Systematic dissection of genomic features determining transcription factor binding and enhancer function

Sharon R. Grossman[a,b,c], Xiaolan Zhang[a], Li Wang[a], Jesse Engreitz[a,d], Alexandre Melnikov[a], Peter Rogov[a], Ryan Tewhey[a,e,f], Alina Isakova[g,h], Bart Deplancke[g,h], Bradley E. Bernstein[a,i,j], Tarjei S. Mikkelsen[a,k,l], and Eric S. Lander[a,b,m,1]

[a]Broad Institute, Cambridge, MA 02142; [b]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Health Sciences and Technology, Harvard Medical School, Boston, MA 02215; [d]Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA 02138; [f]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; [g]Institute of Bioengineering, CH-1015 Lausanne, Switzerland; [h]Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland; [i]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; [j]Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; [k]Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138; [l]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138; and [m]Department of Systems Biology, Harvard Medical School, Boston, MA 02215

Contributed by Eric S. Lander, December 25, 2016 (sent for review December 7, 2016; reviewed by Eran Segal and Itai Yanai)

Enhancers regulate gene expression through the binding of sequence-specific transcription factors (TFs) to cognate motifs. Various features influence TF binding and enhancer function—including the chromatin state of the genomic locus, the affinities of the binding site, the activity of the bound TFs, and interactions among TFs. However, the precise nature and relative contributions of these features remain unclear. Here, we used massively parallel reporter assays (MPRAs) involving 32,115 natural and synthetic enhancers, together with high-throughput in vivo binding assays, to systematically dissect the contribution of each of these features to the binding and activity of genomic regulatory elements that contain motifs for PPARγ, a TF that serves as a key regulator of adipogenesis. We show that distinct sets of features govern PPARγ binding vs. enhancer activity. PPARγ binding is largely governed by the affinity of the specific motif site and higher-order features of the larger genomic locus, such as chromatin accessibility. In contrast, the enhancer activity of PPARγ binding sites depends on varying contributions from dozens of TFs in the immediate vicinity, including interactions between combinations of these TFs. Different pairs of motifs follow different interaction rules, including subadditive, additive, and superadditive interactions among specific classes of TFs, with both spatially constrained and flexible grammars. Our results provide a paradigm for the systematic characterization of the genomic features underlying regulatory elements, applicable to the design of synthetic regulatory elements or the interpretation of human genetic variation.

gene regulation | transcription factor binding | systems biology

Regulatory sequences in DNA encode the information necessary to establish precise patterns of gene expression across cell types and conditions. Although thousands of megabases of potential regulatory sequences have been identified (1, 2), deciphering the regulatory code—that is, being able to recognize and design regulatory sequences corresponding to particular expression patterns based on the underlying sequence—remains a major challenge in biology.

Gene expression is orchestrated by transcription factors (TFs), which bind to cognate binding sites (with characteristic sequence motifs) within regulatory sequences and recruit or modify components of the transcriptional machinery (3, 4). In the past decade, experimental advances have enabled characterization of the binding motifs for hundreds of TFs in vitro (5–9), mapping of the genome-wide binding sites of TFs in vivo (10–14), and functional characterization of the enhancer activity of thousands of genomic sequences (15–19). Comparisons between these experiments, however, have revealed that only a small fraction of the potential TF-binding sites (TFBSs) in eukaryotic genomes are actually occupied by TFs in any given cell type, and that these sites vary substantially across cell types and conditions (3, 19–21). Moreover, only a subset (∼25–50%) of bound TFBSs can drive transcription in reporter assays (17–19, 22). Understanding the regulatory code involves being able to explain the sequence features and mechanisms underlying the ability of enhancers to bind specific TFs and to drive transcription in a given cellular context.

Several features could influence the TF binding and enhancer activity of specific motif sites in vivo. First, variation in the binding and activity of motif sites may reflect differences in the affinity of binding sites, due to the motif sequence (23–25), latent motif preferences induced by cofactors (26, 27), or additional specificity determinants outside the core motifs, such as A/T-rich stretches (28–30). Second, TF access to motif sites may be governed by nucleosomes or the larger chromatin landscape (31–37). Third, additional TF motifs in the surrounding sequence could influence binding directly through protein–protein interactions, or indirectly through cooperative nucleosome displacement (38) or changes to the chromosome structure (39–41). Fourth, TFs at nearby sites could also contribute to transcriptional activation, either independently or through particular combinations of bound TFs acting in concert (e.g., by promoting

Significance

A central question in biology is how transcription factors (TFs) recognize specific binding sites in enhancers and regulate gene expression. In general, only a fraction of potential binding sites for TFs are occupied in a particular cell type. TF affinity for a motif site, local interactions among TFs, and larger-scale chromatin accessibility can influence binding, although the relative contributions of these factors is unclear. Moreover, little is known about how specific combinations of TFs control quantitative gene expression once bound. Here, we use large-scale synthetic biology approaches to explore the features that govern TF binding vs. enhancer activity. This approach provides a paradigm for systematic study of key regulatory sequences within enhancers and how they interact to influence gene expression.

Author contributions: S.R.G., B.D., B.E.B., T.S.M., and E.S.L. designed research; S.R.G., X.Z., L.W., and A.I. performed research; A.M., P.R., R.T., B.D., and T.S.M. contributed new reagents/analytic tools; S.R.G., J.E., and A.I. analyzed data; and S.R.G. and E.S.L. wrote the paper.

Reviewers: E.S., Weizmann Institute of Science; and I.Y., New York University School of Medicine.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE84888).

[1]To whom correspondence should be addressed. Email: lander@broadinstitute.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1621150114/-/DCSupplemental.

better contacts with cofactors and general transcription factors) or in opposition (e.g., by inhibiting each other by disrupting these contacts) (42–44).

The extent to which TF binding and transcriptional activity of an enhancer are controlled by the same or separate factors is generally unclear. In particular, systematically identifying and characterizing TF interactions has proven difficult. Key questions include whether TFs fall into distinct functional groups and what constraints on motif positioning and orientation exist for various interactions.

To address these questions, we focused on peroxisome proliferator-activated receptor γ (PPARγ)-response elements (PPREs) as a model set of regulatory sequences. PPARγ is a nuclear receptor that binds in cooperative fashion as a heterodimer with retinoid X receptor (RXR) to the canonical nuclear receptor direct repeat 1 (DR1) motif (45). It functions as a core regulator in adipocytes, localizing to PPREs during differentiation and primarily acting as a transcriptional activator (46). In mouse adipocytes, PPARγ is bound to only ∼1 in 200 genomic instances of this motif—and, even in regions of open chromatin, to only ∼1 in 16 motif instances. Furthermore, only ∼15% of the genes closest to PPARγ binding sites are up-regulated during adipogenesis (47).

We used massively parallel reporter assays (MPRAs), together with high-throughput in vitro and in vivo binding assays, to systematically manipulate motif affinity, cooperative interactions, and chromatin accessibility across thousands of PPREs and measure the effect on PPARγ binding and expression. MPRA involves testing huge collections of short regulatory sequences (≤150 bp) in parallel by coupling each to a transcription unit containing a matched DNA barcode (48) (Fig. 1 *A* and *B*). In total, we collected data on 32,115 regulatory sequences.

We show that PPARγ binding depends on the affinity of the PPARγ motif and on the larger chromatin landscape, but not significantly on the sequence in the immediate vicinity of the PPARγ binding site, indicating cooperative protein–protein binding interactions are relatively scarce. In contrast, enhancer activity strongly depends on the motifs in the immediate vicinity, particularly a core set of 20–30 additional TF motifs. Notably, we show that, in addition to the individual contributions of these motifs, particular combinations are also an important determinant of expression. We systematically identify and functionally test these interactions and find diverse interaction rules for different pairs of TFs, including additive, inhibitory, and synergistic interactions with varying constraints on motif positioning. Notably, we found consistent interactions between families of TFs, suggesting they may influence transcriptional activation through distinct mechanisms. Together, these experiments present a comprehensive approach to dissect the sequence grammar that determines TF binding and enhancer function. Applying this approach to a broad range of enhancers and cell types will help to determine the prevalence and generality of these rules, and potentially yield universal models to predict expression from regulatory sequences across diverse cellular contexts.

Results

In Vivo PPARγ Binding on Plasmids Is Governed by the Core PPARγ Motif. The mouse genome contains ∼1.5 million PPARγ motif sites, defined based on the canonical 16-base PPARγ/RXR DR1 motif (*Materials and Methods*). Of these, only a small minority—between 5,000 and 10,000 sites—are actually occupied by PPARγ in adipocytes (46, 47, 49). In principle, the binding at specific sites in vivo might depend on the following: (*i*) latent properties of the motif instance, not captured by the consensus sequence (30); (*ii*) cooperative binding in the immediate vicinity by other TFs expressed in the cell type (either directly through protein–protein interactions or indirectly through cooperative competition with nucleosomes) (50); and (*iii*) differences in the accessibility of the sites due to the chromatin landscape (36, 51, 52).

