded from this network (Figure 3H). Hence, while only a subset of TFs bind any individual enhancer or promoter, the combined interactome results in copious binding of all Pu1, Junb, Rela, Runx1, Cebpb, and Irf4, all key regulators of early-induced genes (Donnard et al., 2018; Garber et al., 2012; Link et al., 2018). Such multiway interaction networks involving early response cytokines are also observed in the *Il1a,Il1b*, *Ccl5*,*Ccl3*, and *Ccl17*,*Ccl22* clusters on chromosomes 2, 11, and 8, respectively (Figures S3E–S3G). To expand these observations, we employed a biclustering approach (STAR Methods) that can reveal common multiway interactions that occur frequently across cells in the population while being robust to missing information resulting from dropouts in single clusters (STAR Methods; Figure 3I). For example, biclustering the SIP clusters in the *Itgam* locus (Figure 3I bottom left) revealed seven unique multiway interactions upon biclustering (Figure 3I, bottom right; STAR Methods). At a false discovery rate (FDR) of 5%, we found that 1,895 P-P pairs that included two expressed promoters and at least one enhancer (1,619) or a third promoter (276). When two expressed promoters form a hub with a shared enhancer, they tend to be more correlated than when the interactions involved only two promoters and no enhancer (Figure 3J). Together, our results provide *in vivo* support for the notion that transcriptional hubs can coordinate gene expression (Fukaya, Lim, and Levine, 2016).

Variability in Regulatory Configurations Correlates with Stochasticity in Gene Expression

Because biclustering identifies interactions that are common across the cell population, we reasoned that the genes in such clusters may also have more stable expression. To test this hypothesis, we generated single-cell RNA-sequencing (scRNA-seq) libraries for BMDCs at 0, 1, and 4 h post-LPS stimulation. We focused on highly expressed genes (bulk RNA-seq data maximum TPM >100) for which we could reliably estimate variability in the scRNA-seq data. We focused on genes that were associated with six or more enhancers, representing a complex regulatory landscape (Donnard et al., 2018; González et al., 2015). After this filtering, our analysis concentrated on 2,038 genes, from which we could identify biclusters for 638 genes and only pairwise interactions for the remaining 1,400 genes. While both sets consist of genes have similar expression levels and SIP read coverage (Figures S4A and S4B), genes with consistent multiway interactions have lower expression variability across single cells relative to genes without consistent multiway interactions (Figure 4A), regardless of LPS responsiveness. For example, *Rps28* and *Gnai2* are not responsive to LPS and are expressed at very similar levels; however, *Rps28* has a lower gene expression variation across cells than *Gnai2* (Figure 4B). The inspection of biclusters for both genes revealed that *Rps28* had clear biclusters, while none are detectable for *Gnai2* (Figures 4C and 4D). These results indicate that stability in chromatin interactions confers reduced expression variation across a population of cells (Figure 4E).

Quantitative Induction Predictive Model Identifies the Effect of Enhancer Loss

A recent study explored the impact of naturally occurring genetic variability on the transcriptional response in bone-marrow-derived macrophages (BMDMs) stimulated with Kdo(2)-lipid A (KLA), a TLR4 agonist analogous to LPS, in different mice strains (Link et al., 2018). Because promoter changes could not explain the large differences between LPS responsive genes among strains, the authors hypothesized that distal regulatory elements, rather than promoters, were responsible for these changes. We sought to test this hypothesis using our 3D interaction data. We note that our data were generated in BMDCs, while Link et al. used BMDMs. However, our comparison showed that both cell types have a remarkably similar transcriptional response to LPS (Figure S5A), and both cell types share a very high fraction (70%) of active regulatory elements (Figure S5B).

In this study, we focused on BMDMs derived from C57BL/6 (C57) and SPRET/EiJ (SPRET) strains. These two strains are highly divergent, with an average of one single nucleotide polymorphism (SNP) every 120 bp. BMDMs from the two strains have 4,820 differentially expressed (DE) genes (p-adjusted < 0.01 and absolute fold change > 0.5; STAR Methods) and 11,345 differential H3K27ac regions (absolute fold change > 4 in H3K27ac signal, STAR Methods). For 5,449 of these regions, we could identify significant interactions with at least one promoter. Remarkably, the majority of these enhancers (62%) interact with genes that are DE between strains, compared to genes that are not DE (enrichment of 1.53, Fisher's exact test $p < 10^{-15}$). Importantly, when defining E-P associations by distance, this enrichment disappears, and instead, we observed a slight depletion in enhancers with differential H3K27ac signal associated with DE genes (Figure 5A). Changing the window size or when using an interaction radius defined by the SIP signal did not recapitulate the enrichment observed with SIP-defined E-P associations (Figure 5A; STAR Methods).

Next, we tested our model by focusing on genes that were induced at different levels between the two strains. For example, *Cxcl10* is induced 2.8-fold more in C57 BMDMs compared to

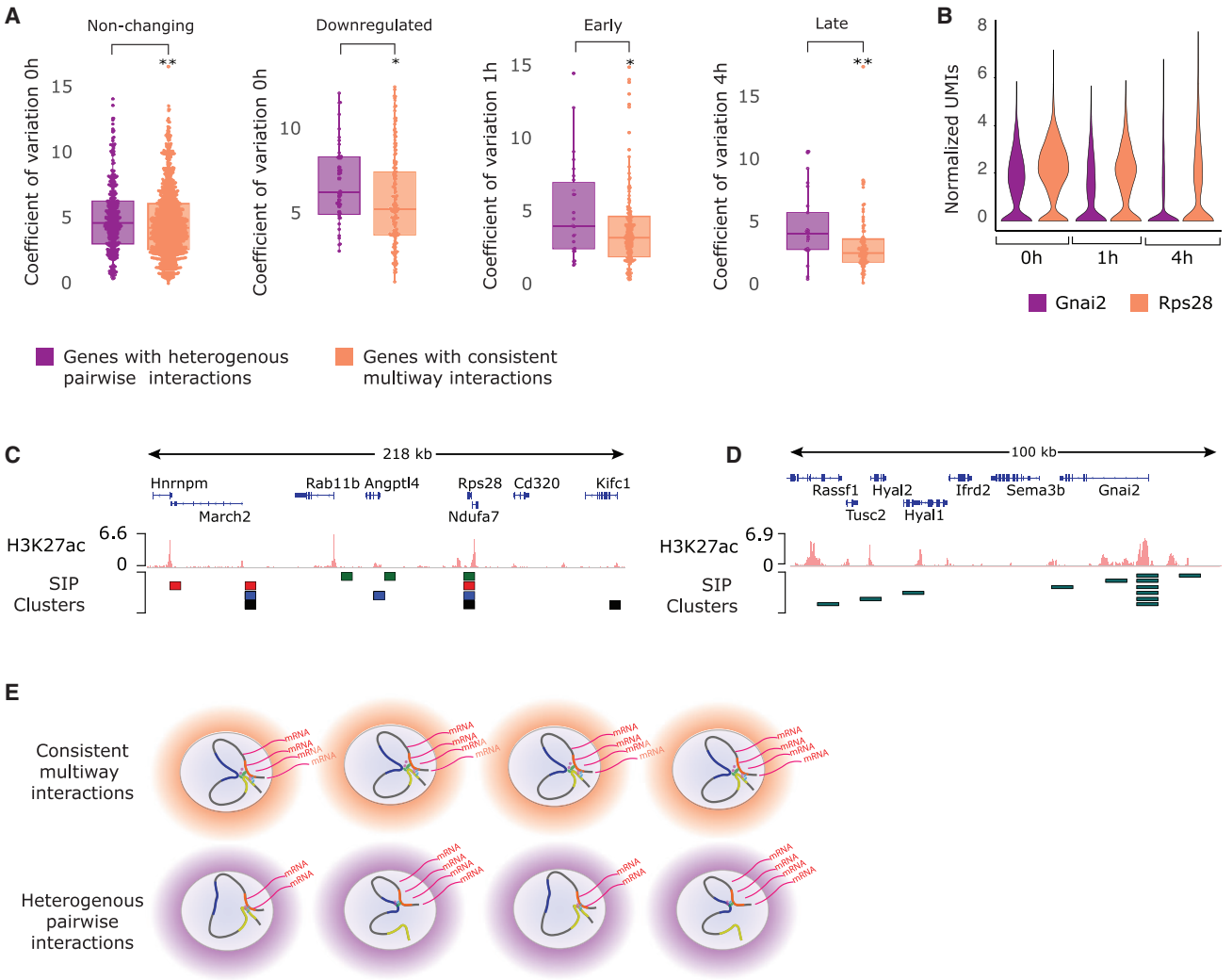


Figure 4. Cell-to-Cell Homogeneity in E-P Reduces Cell-to-Cell Expression Variability

(A) Distribution of the coefficient of variation of expression for genes with (orange) and without (purple) consistent multiway interactions at 0 h for nonchanging genes and downregulated genes, at 1h for early-induced genes, and at 4 h for late-induced genes. Significance was calculated using Wilcoxon rank-sum test.

(B) Gene expression distribution in single cells derived from scRNA-seq at 0, 1, and 4 h after LPS stimulation for *Gnai2* (purple) and *Rps28* (orange).

(C) H3K27ac signal across a 240-kb region corresponding to the *Rps28* gene (top). Higher-order interactions of SIP clusters are shown, where each configuration has at least three unique SIP clusters (bottom).