To investigate the first two possibilities (latent motif affinity and cooperative binding), we used a pooled plasmid-based reporter system to explore PPARγ binding in vivo to genomic motif sites outside of their native chromatin context (Fig. 1 *A* and *B*). We randomly chose 750 of the 6,835 sites we previously identified in ChIP-sequencing (ChIP-seq) experiments ("bound genomic sites") (47), and, for each, we chose a matched site that contained an identical 16-bp PPARγ motif but was not bound by PPARγ in adipocytes ("unbound genomic sites"). We synthesized an oligonucleotide pool (pool 1) containing 145 bp centered on the PPARγ motif from each of these 1,500 sites, as well as 1,500 control sequences, one for each site in which the core PPARγ motif was disrupted by swapping A↔T and G↔C (Fig. 1*C*). We cloned these 3,000 sequences ("candidate enhancers") into plasmids containing a minimal promoter and an ORF containing a unique barcode that identifies its specific upstream enhancer (48).

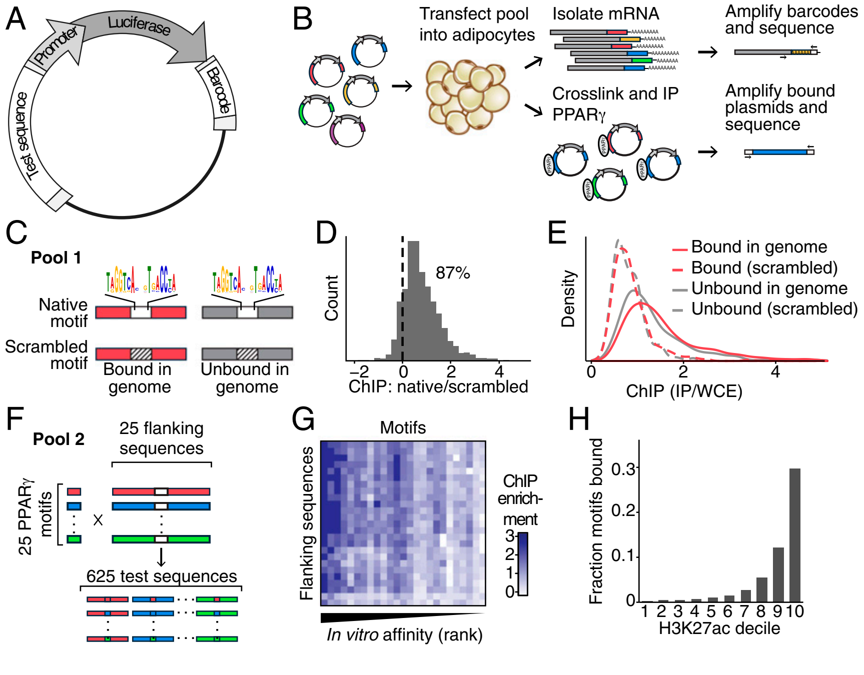

Fig. 1. In vitro PPARγ binding is determined by core motif affinity. (*A* and *B*) Overview of pooled reporter system. (*A*) Candidate sequences were cloned into plasmids upstream of a minimal promoter and barcoded *luc2* ORF. (*B*) Plasmid pools were transfected into adipocytes and assayed for PPARγ binding by ChIP-seq (*Lower*) and for enhancer activity by RNA-seq (*Upper*). (*C*) Schematic of pool 1, containing 750 bound genomic PPARγ motif sites, 750 unbound genomic PPARγ motif sites, and these 1,500 sites with the core PPARγ motif disrupted. (*D*) Log_2-ratio of ChIP enrichment for each genomic sequence with an intact and disrupted central PPARγ motif. (*E*) PPARγ ChIP enrichments for bound and unbound genomic sequences with intact and disrupted core PPARγ motifs. (*F*) Schematic of pool 2. The core PPARγ motif from 25 bound genomic sites was swapped into each of the other 24 flanking sequences, yielding a matrix of 625 enhancer constructs. (*G*) ChIP enrichment for each core motif (columns) and flanking sequence (rows) in pool 2. Core motifs were arranged by affinity measured by MITOMI (*Materials and Methods*). (*H*) Fraction of genomic PPARγ motif sites bound by PPARγ, conditional on the H3K27ac ChIP enrichment score (47).

We transfected this plasmid pool into mouse 3T3-L1 adipocytes 7 d postdifferentiation; grew the cells for 16 h; and measured PPARγ binding by performing ChIP-seq and calculating the relative enrichment of reads corresponding to each sequence. Measurements of relative binding activity were highly reproducible between two biological replicates ($r = 0.93$; *SI Appendix*, Fig. S1*A*). As a control, we also examined PPARγ binding across the genome from the same experiment and confirmed that the observed binding sites were consistent with those identified in our previous ChIP-seq experiments with PPARγ (*SI Appendix*, Fig. S1 *B*–*D*).

For candidate enhancers corresponding to both bound and unbound genomic sites, disrupting the PPARγ motif significantly reduced binding ($P_{\mathrm{Wilcox}} < 2.2 \times 10^{-16}$; Fig. 1*D*). The native sequence showed stronger binding than the disrupted control in 87% of cases. For the other 13%, the difference in binding to the native and disrupted motif sites was negligible [less than the technical variance between replicates (*SI Appendix*, Fig. S2*A*)]. In one-half of these cases, these sequences contained a second PPARγ motif site. In the remaining cases, the native sequence exhibited only weak binding and tended to contain less robust matches of the PPARγ motif, suggesting they have lower affinity for PPARγ (*SI Appendix*, Fig. S2 *B* and *C*).

To our surprise, we found little difference in PPARγ binding between candidate enhancers corresponding to bound genomic sites vs. those corresponding to unbound genomic sites (Fig. 1*E*). More precisely, the bound genomic sites showed slightly higher average binding, but this difference is largely explained by the fact that the sequences flanking the bound sites have a slightly higher average number of PPARγ motif sites (average of 0.5 in bound vs. 0.2 in unbound). For bound and unbound genomic sites with the same number of PPARγ motif sites, the difference is no longer statistically significant ($P_{F\ \mathrm{test}} = 0.67$; *SI Appendix*, Fig. S2*D*).

The fact that the DNA immediately surrounding bound and unbound sites appears to have equivalent ability to bind PPARγ when reintroduced on plasmids into adipocytes suggests that PPARγ binding depends primarily on the core PPARγ motif site and is not significantly influenced by elements or interactions in the immediate surroundings. To test this hypothesis, we selected 25 bound genomic sites from the original pool with a range of predicted motif strengths, and created a second plasmid pool (pool 2) by swapping the central PPARγ motif site from each sequence into the other 24 flanking sequences, generating a “matrix” of 625 candidate enhancers (Fig. 1*F*). Consistent with our hypothesis, in vivo binding to the plasmid pool strongly depended on the precise sequence of the core PPARγ motif site rather than on the flanking sequence (40% vs. 6% of variance explained; Fig. 1*G*).

For each central PPARγ motif sequence, we directly measured the in vitro binding affinity, using a microfluidic device that assays association and dissociation of fluorescently labeled DNA oligonucleotides with PPARγ protein immobilized on the surface of the device (*Materials and Methods*) (53). The binding observed in our cellular ChIP assay was well predicted by the in vitro affinity measurements of the motifs (Fig. 1*G* and *SI Appendix*, Fig. S3*A* and Table S1). Specifically, binding (enrichment in the ChIP assay) fell linearly with affinity [$\log(K_d)$] down to affinity of $\log(K_d) = 6.5$ and remained thereafter. (The predictions were between the 25th and 75th percentiles for all but 4 of the 25 motifs and between the 5th and 95th percentiles for all of the 25 motifs.) Thus, the ability of genomic sequences containing PPARγ motifs to bind PPARγ on episomes is mainly determined—in a quantitative manner—by the core motif site, and therefore not by cooperative binding interactions with elements in the flanking sequence.

In Vivo PPARγ Binding to Genomic PPARγ Motifs Is Closely Related to the Chromatin Landscape. Our results suggest that the explanation for why certain PPARγ motif sites are differentially bound in vivo lies neither in differences in affinities of the core motif site, nor in cooperative protein–protein interactions with TFs that bind in the immediate vicinity. Instead, our data indicate that PPARγ binding at a motif site is strongly correlated with the epigenomic context of the larger genomic locus.

The vast majority (85%) of the PPARγ-bound motif sites lie in regions of open chromatin, defined in terms of DNase hypersensitivity [as assayed by formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing (54)] and marked by H3K4me1/H3K27ac identified in adipocytes using ChIP-seq (47). Moreover, genomic PPARγ binding within open regions was strongly dependent on the quantitative DNA accessibility and the enrichment of chromatin marks such as H3K27ac (Fig. 1*H* and *SI Appendix*, Fig. S4 *A* and *B*). Although only ∼1 in 10 PPARγ motif sites in open regions enriched for active chromatin marks (H3K4me1/2/3 or H3K27ac) are bound, we correctly identify the majority (82%) of bound motif sites with a precision of approximately one in four by using a logistic classifier based on five chromatin modifications (H3K4me1/2/3, H3K27ac, and H3K27me) (*SI Appendix*, Fig. S4*C*). Among sites predicted to be bound, those that are actually bound tend to have motif sites with a better motif match (as measured by position weight matrix score; *SI Appendix*, Fig. S4*D*). This suggests some remaining specificity may be due to differences in motif affinity, consistent with our findings for motif sites on plasmids.