(D) H3K27ac signal across a 105-kb region corresponding to the *Gnai2* gene (top). Pairwise interactions of SIP clusters are shown (bottom).

(E) Consistent multiway E-P interactions drive expression similarly across different cells by bringing together similar regulatory elements compared to heterogeneous, pairwise E-P interactions.

SPRET BMDMs (Figure 5B). All putative enhancers within this locus are shared between C57 and SPRET, except for one that has no H3K27ac signal in SPRET BMDMs (Figure 5C, black box). Interaction data show that although the enhancer is 100 kb away, it physically interacts with *Cxcl10* in C57 BMDCs. Interestingly, the same enhancer also interacts with the adjacent *Naaa* gene, which also has lower induction in SPRET (Figure 5B). This strongly suggests that this enhancer with both of the gene promoters may be important to the higher induction levels observed in C57 BMDMs.

There are a total of 48 genes that show significant differential LPS response between SPRET and C57 BMDMs. 37 genes have a stronger induction in C57 BMDMs, while 11 have a stronger induction in SPRET BMDMs. To validate our models (Figures 3C and 3D), which are trained on C57 sequence and transcriptional response, we set motif scores for the C57 specific enhancers to zero to *in silico* simulate their inactivity in SPRET. The prediction error for fold change in expression is significantly lower than when we use enhancers that are active in SPRET BMDMs ($p < 10^{-5}$, Wilcoxon rank-sum test) (Figure 5D; STAR Methods). Therefore, when using E-P interactions determined from SIP, our model can incorporate the information encoded in TFBMs within enhancers that have variable activity across individuals to estimate the impact of the loss of an enhancer on the expression of its target genes.

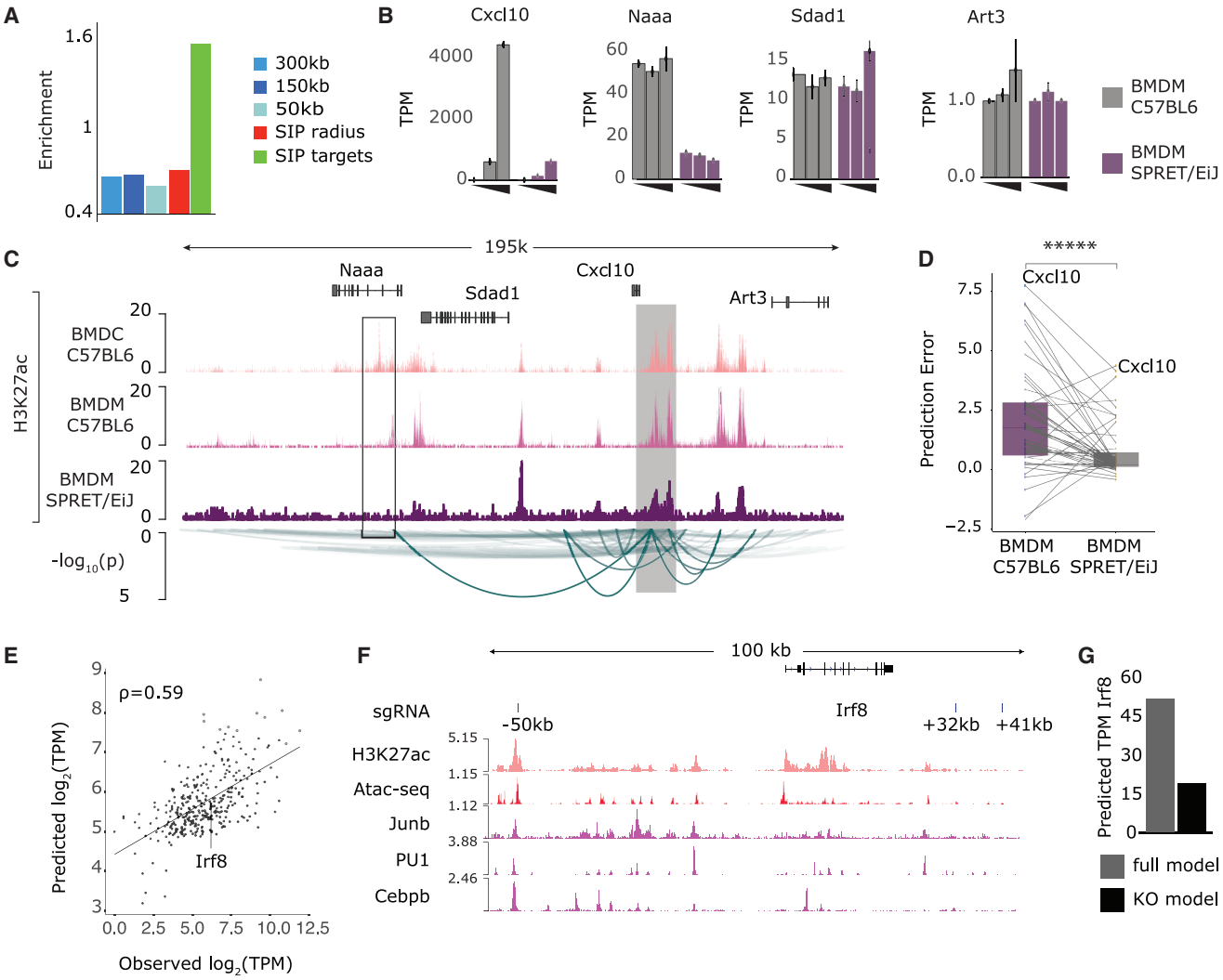


Figure 5. Linear Models of Gene Expression Trained from TFBMs within SIP-Supported E-P Interactions Help Interpret Enhancer Activity Variability in a Population

(A) Enrichment of enhancers that are C57 specific among enhancers that are associated with genes that are DE between C57 and SPRET using distance-defined within 300 kb, 150 kb, 50 kb, the SIP interaction radius, and the SIP interaction signal.

(B) Expression in TPMs of genes in the neighborhood of *Cxcl10* in C57 BMDMs (gray) and SPRET BMDMs (purple) at 0, 1, and 6 h after KLA stimulation.

(C) IGV plot of the *Cxcl10* locus. Tracks from top to bottom show H3K27ac signal in C57 BMDCs at 1 h post-LPS stimulation, H3K27ac signal in C57 BMDMs at 1 h post-LPS stimulation, and H3K27ac signal in SPRET BMDMs at 1 h post-LPS stimulation; interactions with the *Cxcl10* promoter are shown as arcs. Arc height is proportional to the interaction score.

(D) The model trained in C57 using SIP data was run to predict fold changes in SPRET at 1 and 6 h, first using scores for motifs in both shared and C57 specific enhancers (purple) and then setting to zero all motifs within C57 specific enhancers (gray). Boxplot showing the prediction error for each model at the maximum expression time point for DE genes after LPS stimulation is shown. Each dot corresponds to a gene, and lines connect the prediction error for a gene across two models. Significance was calculated using Wilcoxon rank-sum test.

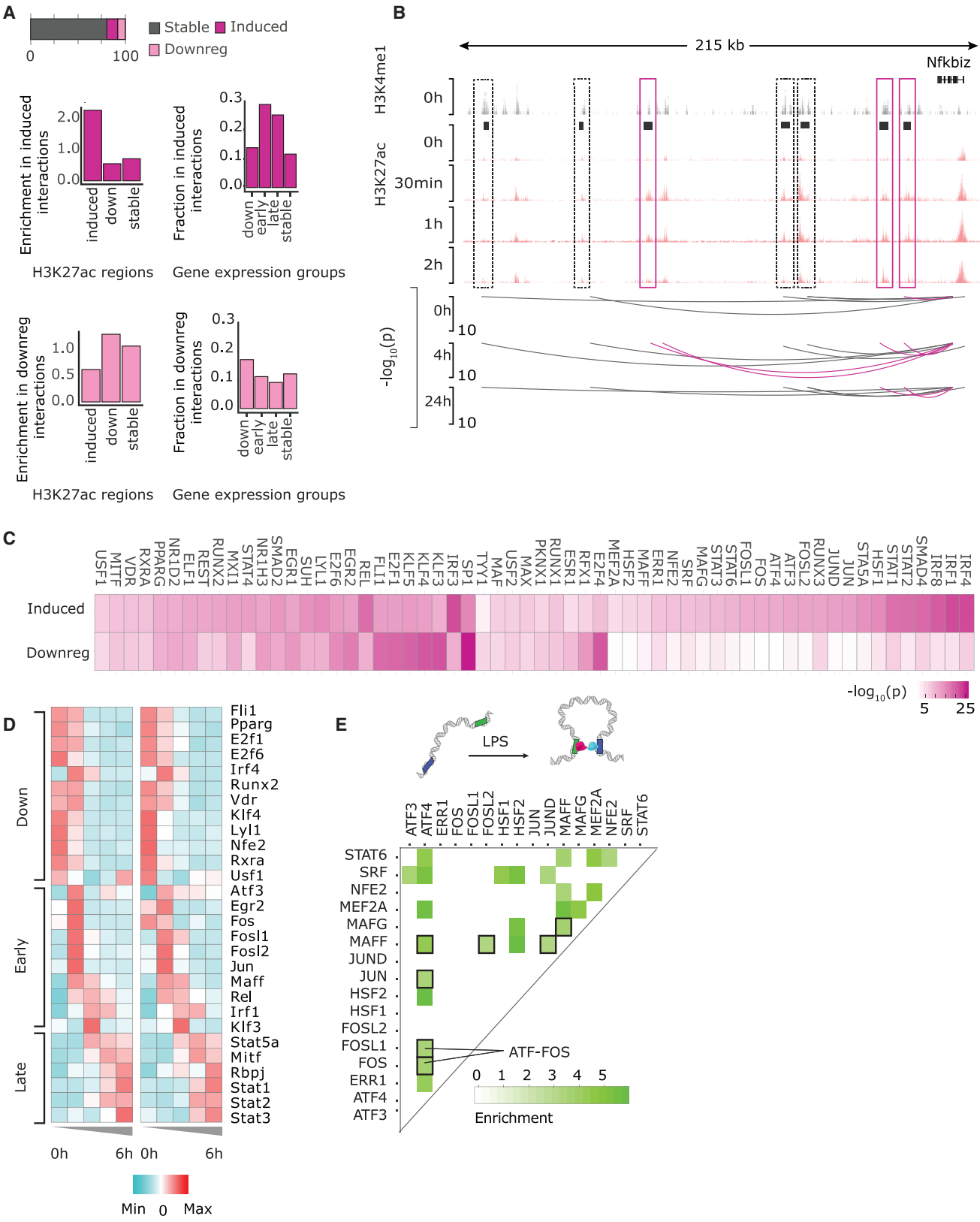
(E) Scatterplot of observed and predicted $\log_2$TPM in unstimulated cells using a regression model that includes TFBMs from the interactions identified by SIP.

(F) IGV plot of the 100-kb region around the *Irf8* promoter. Guide RNAs used in Durai et al. (2019) are shown in the top track, and H3K27ac, ATAC-seq, Junb, Pu1, and Cebpb signals are shown in pink, red, and purple, respectively.

(G) Predicted expression level of *Irf8* using feature matrix that includes TFBMs within the +50-kb enhancer (gray, full model) and using feature matrix that excludes the TFBMs within the +50-kb enhancer (black, knockout [KO] model).

SPRITE Contacts Improve Steady-State Prediction of Expression Levels

Predicting gene expression in steady-state conditions is challenging. As opposed to LPS-dependent genes, which are mostly regulated at the transcriptional level, steady-state mRNA levels depend more heavily on post-transcriptional processes (Rabani et al., 2011, 2014). However, given the improvements in model prediction when using E-P associations inferred from SPRITE



(legend on next page)

data, we next investigated whether improvement in E-P associations also led to improved modeling of steady-state gene expression. We fit a model to medium and high complexity genes, specifically those that are associated with at least four enhancers, to predict gene expression levels of these genes (STAR Methods). Although this model performs worse than the model that predicts LPS-induced changes, when relying on TFBMs in SPRITE-defined E-Ps, the model performs better than when E-P interactions are distance defined ($\rho = 0.59$ versus $\rho = 0.37$ Figures 5E and S5C). We tested this model using data from a loss-of-function analysis of the *Irf8* enhancer (Durai et al., 2019). Three enhancers in the vicinity of the *Irf8* gene were targeted; two of these enhancers, both downstream of *Irf8*, resulted in severe defects in DC development, but neither is active in BMDCs (Figure 5F). A third enhancer, which is 50 kb upstream of *Irf8*, was shown to impact *Irf8* expression levels in macrophages and conventional DCs (Durai et al., 2019). This enhancer is also active in BMDCs. When the TFBM scores within this enhancer are set to zero, to *in silico* represent cells where the enhancer is knocked out, the model predicts almost three times lower expression (TPMs) for *Irf8* gene compared to the model that uses the original scores (51.62 versus 18.89 TPM). As such, this *in silico* inactivation of the enhancer is consistent with the reported drop in Irf8 levels when the enhancer is genetically ablated (Figure 5G).

AP1 Family TFs Mediate the Formation of Inducible Regulatory Interactions

We previously showed that similar to gene expression, accessibility and activity of *cis*-regulatory elements (measured by ATAC and H3K27ac) have well-defined temporal patterns. Up to 30% of regions undergo significant changes in either chromatin accessibility or H3K27ac signal after stimulation (Garber et al., 2012; Donnard et al., 2018). We found that changes in DNA accessibility and histone acetylation are not concomitant with changes in 3D interactions between these regions. In fact, as others have observed in different contexts (Rubin et al., 2017; Ghavi-Helm et al., 2014; Jin et al., 2013), only a small fraction (10%) of interactions have detectable changes upon LPS stimulation (Figure 6A, top; Figures S2E–S2K; STAR Methods). Regardless, LPS-induced interactions are generally enriched in regions with inducible H3K27ac ($p < 10^{-37}$, Fisher's exact test) and are associated with LPS-induced genes (Figure 6A, middle). Similarly, lost interactions are enriched in genes downregulated by LPS ($p < 0.003$, Fisher's exact test) (Figure 6A, bottom). These observations suggest that most E-P interactions involving LPS-inducible enhancers are preestablished prior to LPS stimulation, while enhancers are poised. One of the commonly recognized signatures of poised enhancers is the presence of H3K4me1 signal but lack of H3K27ac signal (Rada-Iglesias et al., 2011). For example, the *Nfkbiz* locus includes seven enhancers with strong H3K4me1 signal. All of these enhancers are induced within 30 min of LPS stimulation. Three of these enhancers (Figure 6B, pink boxes and arcs) form interactions only after LPS stimulation, while the other four interact with the Nfkbiz promoter prior to LPS stimulation (Figure 6B, black boxes and arcs). Genome-wide, we classified enhancers as active or poised based on the H3K27ac and H3K4me1 signals (STAR Methods) and found that 80% of the induced H3K27ac peaks with stable interactions are poised enhancers, indicating that poised enhancers establish promoter interactions prior to their activation.

We next investigated DNA features associated with stimulation-dependent interactions. Motif enrichment analysis revealed 60 TFBMs that are overrepresented in dynamic interactions relative to interactions that do not change after LPS stimulation (STAR Methods). Of these, 25 motifs are enriched exclusively in induced interactions, nine in downregulated interactions, and the remaining 26 motifs are enriched in connections that are both induced and downregulated. IRF, STAT, AP1, and SMAD family motifs are most strongly enriched in induced connections, while KLF motifs are most strongly enriched in connections that are lost (Figure 6C). For more than half (~54%) of TFs associated with induced interactions, the TF-encoding genes are also transcriptionally induced with LPS (Figure 6D), a ~7-fold enrichment of the expected number of induced TFs in a random sample of expressed TFs. This suggests a model where interactions that are signal dependent require signal-dependent TFs.

TFs are typically thought to act synergistically by physically interacting with one another to regulate gene expression (Junion et al., 2012; Garber et al., 2012). We, therefore, investigated whether certain TF pairs mediate the formation of chromatin interactions. To do so, we looked at pairs of motifs occurring within inducible interactions, such that one motif is within one interacting element and the other motif is within the interacting pair (Figure 6E, top). We identified 25 motif pairs that are enriched in induced interactions (p-adjusted < 0.001, binomial test) (Figure 6E). Interestingly, we find that the majority of the motif pairs are for members of the AP1 TF family (Figure 6E, black boxes). Further, five out of seven AP1 family members are transcriptionally induced within 1 h of LPS treatment. Thus, AP1 family heterodimers (Bejjani et al., 2019) may play a critical role in mediating context-specific inducible interactions. AP1 factors have been observed to mediate E-P interactions, since the same TF family

Figure 6. AP1 Factors Mediate LPS-Inducible E-P Interactions

(A) Fraction of interactions that are stable (gray), induced (dark pink), and downregulated (light pink) after LPS stimulation (top). Enrichment of induced, downregulated, and stable H3K27ac regions (left) and fraction of downregulated, early, late, and stable genes (right) in induced interactions (middle) and downregulated interactions (bottom). Significance was calculated using Wilcoxon Fisher's exact test.

(B) IGV plot of temporal changes in E-P interactions in the 215Kb *Nfkbiz* gene neighborhood depicting the . Tracks (top to bottom): H3K4me1 signal at 0h LPS stimulation, H3K27ac at 0h, 30 min, 1h and 2h after LPS stimulation, arcs: interaction score at 0h, 4h, and 24h after LPS stimulation.

(C) TFBMs enriched in induced interactions (top row) and in downregulated interactions (bottom row). The color of each box is proportional to the significance (Fisher's exact test).

(D) Clustering of temporal expression (TPM) of LPS responsive TFs, that bind sequence motifs enriched in LPS-inducible E-P interactions.

(E) TF pairs that are enriched in LPS-inducible interactions.

was previously reported to mediate long-range interaction in other cellular states (Vierbuchen et al., 2017; Phanstiel et al., 2017; Chavanas et al., 2008; Qiao et al., 2015).

DISCUSSION

While it is possible to estimate the effects of coding variants on protein function, they are relatively rare in human populations and are rarely associated with common diseases (Ludwig et al., 2019; Gallagher and Chen-Plotkin, 2018). Instead, most variants reside within noncoding regulatory elements (Nishizaki and Boyle, 2017; Zhu et al., 2017), and the effects of such variants are much harder to predict. Thus, building a framework to interpret the impact of non-coding risk variants is challenging because we do not understand how regulatory elements operate to regulate gene expression.

Several reports showed that incorporating 3D interactions is critical to linking regulatory elements to their target genes (Fulco et al., 2019; Moore et al., 2020). However, simply knowing the target gene(s) of regulatory elements is not sufficient to estimate the impact of their loss. One way to estimate the effect of the loss of a regulatory element is to build quantitative models that predict gene expression from the TFBMs they harbor.