In fact, our observation that within open chromatin TF binding is strongly correlated with the quantitative level of active chromatin marks appears to apply to many TFs. We analyzed 61 sequence-specific TFs profiled in ENCODE in seven cell types (121 total TF–cell type pairs) and found that the binding of 35 of these TFs was significantly correlated with quantitative DNA accessibility (measured by DNase-seq), and 45 were significantly correlated with enrichment of H3K27ac (Bonferroni-corrected $P_{\mathrm{Spearman}} < 0.01$) (*SI Appendix*, Fig. S5). TFs whose binding was not correlated with DNA accessibility include several pioneer factors, such as FOXA1, C/EBP, and NF-YA (46, 55, 56), the silencing factor REST, which maintains a repressive chromatin state (57), and CTCF, which binds insulator elements without active chromatin marks (58).

We note that the fact that TF binding is correlated with activating chromatin marks does not prove the direction of causality: it is possible that PPARγ binding not only depends on but also contributes to chromatin state (59, 60). With respect to DNase hypersensitivity, however, it is known that many (33%) of the genomic sites bound by PPARγ in terminally differentiated adipocytes show DNase hypersensitivity in the first 4 h of adipogenesis, before PPARγ is expressed (61).

Elements in the Sequence Flanking PPARγ Motifs Strongly Affect Gene Expression. We next sought to understand the determinants of enhancer activity for PPARγ motif sites in adipocytes. To explore this question, we measured the transcriptional activity of the 3,000 sequences in pool 1, consisting of 750 bound genomic sites, 750 unbound genomic sites, and their corresponding controls with disrupted core motif sites. We transfected the plasmid pool into 3T3-L1 cells 7 d postdifferentiation; grew the cells for 16 h; and extracted both RNA and DNA. We calculated a “relative enhancer activity” for each candidate enhancer, defined as the ratio of the proportion of total RNA to the proportion of total DNA corresponding to the enhancer (using the median ratio across the unique barcodes for each) (Fig. 2*A*). Measurements of relative enhancer activity were highly robust across three biological replicates ($r = 0.96$–0.97) (*SI Appendix*, Fig. S6 *A* and *B*).

As expected, disrupting the PPARγ motif site substantially decreased the relative enhancer activity in candidate enhancers, consistent with enhancer activity depending on PPARγ binding. Transcription was lower in 71% of all cases and 94% of cases where the expression from the native sequence was above the mean (Fig. 2*A*).

Surprisingly, as described above, although the bound and unbound genomic PPARγ sites showed no significant difference in PPARγ binding affinity (Fig. 1*E*), these sites exhibited sharply different enhancer activity (Fig. 2*A*). One-half of the sequences from bound sites drove expression levels above the 95th percentile for sequences from unbound genomic sites. Moreover, the transcriptional activity from the unbound sites was only weakly affected by disrupting the core PPARγ motif site (median, 1.05-fold decrease), indicating the expression from the unbound genomic sites is close to the background levels. Notably, the sequences from bound sites with disrupted PPARγ motif sites showed higher expression than sequences from the unbound genomic sites with the native PPARγ motifs ($P_{\text{Wilcox}} = 4.7 \times 10^{-13}$), suggesting the bound genomic sites are enriched for critical enhancer elements outside the core PPARγ motif site itself. [This observation holds true even when excluding sequences with addition PPARγ motif sites ($P_{\text{Wilcox}} = 4.3 \times 10^{-6}$).] Moreover, enhancer activity among bound genomic sites was only weakly correlated with PPARγ binding on plasmids ($\rho_{\text{Spearman}} = 0.24$) or in the genome ($\rho_{\text{Spearman}} = 0.12$). Together, these observations show that, although bound and unbound genomic sites do not differ in their inherent ability to bind PPARγ, they differ sharply in their enhancer activity as a result of additional elements in the surrounding sequence.

These results indicate that both the core motif sequence and the flanking sequence contribute to enhancer activity. To assess the relative contributions of each, we measured the enhancer activity of plasmids in pool 2, comprising 25 core motifs inserted into 25 different flanking sequences (described above). Unlike binding activity, enhancer activity was largely explained by the flanking sequence rather than by the core motif sequence (84% vs. 6% of variance explained; Fig. 2*B*). The median transcriptional activity was not substantially affected by changing the affinity of the core motif, although it did drop off for the motif with the lowest affinity as measured in the in vitro assay (*SI Appendix*, Fig. S1*G*), consistent with the effect of disrupting the PPARγ motif site. Thus, unlike PPARγ binding, enhancer activity depends largely on the flanking sequence, provided that PPARγ binding exceeds a threshold level.

Specific TF Motifs Correlate with Transcriptional Activity. To identify the elements in the flanking sequence that determine enhancer activity, we searched the sequences for known TF-binding motifs (Fig. 2*C*). We scanned the sequences using 1,490 vertebrate motifs corresponding to 612 TFs (of which 400 are expressed in adipocytes; *SI Appendix*, Fig. S6*C*) and counted the number of (nonoverlapping) occurrences of each motif. Across the 1,500 candidate enhancers, the number of motif occurrences per TF ranged from 5 (Tcf7l2) to 807 (Sp1) (median, 63; *SI Appendix*, Fig. S6*D*). Enhancer activity showed significant correlations with occurrences of 38 TF motifs, composed of 33 positively correlated motifs and 5 negatively correlated motifs [false-discovery rate (FDR) < 0.001 in permuted datasets] (Fig. 2 *D* and *E*, and *SI Appendix*, Tables S2 and S3).

Several lines of evidence suggest the TFs corresponding to the correlated motifs may functionally contribute to gene regulation in adipocytes.

First, 35 of the 38 TFs are expressed in adipocytes (vs. an expectation of only 26 by chance; $P_{\text{hypergeometic}} = 5.0 \times 10^{-6}$). The three "nonexpressed" TFs [estrogen receptor-like 2 (ERR2), ERR3, and nuclear receptor DAX-1] are nuclear receptors with motifs highly similar to the motif of ERR1, another nuclear receptor expressed in adipocytes (*SI Appendix*, Fig. S6*E*).

Second, consistent with previous observations that functional enhancers often contain homotypic clusters of motif sites (62, 63), the presence of additional PPARγ/RXR motif sites correlated strongly with enhancer activity. Candidate enhancers containing

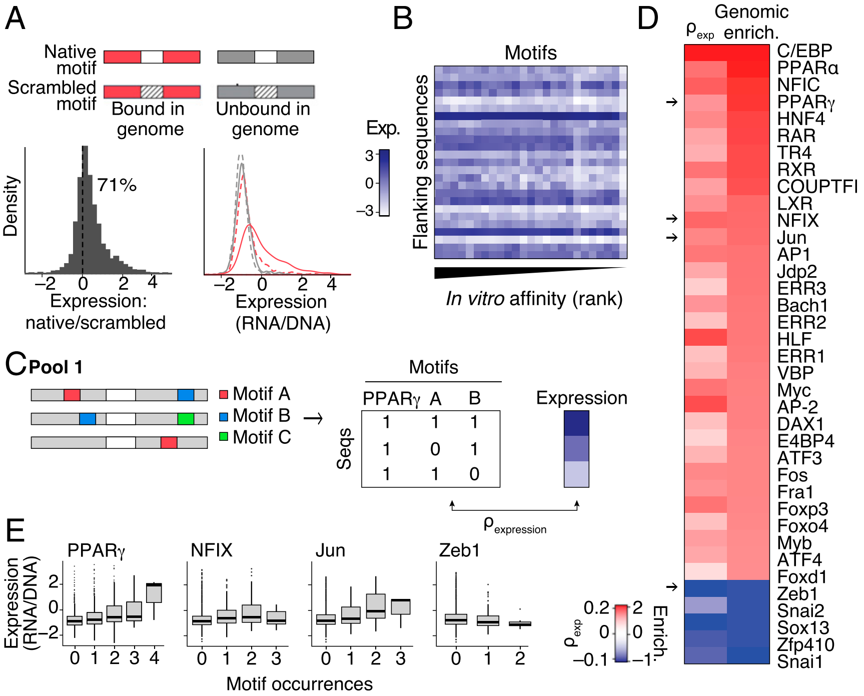

Fig. 2. Elements in flanking sequence govern enhancer activity. (*A*, *Left*) Ratio of expression ($\log_2$[RNA/DNA]) for each genomic sequence with an intact vs. disrupted central PPARγ motif. (*Right*) Expression corresponding to bound and unbound genomic sites with intact and disrupted core PPARγ motifs. (*B*) Expression driven by sequence constructs in pool 2, comprising 25 core PPARγ motifs (columns) swapped into 25 flanking sequences (rows). (*C*) Schematic of identification of TF motifs correlated with enhancer activity. For each TF motif, we calculated the correlation between motif counts and expression in pool 1. (*D*) Counts of 38 motifs were significantly correlated with expression (FDR < 0.01; red). These motifs are enriched or depleted around all 6,835 bound motif sites in the genome (blue). Arrows indicate motifs depicted in *E*. (*E*) Expression of candidate enhancers in pool 1, conditional on the number of occurrences of each motif.

additional PPARγ/RXR motif sites showed nearly twofold higher enhancer activity than those with only a single motif site.