Here, we showed that quantitative models of gene expression are better able to predict the effect of genetic variability and that these models greatly improve when incorporating 3D interaction. Ultimately, such quantitative models can be integrated into frameworks to prioritize disease-associated variants.

We were surprised to find that less than 50% of putative regulatory elements have detectable contacts with promoters. Furthermore, the poor performance of predictive models built using distance-based enhancer associations seems to be mainly due to spurious associations between enhancers and the genes they regulate. The improvement in gene expression prediction is mostly due to the higher specificity of SPRITE-estimated E-P interactions than from connections over large linear distances that exceed distance-based thresholds. Together, these findings strongly suggest that accessible DNA regions, even when they are active (marked by H3K27ac), do not necessarily exert regulatory function and that they may instead be poised to participate in other processes.

Our ability to better define regulatory interactions further helps to interpret the molecular conservation of regulatory elements. Our group and others have reported that a very small fraction of enhancers is shared across mammals (Villar et al., 2015; Donnard et al., 2018; Danko et al., 2018). We previously used motifs within enhancers that were shared between mouse and human DCs to train a random forest to classify genes into LPS-inducible versus LPS-nonresponsive genes. We showed that the classifier could be trained on data generated from mouse DCs and have almost identical performance when applied to human data (Donnard et al., 2018). Without available 3D proximity data, we assigned enhancers to target genes based on proximity. When including SPRITE-based 3D proximity data, we found that a higher fraction (75% versus 63%) of shared enhancers between mouse and human have interactions with promoters compared to enhancers that are mouse specific. This indicates that stable regulatory interactions are more often found in critical or less redundant regulatory elements, and therefore, these are more refractory to inactivation or replacement during evolution.

The single-molecule nature of SPRITE allowed us to directly explore multiway interacting complexes involving multiple enhancers and promoters as a mechanism to coordinate gene expression while also reducing transcriptional noise across a cell population. Further extensions and applications of these models to other systems will establish the practical importance of the models presented here.

Limitations of Study

Sequence variants can impact enhancer function at least two different ways. A variant may disrupt the binding of a single TF while keeping the enhancer activity intact (Claussnitzer et al., 2015). Alternatively, variants may render the enhancer inactive even if they do not directly affect a known binding site but instead affect DNA local properties that reduce the TF affinity (Grossman et al., 2017; Calderon et al., 2019), or they may prevent chromatin remodelers from keeping the enhancer accessible to TF binding (Verlaan et al., 2009). Given the large number of enhancers that are inactive in SPRET compared to C57, we focused our validation on these enhancers and the loss of all TFBMs in SPRET. There has been recent success in estimating the effect of sequencing variants on DNA accessibility (Maurano et al., 2015; Behera et al., 2018). As our understanding of how sequence variants affect DNA accessibility improves, a two-step approach would first predict variants that result in enhancer loss of activity followed by the approach presented here to estimate the effect of this loss on gene expression.

Current SPRITE protocols yield a high fraction of single unclustered reads (~40% in the data presented here), which are uninformative. Reducing or eliminating the number of these singletons could decrease sequencing costs.

TFs do not act alone. We did not model TFBM interactions and instead chose the simplicity and interpretability of linear models. Recent applications of deep learning techniques show promising possibilities in modeling TFBM interactions (Avsec et al., 2020). Extensions of our models that incorporate such terms may significantly improve prediction accuracy.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Cell culture and cell lines used
  - Mouse bone-marrow-derived dendritic cells
  - Mouse embryonic stem cells (V6.5)
- METHOD DETAILS
  - SPRITE
  - SIP method

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.09.005.

ACKNOWLEDGMENTS

We thank Athma Pai, Oliver Rando, John Harris, Kate Fitzgerald, and members of the Garber Lab for valuable discussions and comments. We thank Meenakshi Kagda and Jennifer Jou for help managing data submission to ENCODE. We also thank Inna-Marie Strazhnik for help with graphical abstract. This project was supported by NHGRI grants U01 HG007910 and R21CA236594 (M. Garber), NIH 4DN grants U01 DA040612 and U01 HL130007, and the NYSCF (M. Guttman). M. Guttman is a NYSCF-Robertson Investigator.

AUTHOR CONTRIBUTIONS

Conceptualization, Formal Analysis, and Data Curation, P.V. and M. Garber; Methodology, P.V., R.M., S.A.Q., K.G., P.M., M. Guttman, and M. Garber; Resources, M. Garber, and M. Guttman; Writing, P.V., M. Guttman, and M. Garber; Supervision and Funding Acquisition, M. Garber.

DECLARATION OF INTERESTS

S.A.Q. and M. Guttman are inventors on a provisional patent on the SPRITE method.

Received: February 22, 2020
Revised: June 4, 2020
Accepted: September 4, 2020
Published: September 28, 2020

REFERENCES

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., and Fropf, R. (2020). Base-resolution models of transcription factor binding reveal soft motif syntax. bioRxiv. https://doi.org/10.1101/737981.

Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res. *24*, 999–1011.

Behera, V., Evans, P., Face, C.J., Hamagami, N., Sankaranarayanan, L., Keller, C.A., Giardine, B., Tan, K., Hardison, R.C., Shi, J., and Blobel, G.A. (2018). Exploiting genetic variation to uncover rules of transcription factor binding and chromatin accessibility. Nat. Commun. *9*, 782.

Bejjani, F., Evanno, E., Zibara, K., Piechaczyk, M., and Jariel-Encontre, I. (2019). The AP-1 transcriptional complex: local switch or remote command? Biochim. Biophys. Acta *1872*, 11–23.

Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. Mol. Cell *76*, 473–484.e7.

Bornstein, C., Winter, D., Barnett-Itzhaki, Z., David, E., Kadri, S., Garber, M., and Amit, I. (2014). A negative feedback loop of transcription factors specifies alternative dendritic cell chromatin states. Mol. Cell *56*, 749–762.

Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat. Genet. *51*, 1494–1505.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Carty, M., Zamparo, L., Sahin, M., González, A., Pelossof, R., Elemento, O., and Leslie, C.S. (2017). An integrated model for detecting significant chromatin interactions from high-resolution Hi-C data. Nat. Commun. *8*, 15454.

Chavanas, S., Adoue, V., Méchin, M.-C., Ying, S., Dong, S., Duplan, H., Charveron, M., Takahara, H., Serre, G., and Simon, M. (2008). Long-range enhancer associated with chromatin looping allows AP-1 regulation of the peptidylarginine deiminase 3 gene in differentiated keratinocyte. PLoS ONE *3*, e3408.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO obesity variant circuitry and adipocyte browning in humans. N. Engl. J. Med. *373*, 895–907.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Danko, C.G., Choate, L.A., Marks, B.A., Rice, E.J., Wang, Z., Chu, T., Martins, A.L., Dukler, N., Coonrod, S.A., Tait Wojno, E.D., et al. (2018). Dynamic evolution of regulatory element ensembles in primate $CD4^+$ T cells. Nat. Ecol. Evol. *2*, 537–548.

Dekker, J. (2016). Mapping the 3D genome: aiming for consilience. Nat. Rev. Mol. Cell Biol. *17*, 741–742.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. *14*, 390–403.

Derr, A., Yang, C., Zilionis, R., Sergushichev, A., Blodgett, D.M., Redick, S., Bortell, R., et al. (2016). End Sequence Analysis Toolkit (ESAT) expands the extractable from single-cell RNA-seq data. Genome Res. *26*, 1397–1410.

Dickel, D.E., Zhu, Y., Nord, A.S., Wylie, J.N., Akiyama, J.A., Afzal, V., Plajzer-Frick, I., Kirkpatrick, A., Göttgens, B., Bruneau, B.G., et al. (2014). Function-based identification of mammalian enhancers using site-specific integration. Nat. Methods *11*, 566–571.

Donnard, E., Vangala, P., Afik, S., McCauley, S., Nowosielska, A., Kucukural, A., Tabak, B., Zhu, X., Diehl, W., McDonel, P., et al. (2018). Comparative analysis of immune cells reveals a conserved regulatory lexicon. Cell Syst. *6*, 381–394.e7.

Donnard, E., Shu, H., and Garber, M. (2020). Single cell transcriptomics reveals dysregulated cellular and molecular networks in a fragile X syndrome model. bioRxiv. https://doi.org/10.1101/2020.02.12.946780.

Durai, V., Bagadia, P., Granja, J.M., Satpathy, A.T., Kulkarni, D.H., Davidson, J.T., 4th, Wu, R., Patel, S.J., Iwata, A., Liu, T.T., et al. (2019). Cryptic activation of an Irf8 enhancer governs cDC1 fate specification. Nat. Immunol. *20*, 1161–1173.

Edelman, L.B., and Fraser, P. (2012). Transcription factories: genetic programming in three dimensions. Curr. Opin. Genet. Dev. *22*, 110–114.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. Cell *166*, 358–368.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341–1345.

Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS era: from association to function. Am. J. Hum. Genet. *102*, 717–730.

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol. Cell *47*, 810–822.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. Genome Biol. *17*, 215.

González, A.J., Setty, M., and Leslie, C.S. (2015). Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. Nat. Genet. *47*, 1249–1259.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Grossman, S.R., Zhang, X., Wang, L., Engreitz, J., Melnikov, A., Rogov, P., Tewhey, R., Isakova, A., Deplancke, B., Bernstein, B.E., et al. (2017). Systematic dissection of genomic features determining transcription factor binding and enhancer function. Proc. Natl. Acad. Sci. USA *114*, E1291–E1300.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., et al. (2007). Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome. Nature Genetics *39*, 311–318.

Hsieh, T.-H.S., Slobodyanyuk, E., Hansen, A.S., Cattoglio, C., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D landscape of transcription-linked mammalian chromatin folding. bioRxiv. https://doi.org/10.1101/638775.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E.M. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell *148*, 473–486.

Juric, I., Yu, M., Abnousi, A., Raviram, R., Fang, R., Zhao, Y., Zhang, Y., Qiu, Y., Yang, Y., Li, Y., et al. (2019). MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. PLoS Comput. Biol. *15*, e1006982.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

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.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaöz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. *17*, 691–707.

Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural details of mammalian chromosome architecture. Mol. Cell *78*, 554–565.e7.

Kuhn, M. (2008). Building predictive models in R using the Caret package. J. Stat. Softw. *28*, 1–26.

Kuhn, M. (2011). Caret: classification and regression training. R Package Version 4 (Astrophysics Source Code Library).

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

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. Cell *152*, 1021–1036.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

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.

Li, Y., Rivera, C.M., Ishii, H., Jin, F., Selvaraj, S., Lee, A.Y., Dixon, J.R., and Ren, B. (2014). CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. PLoS One *9*, e114485.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Liaw, Andy, Wiener, Matthew, and Others. (2002). Classification and regression by randomForest. R News *2*, 18–22.

Link, V.M., Duttke, S.H., Chun, H.B., Holtman, I.R., Westin, E., Hoeksema, M.A., Abe, Y., Skola, D., Romanoski, C.E., Tao, J., et al. (2018). Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. Cell *173*, 1796–1809.e17.

Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biol. *15*, 550.

Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X., et al. (2019). Transcriptional states and chromatin accessibility underlying human erythropoiesis. Cell Rep. *27*, 3228–3240.e7.

Maurano, M.T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R., and Stamatoyannopoulos, J.A. (2015). Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. Nat. Genet. *47*, 1393–1401.

Medzhitov, R., and Horng, T. (2009). Transcriptional control of the inflammatory response. Nat. Rev. Immunol. *9*, 692–703.

Moore, J.E., Pratt, H.E., Purcaro, M.J., and Weng, Z. (2020). A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. Genome Biol. *21*, 17.

Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and Super-Enhancers Have an Equivalent Regulatory Role in Embryonic Stem Cells through Regulation of Single or Multiple Genes. Genome Research *27*, 246–258.

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.

Nishizaki, S.S., and Boyle, A.P. (2017). Mining the unknown: assigning function to noncoding single nucleotide polymorphisms. Trends Genet. *33*, 34–45.

Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., et al. (2019). Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science *366*, 1134–1139.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Phanstiel, D.H., Van Bortle, K., Spacek, D., Hess, G.T., Shamim, M.S., Machol, I., Love, M.I., Aiden, E.L., Bassik, M.C., and Snyder, M.P. (2017). Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. Mol. Cell *67*, 1037–1048.e6.

Qiao, Y., Shiue, C.-N., Zhu, J., Zhuang, T., Jonsson, P., Wright, A.P.H., Zhao, C., and Dahlman-Wright, K. (2015). AP-1-mediated chromatin looping regulates ZEB2 transcription: new insights into TNFα-induced epithelial-mesenchymal transition in triple-negative breast cancer. Oncotarget *6*, 7804–7814.

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

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell *174*, 744–757.e24.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N., et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell *159*, 1698–1710.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Raghupathy, N., Choi, K., Vincent, M.J., Beane, G.L., Sheppard, K.S., Munger, S.C., Korstanje, R., Pardo-Manual de Villena, F., and Churchill, G.A. (2018). Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. Bioinformatics *34*, 2177–2184.

Rieder, D., Trajanoski, Z., and McNally, J.G. (2012). Transcription factories. Front. Genet. *3*, 221.

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

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

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.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437–455.

Setty, M., and Leslie, C.S. (2015). SeqGL identifies context-dependent binding signals in genome-wide regulatory element maps. PLoS Comput. Biol. *11*, e1004271.

Snetkova, V., and Skok, J.A. (2018). Chromatin folding and recombination. In Nuclear Architecture and Dynamics, C. Lavelle and J.-M. Victor, eds. (Academic Press).

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Verlaan, D.J., Berlivet, S., Hunninghake, G.M., Madore, A.-M., Larivière, M., Moussette, S., Grundberg, E., Kwan, T., Ouimet, M., Ge, B., et al. (2009). Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am. J. Hum. Genet. *85*, 377–393.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., and Greenberg, M.E. (2017). AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. Mol. Cell *68*, 1067–1082.e12.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Yang, R.Y., Quan, J., Sodaie, R., Aguet, F., and Segre, A.V. (2018). A systematic survey of human tissue-specific gene expression and splicing reveals new opportunities for therapeutic target identification and evaluation. bioRxiv. https://doi.org/10.1101/311563.

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.

Zhu, Y., Tazearslan, C., and Suh, Y. (2017). Challenges and progress in interpretation of non-coding genetic variants associated with human disease. Exp. Biol. Med. (Maywood) *242*, 1325–1334.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software or Algorithms		
Bowtie2	Langmead and Salzberg, 2012	V2.3.2
Samtools	Li et al., 2009	v0.1.19
DESeq2	Love et al., 2014	v1.10.1
Bedtools	Quinlan and Hall, 2010	V2.25.0
MACS2	Zhang et al., 2008	V2
IGVtools	Robinson et al., 2011	V2.3.31
RSEM	Li and Dewey, 2011	v1.3.0
SPRITE processing pipeline	Quinodoz et al., 2018	https://github.com/GuttmanLab/sprite-pipeline
Viewpoint centric analysis	This paper	https://github.com/garber-lab/SPRITE_ViewPoint_Analysis
Single cell RNA-seq alignment and processing	Donnard et al., 2020	https://github.com/garber-lab/inDrop_Processing
Reagents		
anti-H3K4me3 antibody	Millipore	Cat# 05-745R; RRID:AB_1587134
GM CSF	Miltenyi	130-095-735
Barcode and adapters	IDT	Quinodoz et al., 2018
Low Retention Pipette Tips	USA Scientific	1180-x
Reagent Reservoirs	Diversified Biotech	RESE-3000
PCR Strip Tubes	USA Scientific	1402-4700
DNA Lo-bind 1.5mL tube	Eppendorf	22431021
twin.tec PCR plate 96 well LoBind	Eppendorf	30129512
5mL Concial Tube	Olympus	24-285
15mL Conical Tube	Olympus	28-103
50mL Conical Tube	Olympus	28-106
DNA Clean and Concentrator Kit −5	Zymo	D4013
E-Gel EX 2% Agarose Gel	Invitrogen	G402002
BioAnalyzer High Sensitivity DNA	Agilent	5067-4627
Qubit 1x dsDNA HS Assay Kit	Invitrogen	Q33230
Bovine Serum Albumin	American Bio	AB0124300050
Glycine	Millipore Sigma	357002
Trypsin-EDTA (0.05%)	GIBCO	25300-054
Phosphate-Buffered Saline	Corning	29-031-CV
Disuccinimidyl glutarate (DSG); 50 mg bottle	Thermo Scientific	20593
16% Formaldehyde Ampule	Pierce	28906
1,2-Propanediol	Sigma Aldrich	134368
Dynabeads Protein A	Invitrogen	10002D
NHS Activated Magnetic Beads	Pierce	88826
Agencourt AMPure XP Magnetic Beads	Beckman Coulter	A63881
cOmplete, EDTA free Protease Inhibitor Cocktail	Roche	4693159001
Micrococcal Nuclease	NEB	M0247S
Proteinase K	NEB	P8107S
NEBNext End Repair	NEB	E6050S
NEBNext dA-Tailing Module	NEB	E6053S
Instant Sticky-end Ligase Master Mix	NEB	M0370L

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NEBNext Quick Ligation Reaction Buffer	NEB	B6058S
NEBNext® High-Fidelity 2X	NEB	M0541S
Deposited Data		
ChIP-seq (H3K27ac, H3K4me3, H3K4me1) in dendritic cells	Donnard et al., 2018	GEO: PRJNA356880
RNA-seq in dendritic cells	Donnard et al., 2018	GEO: PRJNA356880
ATAC-seq in dendritic cells	Donnard et al., 2018	GEO: PRJNA356880
SPRITE in dendritic cells	This paper	ENCODE: ENCSR027LIN, ENCODE: ENCSR528HTD, ENCODE: ENCSR789HIX
SIP in dendritic cells	This paper	GEO: GSE156379
SIP in mESC	This paper	GEO: GSE156379
Single cell RNA-seq in dendritic cells	This paper	ENCODE: ENCSR562QQA, ENCODE: ENCSR597YMV, ENCODE: ENCSR568FZM
ChIP-seq (H3K27ac) in macrophages cells	Link et al., 2018	GEO: GSE109965
RNA-seq in macrophages cells	Link et al., 2018	GEO: GSE109965

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Manuel Garber (Manuel.Garber@umassmed.edu).

Materials Availability

No unique reagents were generated in this study.

Data and Code Availability

All custom software used in this study is available on GitHub at https://github.com/garber-lab/. All sequencing data are available at NCBI GEO (GSE156379) and ENCODE (ENCSR562QQA, ENCSR597YMV, ENCSR568FZM, ENCSR027LIN, ENCSR528HTD, ENCSR789HIX).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All mice were housed in specific pathogen-free conditions in accordance with the Institutional Animal Care and Use Committee of the University of Massachusetts Medical School. C57BL/6 female mice were euthanized at 7-8 weeks of age to harvest bone marrow.