Third, the TFs that recognize several of the positively correlated motifs are known to promote adipocyte differentiation (64, 65) or regulate gene expression in various stages of adipogenesis (54, 66–69). Conversely, the TFs that recognize several of the negatively correlated motifs are transcriptional repressors involved in inhibiting adipocyte-specific genes (68, 70) or promoting an alternate cell fate (71) (*SI Appendix*, *Supplemental Note*).

Fourth, occurrences of the correlated motifs are enriched in the immediate vicinity of the bound PPARγ sites in the genome. Of the 33 positively correlated TF motifs (detected in the 750 bound sites included in pool 1), 31 were significantly enriched. Conversely, all five negatively correlated TF motifs were significantly depleted across the full set of 6,835 PPARγ-bound genomic sites compared with unbound sites in adipocytes (hypergeometric test, $P = 10^{-6}$ to 10^{-300}) (Fig. 2*D*). Moreover, these 31 positively correlated TFs were the most significantly enriched and the 5 negatively correlated TFs were the most significantly depleted TFs among all 612 TFs tested. [The two TFs that were not significantly enriched, forkhead box protein O4 (Foxo4) and forkhead box protein P3 (Foxp3), have fairly degenerate 4-nt motifs.]

Together, these observations suggest that most of the correlated TF motifs identified in our assay indeed correspond to key regulators of gene expression in the adipocyte lineage.

TF Motifs Directly Influence Transcription. We reasoned that correlation between (*i*) the presence of specific TF-binding motifs in sequences from bound genomic sites and (*ii*) transcriptional activity in our enhancer assay does not necessarily imply that TF binding at these sites plays a causal role in determining transcriptional activity. An alternative possibility, for example, is that some TFBSs might be present at active enhancers because they were used for opening the chromatin before or during adipogenesis but do not contribute to driving transcriptional activity in adipocytes. We therefore next sought to identify motifs directly involved in transcription by deleting and inserting them in controlled contexts.

Disrupting TF Motifs Causes Changes in Expression. We first used an unbiased approach to identify elements that directly affect transcription, either motif sites and otherwise. We created an MPRA pool (pool 4; "block-mutated enhancers") that systematically introduced mutations in sliding windows in 25 of the candidate enhancer sequences from bound genomic sites. First, we disrupted 10-bp blocks, tiled every 5 bp across the sequence (excluding the core PPARγ/RXR motif site). Next, we swapped 20-bp blocks of sequence between the bound genomic site and a matched unbound site (Fig. 3*A* and *Materials and Methods*). For each mutant, we measured the enhancer activity and calculated the change in activity from the wild-type enhancer (Fig. 3*B*).

On average, mutations in the bound genomic sites that disrupted positively correlated TF motif sites reduced expression more than those that did not disrupt such sites (median of 1.7-fold vs. 1.2-fold) and were seven times more likely to cause a major decrease (>2-fold). Inserting blocks containing negatively correlated motifs [such as zinc finger E-box binding homeobox 1 (Zeb1) in the example shown in Fig. 3*B*] into the bound sites also substantially reduced expression (median of 2.0-fold). Finally, inserting blocks containing positively correlated motifs into unbound sites were seven times more likely to cause a major increase in expression. Overall, these data suggest that the correlated motifs account for the majority of elements that strongly contribute to expression in these enhancers.

We next sought to distinguish the contribution of the individual correlated TF motifs to transcription levels. We created an MPRA pool (pool 5; "motif-mutated enhancers") in which we systematically mutated each of the 38 significantly correlated motifs in 375 bound genomic sequences (Fig. 3*A* and *Materials and Methods*). For each of the mutated motif sites, we also created a control by mutating an equally sized block that did not overlap any motif.

For the majority (27 of 33) of the positively correlated TF motifs, the mutations in the motif sites reduced expression significantly more than the mutations in control regions ($P_{\mathrm{Wilcox}} <$ 0.05; *SI Appendix*, Table S2). Mutating activating transcription factor (ATF) and activator protein 1 (AP-1) factors had the largest effect (87% of mutations decreasing expression by more than twofold; Fig. 3*C* and *SI Appendix*, Fig. S7*B*). Mutations overlapping additional PPARγ motif sites surrounding the core PPARγ motif site also substantially reduced expression (64% reducing expression by more than twofold).

Most interesting were the remaining six TFs [nuclear factor I/X (NFIX), nuclear factor I/C (NFIC), Foxo4, Foxp3, forkhead box protein D1 (Foxd1), MYB proto-oncogene transcription factor (Myb), and AP-2], which did not appear to affect transcription in our assay—that is, these motifs are (*i*) enriched in genomic sequences that drive reporter expression but (*ii*) are not required for expression in our assay. A likely explanation is that these TFs act in different cellular contexts or have roles in vivo at these enhancers, such as remodeling chromatin, that are not required for activating transcription in our plasmid-based assay. Consistent with the latter notion, the list includes three Fox family TFs (Foxo4, Foxp3, and Foxd1), which act as pioneer factors that open chromatin during genomic enhancer activation (72–74), and two NFI TFs (NFIX and NFIC) that interact with histones (75, 76) and contribute to remodeling of nucleosome architecture (77, 78). The remaining two TFs (Myb and AP-2) have plausible roles in early adipocyte differentiation, regulating the final cell division (79, 80) and repressing an alternate cell fate (81), respectively.

Together, the mutagenesis data show that (*i*) for the majority of correlated motifs, disruption of a single TF motif site has a strong effect on transcription, whereas (*ii*) the remaining correlated motifs, although not necessary for reporter enhancer activity in adipocytes, may contribute to transcriptional activation in the genome through chromatin remodeling or during different stages of development.

TF Motifs Drive Expression When Inserted into New Contexts. The deletion analysis above revealed which motifs are required for expression. We next investigated the sufficiency of these correlated motifs for driving transcription when inserted into a new sequence context.

First, we substituted binding sites for each motif into existing motif sites in bound genomic sequences with strong activity in our assay. The resulting MPRA pool (pool 6, "motif-substituted enhancers") contained each of the 38 significantly correlated TF consensus motifs substituted into 95 distinct locations, yielding a "matrix" of 3,160 enhancer constructs (Fig. 3*A* and *Materials and Methods*).

Second, we added binding sites for 15 of the positively correlated motifs into sequences with low baseline activity, chosen from genomic regions that are bound by PPARγ in macrophages but not in adipocytes (46), and that contain a central PPARγ motif site but none of the other positive or negative TF motifs identified above. The motif sites were added individually and in pairs to these templates in nine different configurations, with two, four, or six total sites (Fig. 3*A* and *Materials and Methods*). This pool (pool 7; "synthetic enhancers") contained 4,324 sequences.

For each construct in pools 6 and 7, we calculated the "incremental enhancer activity" of the motif-substituted enhancer relative to its background sequence.

The relative strengths of the motif sites in driving transcription in the substituted and synthetic enhancers were highly consistent with the relative effects of their disruption measured previously. Of the 27 motifs whose sites caused significantly reduced expression when mutated, 25 led to increased expression when substituted into the bound genomic sequences, whereas the six TF motifs that did not significantly reduce expression when disrupted