Cell culture and cell lines used

All cells were maintained at 37°C in 5% CO2 humidified incubators.

Mouse bone-marrow-derived dendritic cells

Mouse dendritic cells were derived from bone marrow harvested from 6-8 week old female C57 mice. Bone marrow was then dissociated into single cells and filtered through a 70um cell strainer. The cells were then incubated with the red blood cell lysis buffer for 5 minutes. To differentiate bone marrow to dendritic cells, bone marrow cells were plated at 200,000 cells/mL in untreated tissue culture plates. These cells were supplemented with media on day 2 and day 7. On day 5 cells were harvested and resuspended in fresh media. On day 8 all the floating cells were collected as mouse bone-marrow-derived dendritic cells. The media used for culturing and differentiating contains RPMI (GIBCO) supplemented with 10% heat-inactivated FBS (GIBCO), β-mercaptoethanol (50uM, GIBCO), MEM non-essential amino acids (1X, GIBCO), sodium pyruvate (1mM, GIBCO), and GM-CSF (20 ng/ml; Miltenyi).

Mouse embryonic stem cells (V6.5)

V6.5 mESCs, DGCR8 knockout mESCs, and Dicer knockout mESCs were cultured in Dulbecco's Modified Eagle's Medium (Thermo Scientific) supplemented with HEPES pH 7.0, 15% fetal bovine serum (FBS), 1000 U/ml leukemia inhibitory factor (Chemicon/Millipore), 0.1 mM l-glutamine, penicillin and streptomycin, and 0.11 mM β-mercaptoethanol.

METHOD DETAILS

SPRITE

SPRITE protocol was performed as previously described in Quinodoz et al. for mouse bone-marrow-derived dendritic cells stimulated for 0 hr, 4hrs and 24 hr. After the first barcoding step we pooled the complexes from the three time points and continited with the rest of the protocol. We also aliquoted our final crosslinked material into seven aliquots to contain approximately 200 million DNA molecules each (assuming no losses). Libraries were sequenced to yield about 1.2B paired-end reads on Illumina NextSeq 500.

SIP method

Crosslinking and Chromatin isolation

Cells were cross-linked in a single-cell suspension following steps listed in Quinodoz et al. Crosslinked cells were frozen in 5 or 10 million cell aliquots at −80C until used.

Crosslinked cell pellets (5-10 million cells) were lysed as described in Quinodoz et al. with few modifications. Specifically, cell pellets were first resuspended in 1.4mL per 10M cells of Nuclear Isolation Buffer A (50 mM HEPES pH 7.4, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 140 mM NaCl, 0.25% Triton-X, 0.5% NP-40, 10% Glycerol, 1X PIC) and incubated for 10 min on ice. Cells were pelleted at 850 g for 10 min at 4°C. The supernatant was removed, 1.4mL per 10M cells of Lysis Buffer B (10 mM HEPES pH 7.4, 1.5 mM EDTA, 1.5 mM EGTA, 200 mM NaCl, 1X Protease inhibitor cocktail) was added and incubated for 10 min on ice. Nuclei were obtained after pelleting and supernatant was removed (as above), and 550 μL of Lysis Buffer C_1 (20mM HEPES pH 7.5, 1.5mM EDTA, 100mM NaCl, 0.1% NaDOC, 0.1% Igepal CA360, 1x PIC) was added and incubated for 10 min on ice prior to sonication.

Chromatin digestion

After nuclear isolation, chromatin was sheared via sonication of the nuclear pellet using a Branson needle-tip sonicator (3 mm diameter (1/8" Doublestep), Branson Ultrasonics 101-148-063) at 4°C for a total of 1 min at 4-5 W (pulses of 0.7 s on, followed by 3.3 s off). DNA was further digested using 0.005 - 0.01 uL of Micrococcal Nuclease (MNase) (NEB M0247S) per 10 μL of the sonicated lysate (equivalent to ~200,000 cells), in Buffer C_1 supplemented with 5mM Cacl2 at 37°C for 20 min. Concentrations of MNase were optimized to obtain DNA fragments of mean size between 150-300 bp in length. MNase activity was quenched by adding 50mM EGTA and 0.1% SDS final concentration.

Immunoprecipitation

After digesting, the lysate is precleared with 60uL of protein A beads (at 10mg/mL concentration) per 1ug of chromatin by rotating at 4°C for at least 2 hours. Protein A beads were captured with a magnet and the lysate was transferred to a fresh tube. An equal volume of Adjustment buffer (80 mM HEPES pH 7.5, 200 mM NaCl, 1.5 mM EDTA, 50mM EGTA 1.8% Igepal CA630, 0.9 NaDOC, 0.1% SDS and 0.5mM PMSF) was added to the lysate. Next, the H3K4me3 antibody (Millipore Cat # 05-745R Lot # 288116) was added to the lysate (in this experiment a ratio of 1ug of antibody per 1ug of lysate was used) and rotated overnight at 4°C. The antibody-chromatin complexes were then captured using 25uL of blocked protein A beads (beads were blocked by rotating overnight at 4°C in 1xPBS, 0.5% BSA, 0.5% Tween 20) and washed in ChRIPA buffer (1X PBS, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Igepal CA630, 0.5% NaDOC, 0.1% SD). Chromatin is then eluted off the protein A beads using PBS-DEB buffer (1x PBS, 5 mM EDTA, 0.5% SDS, 10 mM DTT (added fresh)). 5% of the eluate was then reverse crosslinked using Proteinase K (NEB P8107S) in RNK-400 buffer (20mM Tris HCL pH8, 400mM NaCl, 10mM EDTA, 10mM EGTA, 0.5% Triton-X, 0.2% SDS) to be used for a ChIP-Seq library. The remaining eluate was then coupled to NHS beads after estimating the molarity as previously described in Quinodoz et al.

Split-pool ligation

Split-pool ligation was also performed as previously described in Quinodoz et al. except that we used 10X lower amount of adapters in each round (4.5uM of adapters instead of 45uM). Libraries were sequenced at 1.5B reads to yield 1.2B valid reads (reads that pass all the quality filters and have all 5 barcodes as described in Quinodoz et al.).

The detailed SIP protocol is available at https://www.umassmed.edu/garberlab/protocols/.

Comparison of SIP and HiChIP:

SIP and HiChIP employe similar enrichment strategies with antibodies against proteins of interest but both the assays differ in several details

Distance distribution:

We observed that applying our viewpoint centric analysis to SIP and HiChIP libraries, SIP captures longer-range interactions compared to HiChIP (Figure S2L). Also in our previous publication (Quinodoz et al., 2018), we showed that the SPRITE can capture longer-range interactions compared to Hi-C. Two reasons that can explain this difference is crosslinking and ligation.

Cross Linking:

SIP uses a similar cross linking strategy as SPRITE where the cells are cross-linked using 3% formaldehyde and DSG, while HiChIP uses 1% formaldehyde. The higher percentage of formaldehyde and additional cross linker such as DSG can capture interactions that are part of large structures that are observed in cells (Cho et al., 2018; Quinodoz et al., 2018).

Digestion:
In SIP we used MNase to digest chromatin while HiChIP uses restriction enzymes. We carefully titrated the concentration of MNase such that our average fragment size is around 150bp. The usage of MNase gives us higher resolution than restriction enzymes (Hsieh et al., 2020; Krietenstein et al., 2020).

Non-proximity ligation:
Ability of SPRITE based methods to capture chromatin interactions without using proximity ligation also enables us to detect all the interacting DNA molecules while proximity-based methods can only capture 2 of the interacting DNA molecules at a time. For example, for a cluster of 3 interacting regions population-based proximity ligation methods can capture any 2 of the interactions, this leads to low signal on each of the pairs (for fixed sequencing depth and cell number), while SPRITE based methods are able to capture all the possible pairwise interactions resulting in higher signal for each pair. This advantage also can be a reason for us to detect interactions at a higher resolution. From our data, we observe that about 30% of the clusters we sequenced have more than 2 DNA molecules interacting (Figure S1A).

***In situ*:**
HiChIP protocol is built on *in situ* HI-C, where crosslinking, digestion, and ligation are all done in the permeabilized nucleus reducing the number of false contacts, while the current version of SIP digestion and split pool barcoding is done on digested chromatin in tubes. This can lead to some proteins bound chromatin complexes to affinity bind in solution leading to a spurious larger complex that will get barcoded as one complex. To overcome this possibility and, in general, to filter out interactions that are not recurrent in the cell population, we require that each interaction is observed at least 5 times.

Many Abs:
HiChIP protocol has also been optimized for many different antibodies and cell-types while for SIP this has to be done yet.

QUANTIFICATION AND STATISTICAL ANALYSIS

SPRITE/SIP data processing and cluster generation

Barcode identification was done as previously described (Quinodoz et al., 2018). The genomic reads from both the SPRITE and the SIP libraries from BMDCs and mESCs were aligned to the mm10 genome using Bowtie2 (V2.3.2) with–local–trim5 11. The SAM files were converted to BAM using Samtools V1.4. BAM files are then filtered to keep only reads with all 5 barcodes identified and which are less than 2 mismatches to the reference genome. The files are then further filtered to keep all the reads with MAPQ > 10. Clusters of interacting DNA are generated from the filtered BAM files by identifying all the reads that have the same 5 barcodes using the custom script (get_clusters.py) which are available at https://github.com/GuttmanLab/sprite-pipeline. We obtained about 65% to 80% valid reads from SPRITE and SIP libraries respectively (Table S1).