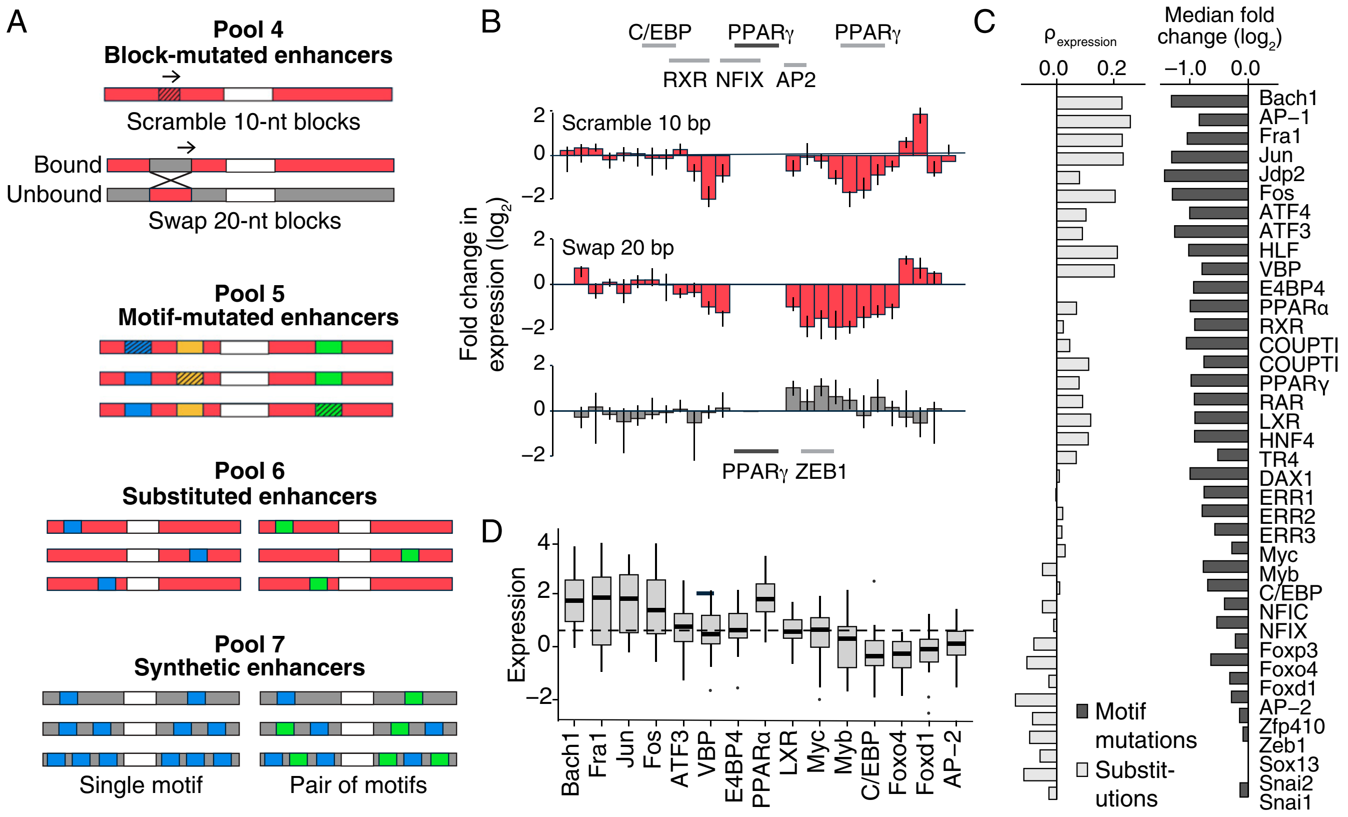

Fig. 3. Disrupting TF motifs affects enhancer activity. (*A*) Schematic of motif deletion pools. Pool 4 (block-mutated enhancers): for 25 bound genomic sites, we disrupted 10-bp blocks tiled every 5 bp across the sequence (*Top*) and swapped 20 bp blocks tiled every 5 bp across the sequence between bound (*Middle*) and unbound (*Bottom*) genomic sites, matched by the sequence of the central PPARγ motif. In each case, the central PPARγ motif was left intact. Pool 5 (motif-mutated enhancers): each occurrence of the 38 significantly correlated motifs were disrupted across 375 bound genomic sites. Pool 6 (substituted enhancers): motif sites for each of the 38 correlated motifs were substituted into 90 existing motif sites in bound genomic sites. Pool 7 (synthetic enhancers): motif sites for 15 of the positively correlated motifs were added individually (*Left*) and in pairs (*Right*) to three neutral templates in various configurations (see *SI Appendix*, *Supplemental Methods*). (*B*) Example of changes in expression caused by tiled mutations in a bound sequences (red, chr8:90491327–90491472) and unbound sequence (gray, chr14:57223369–57223514). Bars represent the $\log_2$ ratio of the mutant and wild-type expression for the block centered at that position. (*C*, *Right*) Median change in expression due to mutations in each motif across 375 bound genomic sites (Fig. 3*C*). (*Left*) Correlation between change in counts and change in expression for each motif in the substituted enhancers. (*D*) Expression of synthetic enhancers containing multiple copies of one motif.

also did not affect expression when substituted (Fig. 3*C* and *SI Appendix*, Fig. S7*A* and Table S2). The five negatively correlated motifs detected in the native enhancers strongly reduced transcription in the substituted enhancers. Similarly, synthetic enhancers containing motif sites associated with reduced expression in the mutated enhancers had activity significantly above the background level in most cases (seven of nine), whereas synthetic enhancers containing motif sites not associated with reduced expression did not (six of six cases) (Fig. 3*D*). Moreover, the average quantitative effects of the motifs on enhancer activity were highly concordant in the substituted and synthetic enhancers (Spearman $\rho = 0.85$; *SI Appendix*, Fig. S7 *C* and *D*, and Table S2).

Overall, 70% of these synthetic enhancers had higher transcriptional output than the template. Moreover, the proportion increased with the number of copies of the motif: 54% with two motifs, 73% with four motifs, and 81% with six motifs (*SI Appendix*, Fig. S7*E*). Thus, these motifs in combination with a central PPARγ motif are sufficient to drive expression independent of positioning and sequence background.

The Composition of TF Motif Sites Predicts Quantitative Expression Levels. We next sought to explore how much of the quantitative enhancer activity could be explained by the composition of TF motif sites (i.e., number and identity) in the enhancers. To explore how well this model captures enhancer activities in the naturally occurring enhancers, we fit a linear regression (in which each motif contributes additively to the total expression level) to predict the log-transformed transcript levels associated with the original candidate enhancer from 750 bound and 750 unbound genomic sites (pool 1) based on the number of motif sites for each of the 38 TFs identified above, as well as overall GC content [which is often elevated at TFBSs (18, 82, 83)]. Because some TFs have similar binding motifs, we removed redundant variables using stepwise variable selection to minimize the Akaike information criterion and evaluated the performance of the selected model using 10-fold cross-validation.

The selected linear model included 23 of the 38 TFs and explained one-fifth of the variance in enhancer activity in the training dataset (cross-validation $r^2 = 0.20$; *SI Appendix*, Fig. S8 *A*, *D*, and *E*). The model explains 15% of the variance among the bound genomic sites alone, indicating it is not simply differentiating between the bound and unbound class. To validate the model, we created an MPRA pool (pool 3) containing a new set of 750 bound motif sites and 750 unbound motif sites and measured their enhancer activity (*SI Appendix*, Fig. S9 *A* and *B*). The model explained an equal amount of variance in enhancer activity (20%) among the test set (Fig. 4*A*). (We note that ∼5% of the variance is due to inherent noise in the assay, as determined from comparison of biological replicates.)

Potential sources of the remaining variance include differences in motif site affinities, interactions between motif sites, effects of motif positioning, and additional features in the background sequence below our detection limit.

We thus next sought to determine how much additional variance could be explained if we held the affinities of the motif sites constant and controlled for the activity of the background sequence. We fit linear models to predict the incremental enhancer activity (described above) of the enhancers with inserted motif sites (pools 5 and 6) based on the changes in the number of motif sites relative to its background sequence. (As before, we removed redundant variables using stepwise variable selection and evaluated the performance of the selected models using 10-fold cross-validation.) The selected models explained 56% of the incremental activity of the substituted enhancers with a single inserted motif site (pool 5), and 45% of the incremental activity of the synthetic enhancers with pairs of sites inserted (pool 6) (*SI Appendix*, Fig. S8 *B* and *C*). These results show that motif affinity and background sequence contribute to variation in enhancer activity, but also suggest a substantial role for such features as nonadditive interactions and motif positioning.

Interactions Between Motif Sites Explain Additional Quantitative Variance in Expression. Under our simple linear model, each motif site contributes a fixed amount to enhancer function, independent of the other sites in the immediate region and their arrangement. We next wondered whether pairwise combinations of motifs could account for a substantial fraction of the variance not explained by the simple additive model. To the linear models above, we added second-order interaction terms for motif pairs that co-occur in at least 10 sequences. Because only a minority of the potential interaction terms are likely to be relevant, we used a Lasso regression model, which selects sparse models, and optimized the tuning parameter by 10-fold cross-validation.

The models with interactions explained substantially more variance than the linear models for both the natural enhancers and the enhancers with inserted motifs. For the natural enhancers, the selected model included 73 out of 384 possible interaction terms, and explained 40% of the variance in the training data in 10-fold cross-validation (pool 1) and 38% in the test data (pool 3) (Fig. 4*B* and *SI Appendix*, Fig. S10*A* and Table S4). For the substituted and synthetic enhancers, the selected models explained 74% and 52% of the incremental expression in 10-fold cross-validation, respectively (*SI Appendix*, Fig. S10 *B* and *C*). The improvement in the performance of this model compared with the additive model suggests that TF interactions play an important role in the function of these enhancers to generate combinatorial enhancer activity. (Because the Lasso process involves some arbitrary choices among correlated variables, the specific terms in the model should not be regarded as a comprehensive list of biologically meaningful interactions. Below, we consider interactions between specific TFs.)

Synergistic and Inhibitory Interactions Occur in Synthetic Enhancers. To explore combinatorial interactions between specific TFs, we first focused on interactions present in synthetic enhancers discussed above, containing all pairwise combinations of motif sites for 15 TFs inserted into inactive template sequences with a central PPARγ motif. Although motifs may co-occur in active native enhancers for a variety of reasons, only those pairs of TFs that functionally interact to drive enhancer function do not require specific positioning will be detected in our synthetic enhancers.

Using ANOVA to study interactions between pairs of TF motif sites, we identified 21 significant positive and negative interactions among the 15 TF motifs tested in the synthetic enhancers (Bonferroni-corrected $P_{F\ \mathrm{test}} < 0.01$; Fig. 4 *C* and *D*, and *SI Appendix*, Fig. S11*A* and Tables S5 and S6). (For simplicity below, we refer to these as interactions between the TFs expected to bind to the

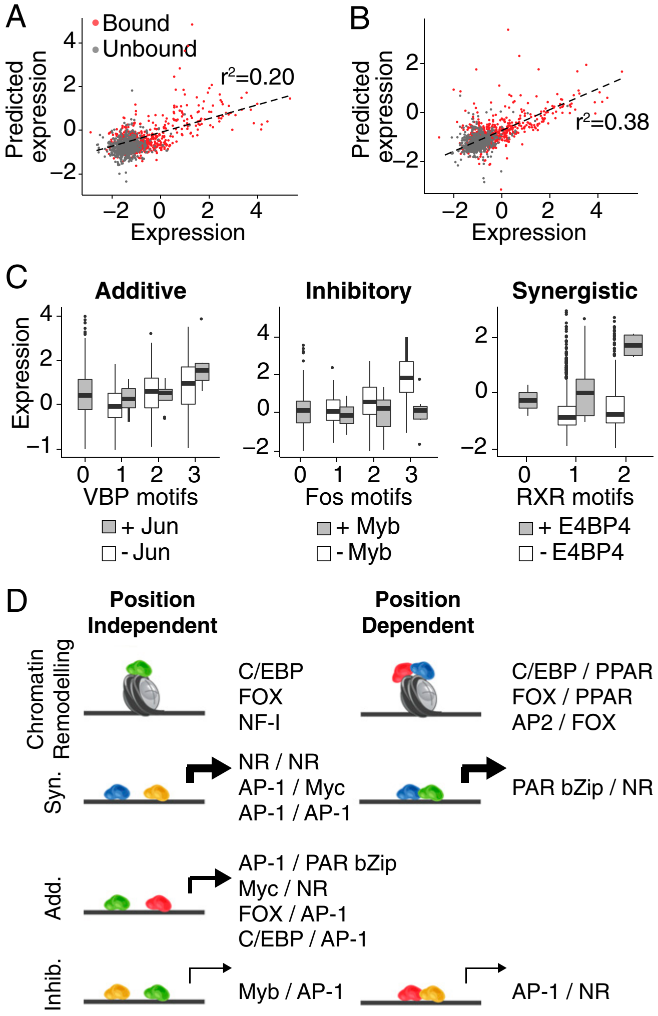

Fig. 4. Interactions between motifs contribute substantially to enhancer activity. (*A* and *B*) Performance of linear model (*A*) and Lasso model (*B*) predicting expression levels based on motif counts in independent test dataset (pool 3). (*C*) Boxplots represent expression of sequences in pool 1 (synergistic plot, *Left*) or pool 5 (additive and inhibitory plots, *Left* and *Center*), conditioned on counts of the two motifs. (*D*) Modes of interaction between TFs. Pioneer factors (*First Row*) are required to open chromatin at enhancers in the genome, but do not contribute strongly to transcriptional activation. Some pairs of TF enhance each other's activity, resulting in superadditive transcriptional output (*Second Row*). Other pairs of TFs function independently of each other, contributing additively to the transcriptional output (*Third Row*). Finally, some TFs mutually inhibit each other's activity, resulting in subadditive transcriptional output (*Fourth Row*). Add, additive; Inhib, inhibitory; and Syn, synergistic.

motif sites in adipocytes; however, we note that the presence of the motif site does not necessarily imply that the corresponding TF is bound in all cases.) These interactions fall into four main classes (Fig. 4 *C* and *D*). The first two classes comprised synergistic interactions between various AP-1 family factors (six pairs) and between AP-1 factors and protooncogene c-Myc (Myc) (two pairs). AP-1 family members are strong activators that are able to interact with a wide range of other TFs (84), and pairs of AP-1 factors are known to cooperatively induce DNA bending (85). The third class consisted of repressive interactions between AP-1 factors and nuclear receptors (seven pairs). Consistent with this result, AP-1 factors and the nuclear receptors PR, GR, and ER have been shown to mutually inhibit each other's ability to activate transcription (86, 87). Finally, the fourth class contained repressive interactions involving Myb (four pairs; Fig. 4*C*, *Center*). Myb contains a repressive domain and can function as a transcriptional repressor in some contexts (88). In the synthetic enhancers in our assay, Myb appears to act as a dominant repressor, returning transcription close to baseline levels. We also detected two interactions that did not fall into any of these four classes: a synergistic interaction between two nuclear receptors, PPAR and liver X receptor (LXR), and a repressive interaction between CCAAT/enhancer binding protein (C/EBP) and PPAR.

To confirm the interactions, we examined the results from the mutated and substituted enhancer pools (pools 5 and 6) to assess the effect of adding or deleting one of a pair of interacting motifs at several hundred sites in active naturally occurring sequences. For the majority of identified pairs, the effect of disrupting or inserting an interacting motif site differed significantly in sequences that contained the partner motif compared with sequences that did not (82% for motif disruptions and 86% for motif insertions; $P_{\mathrm{Wilcox}} < 0.0001$), and the direction of the effect was consistent with the interaction detected in the synthetic enhancers.

Finally, we tested whether the 21 interacting pairs showed significant interaction with respect to transcriptional activity of the naturally occurring enhancers (pool 1). Eleven of the pairs (including most AP-1/AP-1, AP-1/Myc, and Myb/TF pairs) co-occurred too rarely in the native enhancers (frequency < 0.01) to allow meaningful analysis. For the remaining 10 pairs, we expected to have power to see three to five significant interactions (Bonferroni-corrected $P_{F\ \mathrm{test}} < 0.05$) based on the effect sizes and counts in the natural enhancers (*Materials and Methods*). Consistent with this, we found five pairs with significant interaction terms, all of which had the same sign as the interaction in the synthetic enhancers (*SI Appendix*, Table S5).

Naturally Occurring Enhancers Contain Additional Classes of Interactions. We next examined native enhancer sequences (pool 1) to identify additional TF interactions not detected in the synthetic enhancers. Such interactions might fall into three classes: (*i*) pairs of motifs that do not involve the 15 TFs used in creating synthetic enhancers on a neutral template, (*ii*) pairs involving TFs such as pioneer factors that are correlated in the genome with an effective enhancer but not necessary for expression in adipocytes, or (*iii*) pairs that require specific spacing and orientation, which were not imposed in the synthetic enhancers.

Among the 38 TFs correlated with expression, we detected 25 significant positive and negative interactions beyond those seen in the synthetic enhancers (Fig. 4 *C* and *D*, and *SI Appendix*, Fig. S11*A* and Tables S5 and S7). They fell into each of the three classes. The first class included 11 pairs of inhibitory AP-1/NR pairs that were not tested in the synthetic enhancers. The second class (TFs not required for reporter expression) consisted of two groups: the first group (six pairs) involves synergistic interactions between nuclear receptors and FOX TFs, which have pioneering ability (72–74); the second group (four pairs) involves interactions between various TFs and AP-2, which may play a role in early adipocyte differentiation by repressing an alternate cell fate (81), but is down-regulated in terminally differentiated adipocytes.

To study the third class (spatially constrained interactions), we evaluated whether the two motifs occurred adjacently (<10 bp apart) more often than would be expected by chance in naturally occurring enhancers. Of the 25 interacting pairs, 8 showed significant enrichment of adjacent co-occurrences in bound PPARγ enhancers in the genome (Bonferroni-corrected $P_{\mathrm{Fisher}} < 0.01$; *SI Appendix*, Fig. S11*B*). The eight pairs are CEBP/ATF-3, AP-2/Foxo4, three AP-1/NR pairs, and three FOX/NR pairs. The first pair is known to bind as a heterodimer to composite motifs in the genome (89, 90), suggesting that the enriched configurations reflect functional physical interactions. Furthermore, FOXO family TFs have also been shown to physically interact with a number of nuclear receptors, often in a ligand-dependent manner, resulting in changes in the activity of the two TFs (91). AP-1 and NRs both directly interact with CBP/p300 to activate transcription (92), and could interfere with each other's interaction when in close proximity.

Of the five adjacent pairs that include an asymmetric motif, four were enriched for a specific orientation of the two motifs relative to each other. Interestingly, eight of the interactions detected in the synthetic enhancers were also biased toward a specific configuration in the natural enhancers, suggesting that these pairs may interact more efficiently in one orientation.