Viewpoint Centric analysis and E-P interactions

To map the E-P interactions which are mostly short-range interactions, we used small (2-10) and medium (11-100) size SPRITE clusters (Quinodoz et al., 2018). SPRITE/SIP clusters are groups of DNA-molecules that have the same 5 barcodes. Pairwise contacts are calculated at 5kb resolution using the script get_sprite_contacts.py (https://github.com/GuttmanLab/sprite-pipeline) with *–down-weighting none* option. If a sprite cluster has two or more DNA molecules in the same 5kb genomic region (bin) it is counted only once. We then employed a viewpoint centric analysis where every promoter is set as a viewpoint and all interactions occurring with it are used. To determine bins with significant interactions with a viewpoint of interest (i.e., a promoter or an enhancer), computed by fitting a negative binomial (NB) generalized linear model (GLM) that models counts based on the distance decay of the signal (log distance), and the local SPRITE read coverage (log coverage). We note that GC content did not improve the GLM fit (r2 of 0.815 versus 0.811) and we therefore did not include it in the final GLM. We next calculated the p value for each bin pair as $p_{ij} = NB(X > x_{ij}/e_{ij})$ where e_{ij} is the expected frequency of the interaction from the background model. Similar to HiC-DC, MAPS, and Fit-Hi-C (Carty et al., 2017; Juric et al., 2019; Ay, Bailey, and Noble 2014) analysis, to better capture the null distribution, we re-fit our GLM based NB background model by excluding the top 15% scoring bins and recomputed p values with the new fit and adjusted for FDR using the Benjamini-Hochberg procedure. Using the view point centric approach we considered the 2MB region around each promoter (±2Mb) and searched for over-represented interactions with all the active promoters. This procedure was implemented as an R script available at https://github.com/garber-lab/SPRITE_ViewPoint_Analysis. The Interaction score between an viewpoint and a distal element is then calculated by dividing the observed number of SIP/SPRITE clusters containing the viewpoint and the distal element by the number of expected SIP/SPRITE clusters containing the distal element and the viewpoint under the generalized linear model.

The enriched interactions called using SPRITE data are highly reproducible in SIP data as well have highly correlated interaction frequency ($\rho >= 0.66$) (Figures 2C and S2D). The enriched interactions are also highly reproducible across time points in SIP data with > 90% being called at any two time points and also have high correlation in interaction frequency ($\rho >= 0.77$) (Figures S2E–S2J).

To call interactions we first merged valid reads across all time points and also at each time point separately using the model described above. Our calls per time point basically yielded a subset of interactions (86%) we called when all the data was merged.

Null model for coregulation of promoter-promoter pairs

To assess if the Promoter-Promoter pairs mapped using SPRITE/SIP are coregulated in response to LPS we generated null model by choosing noninteracting promoter pairs such that the expression level of the two genes is in the same tertile as that of the genes with interacting promoters and the distance between two promoters of a pair is similar to the distance between the promoters in the interacting promoter pairs.

Motif instances

We used class A motifs from the mouse Hocomoco v11 database. All instances were detected across all ATAC peaks (promoters and enhancers) using Fimo (Grant et el., 2011), with a q-value threshold of 1e-4.

Random Forest

In order to account for different numbers of enhancers that are associated with genes, we classified genes into tertiles of complexity as previously described (Donnard et al., 2018). Within each tertile we filtered out genes that were expressed less than 50 TPMs and used genes in high complexity tertile for this analysis.

All model training and evaluation were done in R 3.5.1, using the caret (v6.0.77) (Kuhn, 2008) and randomForest (v4.6.12) (Liaw et al., 2002) packages. For each feature set, we evaluated the accuracy of the model on the training data with 10-fold cross-validation. For each one of the training data in the cross-validation hyperparameters tuning was performed using 10-fold inner cross-validation with the "train" command, using the following parameters: tuneLength = 20, metric = "ROC."

Linear regression

Peak dynamics

All accessible peaks were categorized into early, late, down or non-changing as previously described (Elisa Donnard et al., 2018).

Motif calling and scoring

Motif calling and scoring was performed for each ATAC temporal group. Because each accessible site can be of variable length (after merging from all time points), we only considered 200 bp windows centered on the summit. For the background set of regions we randomly shuffled each 200bp window within 10,000 bp on each side to pick a random region of 200 bp that matched the GC content of the test set. We then used SeqGL (Setty and Leslie 2015) to identify motifs that are enriched in accessible regions within regulatory elements of early, late, downregulated and non-changing ATAC groups, with 250 groups and 30,000 features, similar to the parameters used to analyze DNase-seq data in the original SeqGL publication.

Gene selection

Models that predict transcriptional changes:

In order to account for different numbers of enhancers that are associated with genes, we classified genes into tertiles of complexity as previously described (Donnard et al., 2018). Within each tertile we filtered out genes that were expressed less than 50 TPMs and fitted an elastic net regression model for genes in high-complexity tertile (detailed below) to predict fold change at 1 hr or 6 hr after LPS treatment.

Interacting promoters as distal regulatory elements:

To determine the importance of promoters of interacting genes as distal regulatory elements we measured the change in above models performance using a feature matrix that excludes the scores of TFBMs from the interacting promoters (Figure 3G). We were able to test this only at 6hrs after LPS stimulation and not at 1hr post LPS as there were only few promoter-promoter interactions of expressed promoters at this time point.

Models that predict steady state expression:

To predict expression at baseline i.e., without any LPS stimulation we focused on genes that are not ubiquitously expressed. In this regard we excluded all genes with tissue specificity score less than 3 from the list published in (Yang et al., 2018) and that are expressed at less than 50 TPMs and fitted an elastic net regression model for genes in high and medium complexity tertile to predict gene expression.

ElasticNet regression

We used the sum of peak scores reported by SeqGL across each peak category per k-mer group as features for elastic net regression. We then used this feature matrix to predict the fold change of genes at early time point (1 hr), late time point (6 hr) or TPM at steady state. The models were trained using 70% of the data and evaluated using the remaining 30% using the caret (v6.0.77) (Kuhn, 2011)and glmnet. The hyperparameters tuning was performed using 10-fold inner cross-validation with the "train" command, using the following parameters: method = "repeatedcv," number = 10, allowParallel = T, repeats = 100, tuneGrid = expand.grid(.alpha = seq(0, 1, by = 0.05), .lambda = seq(0,10, by = 0.5)), metric = "RMSE."

Effect of interaction Distance on model performance

When performing viewpoint centric analysis, SPRITE, as well as HiChIP, the signal is found across a large area. An enrichment statistic is used to determine which loci are in physical proximity to the viewpoint. Although our models dramatically improved when

restricting TFBMs to enriched loci, it could also be that the model could have been further improved by using the area with SPRITE signal which varies greatly from promoter to promoter. We also tested models in which we assigned enhancers to genes up to the neighborhood window defined by the SPRITE data. The model performance did not improve with these newly defined neighborhood windows (Figure S1E). This strongly suggests that the SPRITE signal is not just defining an interaction window. It is also possible that interaction data derived from SPRITE is useful for defining longer range target genes than the short-range. To test this hypothesis, we also tested the difference between SPRITE based target gene assignments to distance-based assignments for enhancers that are within the 50kb window of the target gene. We considered motifs within the 50kb window for either distance-based assignment or SPRITE based assignment. The model performance is still better for target genes assigned using SPRITE compared to distance-based (Figure S1F). It is interesting to note that this shorter window assignment has lower performance than the 2MB (0.66 versus 0.79 ROC-AUC). This is probably because only about 30% of the interactions assigned using SPRITE are within 50kb of the target gene.

Dependence on the kinetics of DNA accessibility and epigenomic features on model performance

Given that the model performance improved drastically when using the interactions defined by the SPRITE signal, we next tested which aspects of our feature matrix such as the DNA accessibility kinetics, intensity of accessibility measured from ATAC-seq signal, the intensity of H3K27ac-signal are key contributors for the improvement.

First, we removed the dependence on the timing of DNA accessibility by constructing a new feature matrix that uses the sum of the TFBMs scores across all the interacting regulatory elements for a given gene instead of adding the TFBMs scores across all regulatory elements with similar DNA accessibility kinetics. The model built using this feature matrix reduced the predictive model performance (Figure S3C). This suggests that the kinetics of DNA accessibility are one of the important features for the model performance.

Next, we wanted to test to what extent the local epigenomic signals measured by ATAC signal intensity and the H3K27ac signal intensity contribute to the predictive models. For this, we multiplied the TFBM scores with the z-score normalized signal of the ATAC peak or the H3K27ac peak in which they occur. We constructed feature matrices as before, by the adding of these scores across all regulatory elements with similar DNA accessibility kinetics. Predictive models using these feature matrices have no change in their model performance compared to the models that use just the TFBM scores itself (Figure S3D). This suggests that the local epigenomic features are not directly impacting the model performance but rather may be affecting the establishment of E-P interactions as shown by a recent study (Fulco et al., 2019).

Reproducibility of SIP/SPRITE interactions:

Our SIP and SPRITE data provide two replicates. We performed SIP and SPRITE on cells derived from different mice and observed highly comparable results across each.