Discussion

Deciphering the regulatory code of enhancers requires understanding how the combinatorial input of different TFBSs lead to precise TF-binding patterns and gene expression outputs. Here, we use a series of MPRA experiments, involving 32,115 distinct enhancer constructs, to systematically evaluate the factors that govern PPARγ binding and regulation in adipocytes. We demonstrate that (*i*) the PPARγ motif affinity (and not cooperative elements in the immediately flanking sequence) largely determines PPARγ binding to genomic sequences when removed from their chromatin context; (*ii*) enhancer activity depends not only on PPARγ binding but also on a network of 20–30 TF motifs in the flanking sequence that have distinct quantitative contributions to expression; and (*iii*) various pairs of motifs interact in additive, inhibitory, and synergistic ways with varying constraints on motif positioning. Although in this study we measured enhancer activity in an episomal context, a recent study found that enhancer activity was highly concordant between episomal and genomic contexts (r = 0.86 across 2,236 candidate enhancers vs. 0.90–0.98 for replicates within each context) (93). Importantly, our results show that PPARγ binding and enhancer activity are independently regulated.

Studies of several TFs, including PPARγ, have observed strong correlations between DNA accessibility and TF binding, leading to the hypothesis that TF binding for nonpioneer factors is largely governed by nucleosomes or the larger chromatin landscape (31–33, 36, 51, 52, 94). In this model, pioneer factors bind to sites in closed chromatin and displace surrounding nucleosomes, allowing other TFs to bind to neighboring sites, which may then reinforce nucleosome exclusion. Our results support this model, demonstrating that, in an episomal context, both bound and unbound genomic motif sites bind PPARγ equally well (excluding the possibility of latent features controlling motif affinity) and that binding is largely independent of sequences immediately surrounding the PPARγ motif (excluding a major role for direct cooperative binding). Although the presence of H3K27ac at bound PPARγ sites has been widely appreciated, our results suggest a graded effect even among open sites, whereby stronger quantitative chromatin accessibility is associated with more frequent TF occupancy. This relationship appears to be general to other TFs: we see a similar quantitative correlation between quantitative H3K27ac signal and TF binding for nearly all of the TFs and cell lines profiled in the ENCODE Project.

Our study identifies a collection of ∼20 TF motifs that are correlated with higher enhancer activity in naturally occurring enhancers and, with the exception of pioneer factors, play direct roles in enhancer activity (as assayed by mutational perturbation). Intriguingly, the TF motifs that affect enhancer activity correspond closely to those that are most enriched in the genomic sequences of PPARγ binding sites in adipocytes. If this observation can be confirmed for some additional TFs and cell types, it may allow the use of motif co-occurrences in the genome to be used to predict the functional activities of TFs.

Although cooperative binding of TFs to composite motif sites has been studied in depth, much less is known about how sets of TFs, once bound, influence gene expression. Characterizing such interactions is difficult due to the large number of possible combinations and uncertainty about spatial constraints. By choosing a related set of enhancers and focusing on TFs that correlate individually with enhancer activity, our approach yields a tractable number of potential interactions for functional characterization, facilitating the identification of a basic set of grammatical rules governing the activity of these enhancers.

Using this approach, we detect examples of subadditive, additive, and superadditive interactions between different pairs of TFs with varying degrees of spatial constraint. The identified interactions fall into several classes, comprising TFs from specific structural families that interact similarly with TFs from other families. We reproduce several types of interactions supported by previous studies, such as mutual inhibition of nuclear receptors and AP-1 factors and synergistic interactions between pairs of AP-1 factors (84, 86), and identify additional intriguing interactions, such as quenching of enhancer activity by Myb. Our results highlight the need for better understanding of the molecular and biochemical basis of TF activity to understand the mechanisms underlying TF cooperativity and combinatorial transcriptional activation.

The approach described here provides a framework to dissect the regulatory grammar underlying enhancer function by (*i*) systematically identifying TFBSs correlated with activity, (*ii*) isolating the independent quantitative contributions of each TF to enhancer output by disrupting and inserting binding sites in controlled contexts, (*iii*) identifying interactions between TFBSs in synthetic and natural enhancers that influence enhancer activity, and (*iv*) characterizing spatial constraints on the identified interactions. Our approach is readily applicable to many TFs and cell types to understand the regulatory grammar of diverse sets of regulatory elements, revealing the prevalence and generality of these rules. Understanding this regulatory code is critical in understanding how gene expression drives fundamental biological processes such as differentiation and development, as well as interpreting the increasing number of variants in regulatory regions implicated in cancer and other diseases.

Materials and Methods

Design and Synthesis of Plasmid Pools. Oligonucleotide libraries containing the 145- or 150-bp candidate enhancers were synthesized (Agilent Technologies) and unique barcodes were added by PCR. The oligonucleotide libraries were cloned into a plasmid backbone with a minP-luc2 insert, as previously described (48, 95).

Cell Culture and Transfections. 3T3-L1 cells were cultured and differentiated as described in ref. 96. The plasmid libraries were transfected into differentiated 3T3-L1 adipocytes using a Nucleofector II Device with Cell Line Kit SE (Lonza).

PPARγ ChIP-Seq. ChIP was performed with a PPARγ antibody (Cell Signaling) as described in ref. 97. The ChIP and whole-cell extract (WCE) fragments derived from each oligo were counted, and enrichment was calculated as the $\log_2$-transformed ratio of ChIP and WCE counts. Oligos with fewer than 100 WCE counts were removed.

In Vitro PPARγ Binding Assay. Fluorescently labeled 26-bp oligos containing genomic PPARγ motifs (16 bp) flanked by 5 bp on either side were synthesized (Invitrogen). These oligos were used to generate fluorescently labeled double-stranded oligos. Equilibrium recombinant PPARγ (Promega) binding to each sequence was measured for a range of input DNA concentrations by MITOMI as described previously (45, 53).

RNA Preparation and Sequencing of RNA-Derived Barcodes. RNA was harvested and the barcodes were isolated as previously described (48). mRNA and plasmid reads for each barcode were counted, and the counts for all barcodes corresponding to each enhancer construct were summed. Activity was calculated as the median $\log_2$-transformed ratio of RNA and DNA.

Motif Analysis. Motif instance matching was performed using FIMO (98) for all vertebrate motifs from TRANSFAC (99) and JASPAR (100). Overlapping motif matches corresponding to the same TF were merged.

ACKNOWLEDGMENTS. We thank Aviv Regev for valuable advice regarding the design and interpretation of the experiments, and Brian Cleary, Tim Wang, Mitch Guttman, Chris Burge, members of the E.S.L. Laboratory, and numerous colleagues for helpful comments and discussion. This work was supported by National Human Genome Research Institute Grant 2U54HG003067-10 (to E.S.L.), National Institute of General Medical Sciences Grant T32GM007753 (to S.R.G.), and National Institutes of Health Grant R01 HG006785 (to T.S.M.), as well as funding by SystemsX.ch (B.D.) and internal support by École Polytechnique Fédérale de Lausanne (B.D.).

1. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
2. Kundaje A, et al.; Roadmap Epigenomics Consortium (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–330.
3. Spitz F, Furlong EE (2012) Transcription factors: From enhancer binding to developmental control. *Nat Rev Genet* 13(9):613–626.
4. Ptashne M, Gann A (1997) Transcriptional activation by recruitment. *Nature* 386(6625):569–577.
5. Badis G, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324(5935):1720–1723.
6. Badis G, et al. (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* 32(6):878–887.
7. Grove CA, et al. (2009) A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell* 138(2):314–327.
8. Jolma A, et al. (2013) DNA-binding specificities of human transcription factors. *Cell* 152(1-2):327–339.
9. Zhu C, et al. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res* 19(4):556–566.
10. Gerstein MB, et al. (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489(7414):91–100.
11. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830):1497–1502.
12. Ren B, et al. (2000) Genome-wide location and function of DNA binding proteins. *Science* 290(5500):2306–2309.
13. Robertson G, et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4(8):651–657.
14. Wei CL, et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell* 124(1):207–219.
15. Kheradpour P, et al. (2013) Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23(5):800–811.
16. Kheradpour P, Kellis M (2014) Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* 42(5):2976–2987.
17. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA (2014) High-throughput functional testing of ENCODE segmentation predictions. *Genome Res* 24(10):1595–1602.
18. White MA, Myers CA, Corbo JC, Cohen BA (2013) Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci USA* 110(29):11952–11957.
19. Whitfield TW, et al. (2012) Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* 13(9):R50.
20. Biggin MD (2011) Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 21(4):611–626.
21. Landolin JM, et al. (2010) Sequence features that drive human promoter function and tissue specificity. *Genome Res* 20(7):890–898.
22. Fisher WW, et al. (2012) DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 109(52):21330–21335.
23. Rowan S, et al. (2010) Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* 24(10):980–985.
24. Jiang J, Levine M (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* 72(5):741–752.
25. Gaudet J, Mango SE (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* 295(5556):821–825.
26. Siggers T, Duyzend MH, Reddy J, Khan S, Bulyk ML (2011) Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol* 7:555.
27. Slattery M, et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147(6):1270–1282.
28. Dror I, Golan T, Levy C, Rohs R, Mandel-Gutfreund Y (2015) A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res* 25(9):1268–1280.
29. Gordân R, et al. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep* 3(4):1093–1104.
30. Levo M, et al. (2015) Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res* 25(7):1018–1029.
31. Barozzi I, et al. (2014) Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell* 54(5):844–857.
32. Mirny LA (2010) Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci USA* 107(52):22534–22539.
33. Thurman RE, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82.
34. Raveh-Sadka T, et al. (2012) Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat Genet* 44(7):743–750.
35. Guertin MJ, Lis JT (2013) Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev* 23(2):116–123.
36. John S, et al. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 43(3):264–268.

37. Li XY, et al. (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol* 12(4):R34.
38. Polach KJ, Widom J (1996) A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J Mol Biol* 258(5):800–812.
39. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–589.
40. Buck MJ, Lieb JD (2006) A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nat Genet* 38(12):1446–1451.
41. Zeitlinger J, et al. (2003) Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* 113(3):395–404.
42. Blau J, et al. (1996) Three functional classes of transcriptional activation domain. *Mol Cell Biol* 16(5):2044–2055.
43. Han K, Levine MS, Manley JL (1989) Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. *Cell* 56(4):573–583.
44. Scholes C, DePace AH, Sanchez A (2016) Integrating regulatory information via combinatorial control of the transcription cycle. bioRxiv:039339.
45. Isakova A, Berset Y, Hatzimanikatis V, Deplancke B (2016) Quantification of cooperativity in heterodimer-DNA binding improves the accuracy of binding specificity models. *J Biol Chem* 291(19):10293–10306.
46. Lefterova MI, et al. (2008) PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev* 22(21):2941–2952.
47. Mikkelsen TS, et al. (2010) Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143(1):156–169.
48. Melnikov A, et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 30(3):271–277.
49. Nielsen R, et al. (2008) Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev* 22(21):2953–2967.
50. Berman BP, et al. (2002) Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc Natl Acad Sci USA* 99(2):757–762.
51. Guertin MJ, Lis JT (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* 6(9):e1001114.
52. Robertson AG, et al. (2008) Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res* 18(12):1906–1917.
53. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315(5809):233–237.
54. Waki H, et al. (2011) Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. *PLoS Genet* 7(10):e1002311.
55. Lupien M, et al. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132(6):958–970.
56. Sherwood RI, et al. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* 32(2):171–178.
57. Ballas N, et al. (2001) Regulation of neuronal traits by a novel transcriptional complex. *Neuron* 31(3):353–365.
58. Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28(8):817–825.
59. Gelman L, et al. (1999) p300 interacts with the N- and C-terminal part of PPARgamma2 in a ligand-independent and -dependent manner, respectively. *J Biol Chem* 274(12):7681–7688.
60. Blanco JC, et al. (1998) The histone acetylase PCAF is a nuclear receptor coactivator. *Genes Dev* 12(11):1638–1651.
61. Siersbæk R, et al. (2011) Extensive chromatin remodelling and establishment of transcription factor "hotspots" during early adipogenesis. *EMBO J* 30(8):1459–1472.
62. Gotea V, et al. (2010) Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res* 20(5):565–577.
63. Lifanov AP, Makeev VJ, Nazina AG, Papatsenko DA (2003) Homotypic regulatory clusters in *Drosophila*. *Genome Res* 13(4):579–588.
64. Rosen ED, et al. (2002) C/EBPalpha induces adipogenesis through PPARgamma: A unified pathway. *Genes Dev* 16(1):22–26.
65. Yu K, et al. (2014) Activating transcription factor 4 regulates adipocyte differentiation via altering the coordinate expression of CCATT/enhancer binding protein β and peroxisome proliferator-activated receptor γ. *FEBS J* 281(10):2399–2409.
66. Dahle MK, et al. (2002) Mechanisms of FOXC2- and FOXD1-mediated regulation of the RI alpha subunit of cAMP-dependent protein kinase include release of transcriptional repression and activation by protein kinase B alpha and cAMP. *J Biol Chem* 277(25):22902–22908.
67. Distel RJ, Ro HS, Rosen BS, Groves DL, Spiegelman BM (1987) Nucleoprotein complexes that regulate gene expression in adipocyte differentiation: Direct participation of c-*fos*. *Cell* 49(6):835–844.
68. Patel YM, Lane MD (2000) Mitotic clonal expansion during preadipocyte differentiation: Calpain-mediated turnover of p27. *J Biol Chem* 275(23):17653–17660.
69. Seo J, et al. (2009) Atf4 regulates obesity, glucose homeostasis, and energy expenditure. *Diabetes* 58(11):2565–2573.
70. Lee YH, et al. (2013) Transcription factor Snail is a novel regulator of adipocyte differentiation via inhibiting the expression of peroxisome proliferator-activated receptor γ. *Cell Mol Life Sci* 70(20):3959–3971.
71. Cameron TL, Belluoccio D, Farlie PG, Brachvogel B, Bateman JF (2009) Global comparative transcriptome analysis of cartilage formation in vivo. *BMC Dev Biol* 9:20.
72. Cuesta I, Zaret KS, Santisteban P (2007) The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol Cell Biol* 27(20):7302–7314.
73. Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS (2009) Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev* 23(7):804–809.
74. Zaret KS, et al. (2008) Pioneer factors, genetic competence, and inductive signaling: Programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol* 73:119–126.
75. Dusserre Y, Mermod N (1992) Purified cofactors and histone H1 mediate transcriptional regulation by CTF/NF-I. *Mol Cell Biol* 12(11):5228–5237.
76. Alevizopoulos A, et al. (1995) A proline-rich TGF-beta-responsive transcriptional activator interacts with histone H3. *Genes Dev* 9(24):3051–3066.
77. Ferrari S, et al. (2004) Chromatin domain boundaries delimited by a histone-binding protein in yeast. *J Biol Chem* 279(53):55520–55530.
78. Hebbar PB, Archer TK (2003) Nuclear factor 1 is required for both hormone-dependent chromatin remodeling and transcriptional activation of the mouse mammary tumor virus promoter. *Mol Cell Biol* 23(3):887–898.
79. Pittenger MF, et al. (1999) Multilineage potential of adult human mesenchymal stem cells. *Science* 284(5411):143–147.
80. Joaquin M, Watson RJ (2003) Cell cycle regulation by the B-Myb transcription factor. *Cell Mol Life Sci* 60(11):2389–2401.
81. Huang Z, Xu H, Sandell L (2004) Negative regulation of chondrocyte differentiation by transcription factor AP-2alpha. *J Bone Miner Res* 19(2):245–255.
82. Erwin GD, et al. (2014) Integrating diverse datasets improves developmental enhancer prediction. *PLoS Comput Biol* 10(6):e1003677.
83. Wang J, et al. (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* 22(9):1798–1812.
84. Chinenov Y, Kerppola TK (2001) Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* 20(19):2438–2452.
85. Kerppola TK, Curran T (1993) Selective DNA bending by a variety of bZIP proteins. *Mol Cell Biol* 13(9):5479–5489.
86. Shemshedini L, Knauthe R, Sassone-Corsi P, Pornon A, Gronemeyer H (1991) Cell-specific inhibitory and stimulatory effects of Fos and Jun on transcription activation by nuclear receptors. *EMBO J* 10(12):3839–3849.
87. Herrlich P (2001) Cross-talk between glucocorticoid receptor and AP-1. *Oncogene* 20(19):2465–2475.
88. Oh IH, Reddy EP (1999) The *myb* gene family in cell growth, differentiation and apoptosis. *Oncogene* 18(19):3017–3033.
89. Vinson CR, Hai T, Boyd SM (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design. *Genes Dev* 7(6):1047–1058.
90. Vallejo M, Ron D, Miller CP, Habener JF (1993) C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to cAMP response elements. *Proc Natl Acad Sci USA* 90(10):4679–4683.
91. van der Vos KE, Coffer PJ (2008) FOXO-binding partners: It takes two to tango. *Oncogene* 27(16):2289–2299.
92. Kamei Y, et al. (1996) A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* 85(3):403–414.
93. Inoue F, et al. (2017) A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* 27(1):38–52.
94. Simicevic J, et al. (2013) Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat Methods* 10(6):570–576.
95. Tewhey R, et al. (2016) Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* 165(6):1519–1529.
96. Eguchi J, et al. (2008) Interferon regulatory factors are transcriptional regulators of adipogenesis. *Cell Metab* 7(1):86–94.
97. Mikkelsen TS, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153):553–560.
98. Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.
99. Matys V, et al. (2006) TRANSFAC and its module TRANSCompel: Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34(Database issue):D108–D110.
100. Mathelier A, et al. (2014) JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42(Database issue):D142–D147.