The enriched interactions are highly replicable across protocols (SIP versus SPRITE) in terms of the number of interactions called (i.e 74% of interactions identified in SPRITE are also called in SIP) and also the correlation of interaction frequency ($\rho = 0.66$) (Figures 2C and S2D). Across time points the number of promoter enhancer interactions called from SIP data is overwhelmingly similar with over 90% of contacts identified at any time point being also identified at other time points (Figures S2E–S2K). Across time points the number of promoter enhancer interactions called from SIP data is overwhelmingly similar with over 90% of contacts identified at any time point being also identified at other time points (Figures 6A and S2E–S2K). Further for the contacts identified at any two time-points have a high correlation of interaction frequency ($\rho >= 0.77$).

SPRET data processing

We reprocessed previously published data for RNaseq and H3K27ac ChIP-Seq (Link et al., 2018) from BMDMs derived from SPRET and C57.

The custom genome was generated for SPRET by incorporating alleles reported in the VCF files from the Mouse genome project (version v5) (Keane et al., 2011) using EMASE (Raghupathy et al., 2018). VCF files were filtered to keep the SNPs that passed VCF quality control and when an SNP and an indel overlapped we kept the variant with the best quality. We generated reference files using bowtie2 using the command *bowtie2-build* with default parameters.

The custom transcriptome was generated for SPRET by using liftover to remap the mm10 GTF file coordinates onto the SPRET genome created. We next generated reference files using RSEM (v1.3.0) using the command *rsem-prepare-reference* with the option–polyA-length 67.

RNaseq

We used RSEM (v1.3.0) to estimate gene expression in Transcripts per Million (TPM) with parameters *–no-bam-output–bam*. RSEM was configured to use bowtie (v1.2.2). Genes with more than 15 TPMs at any time point were considered as expressed.

We used counts per gene to identify differentially expressed genes by at least absolute $\log_2$ fold change of 0.5 between BMDMs derived from C57 at corresponding time point whose change in expression was significant (p-adjusted < 0.05) according to the package DESeq2 (v1.10.1) (Love et al., 2014) in R (v3.5.1). Due to the large transcriptional changes observed in this system, we turned off the fold change shrinkage in DESeq2 with *betaPrior = FALSE* and we added a pseudo count of 32 to all time points to avoid spurious

large fold change estimates from lowly abundant genes. Genes were then classified based on their response to KLA stimulation in each species (induced, downregulated or non-responsive).

ChIPseq

We used bowtie2 to map the reads to the custom SPRET genome generated with default parameters. We used macs2 to call peaks with–extsize 300 option. We next merged all the peaks within 200bp of each other using bedtools *slop* and bedtools *merge*. We next built a master list of peaks by merging peaks from all time points (0h,1h,6h post-KLA). The peaks were then quantified by calculating the coverage per bp at each time point and normalizing for the size of the library using DESeq2 *estimatesizefactors*. Then we filtered the peaks to keep all the peaks that have at least 10 normalized counts.

To get the list of peaks that are conserved between BMDMs derived from C57 and SPRET we used liftover to map the coordinates of SPRET onto C57 genome and all the peaks that overlapped another H3K27ac peak in C57 are considered conserved. The peaks that did not overlap the C57 H3K27ac peaks but were mappable were considered non-conserved.

Next, we compared the signal strength of conserved peaks at each time point between BMDMs derived from C57 and SPRET mice. The peaks that have at least a log2 fold change of 2 between the two species we considered differential.

Enrichment of differential enhancers

The enhancer assignment for Figure 5A: Enhancers are assigned to the closest expressed gene within either 300kb, 150kb or 50kb or enhancers are assigned based on the interaction radius identified by SPRITE/SIP data for that given gene. Enrichment is calculated using contingency table shown below

	Connected	Not Connected
Differential activity	**5449**	**3340**
Similar activity	**3441**	**3219**

ElasticNet Regression

We recomputed the motif scores for the motifs identified in different ATAC-seq peak categories defined in C57 BMDCs. Given the overall high conservation of active regulatory elements and gene expression profiles, we assumed that the enhancer promoter interactions are also conserved between BMDCs and BMDMs. We predicted the fold change at 1h and 6h using our model that was trained in C57, when we applied to SPRET we still used motif scores computed from C57 sequences setting to zero any motif lying within a region that was inactive in SPRET. To further address any discrepancies that may arise if SPRET sequences are used, we also applied our model after computing motifs scores using SPRET sequences. The results are almost identical ($\rho = 0.61$ versus $\rho = 0.59$).

Temporal changes in pairwise interactions

The interactions counts for all the pairwise interactions are split based on the barcodes of 0, 4 and 24h LPS stimulation. For each interaction, we performed Fisher's exact test to compare the counts at 4h or 24h against 0h to assess if the interaction is induced or downregulated or stable. We then considered any interaction to be significantly changing if the p values obtained from Fisher's exact test is ≤ 0.05.

Defining poised Enhancers

To define poised enhancer regions we used H3K4me1 and H3K27ac data that was previously published (Garber et al., 2012). The data was processed to align to mm10 genome and ChIP-Seq peaks for both H3K27ac and H3K4me1 were called using MACS2 as described previously in (Donnard et al., 2018). ChIP-Seq peaks were filtered to retain only the peaks that are two-fold enriched over input. Next, poised enhancers are defined as regions having H3K4me1 but not H3K27ac mark in unstimulated DCs, but acquire H3K27ac after stimulation with LPS. To do this we first generated a table of H3K4me1 regions that do not overlap H3K27ac regions in unstimulated DCs. Next we intersected these peaks with H3k27ac peaks from stimulated DCs. All the intersections were done using *bedtools intersect*.

Motif enrichment in dynamic interactions

We performed motif enrichment in induced interactions and downregulated interactions using Fisher's Exact test on the motif instances called using FIMO (section Motif Instances in STAR Methods) . The p values obtained were corrected for multiple hypothesis testing using FDR in R (v3.3.5).

3D Cis-regulatory Modules

We generated all the possible combinations of motifs doublets from the significant motifs in dynamic interactions. For each pair of motifs we conditioned that they have instances in each interacting pair. We then used the FDR corrected binomial p values test to

assess their co-occurrences. For example, to compute the co-occurrences of motif x with motif y, we define the number of co-occurrences of this pair such that motif x occurs in pair1 and motif y occurs in pair2. The background probability set to the product of the probability of motif x and motif y.

Biclustering for finding higher-order interactions

On average there are 6 cis-elements interacting per promoter, which makes the total number of possible combinations, assuming all interactions are possible, to be $2^6 = 64$. Given this large number of possible configurations, we reasoned that a clustering approach would reveal complex interactions whenever they are frequent in the cell population. For each promoter, we built a boolean interaction matrix where every SPRITE/SIP cluster is a row and genomic bins around the promoter are columns. The value of a cluster and a genomic bin is 0 or 1 depending on whether or not the bin overlaps a read in the cluster. We biclustered this matrix using the biclust (v2.0.1) package in R (v3.5.1) such that every bicluster should have at least 2 distal elements and must be represented in 5 unique SPRITE/SIP clusters. To assess the random chance of getting these biclusters we permuted the values within each row, so that the total number of 1 s and 0 s remained the same. of the interaction matrix and performed biclustering (as described above) on the permuted matrix. We repeated this 1000 times and to get the p value we used the permutation test.

Single-cell sequencing data

Single-cell RNA-Seq data was collected using in house built inDROP (Klein et al., 2015) systems for BMDCs stimulated with LPS for 0,1 and 4 hr.

Library generation

First, cells were resuspended in 15% OptiPrep in 1x PBS at a density of ~80,000 cells/mL then run through the microfluidic chip of the instrument. Along with cells, beads, RT mix (containing SuperScript III Reverse Transcriptase from Invitrogen 18080093) and a carrier oil (HFE 7500) were run through the microfluidic chip to capture single cells with single beads in an aqueous droplet. Following collection, the oligos were released from the beads by a 7 minute UV light exposure, then 2 hour for reverse transcription at 55C, 15-minute heat kill at 70C, followed by emulsion breaking. Next, the samples are cleaned before second strand synthesis, IVT, and RT using random hexamer primers. Then final library amplification was performed while incorporating sequencing adapters.

Alignment and processing

To generate fastq files we used bcl2fastq with–use-bases-mask y58n*,y*,I*,y16n* parameters. Next we extracted all the reads that contained a valid cell barcode and the unique molecular identifier (UMI) sequence had no Ns. The cell barcode and the UMI is appended to the read 1 header using custom scripts. The read 1 fastq files are then aligned to the mouse genome mm10 using TopHat (v2.1.1) (Kim et al., 2013) with default parameters. The bam files are then filtered to keep the cell barcodes that contained > = 3000 reads. The filtered bam files are then processed with ESAT single-cell analysis mode (-scPrep) (Derr et al., 2016) to generate a matrix of counts per gene per cell, using the UCSC gtf file for mouse genome built mm10. At this stage, we also extended the 3′ annotations of the transcriptome file upto 1000 bases, discarded all the multi-mapped reads (-wExt 1000, -task score3p, -multimap ignore). Finally, we merged all the UMIs that have one count and are one hamming distance away from another UMI that has two or more counts. All the scripts used for this processing are available through https://github.com/garber-lab/inDrop_Processing.

Variability of gene expression across cells

The coefficient of variation (CV) is computed for all the genes at every time point. Genes are filtered to keep the ones that are detected in at least 20 cells at any time point. We then compared the CV for genes with higher-order interactions and genes without grouping genes by maximum expression and number of enhancers.