High-throughput functional comparison of promoter and enhancer activities

Thomas A. Nguyen, Richard D. Jones, Andrew R. Snively, Andreas R. Pfenninger, Rory Kirchner, Martin Hemberg, and Jesse M. Gray

1Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 2Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; 3Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA; 4Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom

Promoters initiate RNA synthesis, and enhancers stimulate promoter activity. Whether promoter and enhancer activities are encoded distinctly in DNA sequences is unknown. We measured the enhancer and promoter activities of thousands of DNA fragments transduced into mouse neurons. We focused on genomic loci bound by the neuronal activity-regulated coactivator CREBBP, and we measured enhancer and promoter activities both before and after neuronal activation. We find that the same sequences typically encode both enhancer and promoter activities. However, gene promoters generate more promoter activity than distal enhancers, despite generating similar enhancer activity. Surprisingly, the greater promoter activity of gene promoters is not due to conventional core promoter elements or splicing signals. Instead, we find that particular transcription factor binding motifs are intrinsically biased toward the generation of promoter activity, whereas others are not. Although the specific biases we observe may be dependent on experimental or cellular context, our results suggest that gene promoters are distinguished from distal enhancers by specific complements of transcriptional activators.

High-throughput comparison of promoter and enhancer activities using Massively Parallel Reporter Assays (MPRAs)

To identify DNA sequences that might distinctly encode enhancer and promoter activities, we compared these two activities across a library of sequences derived from distal enhancers and gene promoters (Fig. 1A). We chose 253 distal enhancer and 234 gene promoter loci based on their chromatin state in mouse cortical neurons, a well-characterized experimental system that we previously used to uncover widespread transcription at enhancers (Kim et al. 2010). Our primary selection criteria were binding of the coactivator CREBBP (Kim et al. 2010) and an increase in histone H3 acetylated on Lys-27 (H3K27ac) upon neural activation (Malik et al. 2014). We also required each locus to have a human counterpart.

© 2016 Nguyen et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
mid library rather than from cells, because we found MPRA read counts from plasmid DNA and DNA extracted from neurons to be indistinguishable.

We typically sequenced control DNA amplified from the plasmid pool into AAV for infection of mouse cortical neurons and quantified cDNA and control DNA barcode abundance by sequencing. We typically sequenced control DNA amplified from the plasmid pool into AAV for infection of mouse cortical neurons and quantified cDNA and control DNA barcode abundance by sequencing.

Figure 1. Measuring thousands of promoter and enhancer activities in primary neurons using Massively Parallel Reporter Assays (MPRAs). (A) For each of 234 CREBBP-bound promoter and 253 CREBBP-bound distal enhancer loci, we synthesized 11 139-bp tiles whose start sites are spaced at 20-bp intervals. The center tile at each locus is centered on the called CREBBP peak. Plots show tile coverage as a count of quantifiable tiles versus position relative to the locus center. Distal enhancers are at least 500 bp from an annotated TSS. (B) After a series of cloning steps (Supplemental Figs. S1, S2), we tagged each tile with one or more unique 18-bp barcodes. In the enhancer test, the abundance of each barcode in a cellular mRNA pool depends on the enhancer activity of its associated tile. The library contains a mix of tiles from enhancer (black) and promoter (orange) loci. pFos is a 100-bp basal FOS promoter. (C) In the promoter test, the abundance of each barcode depends on the promoter activity of its tile. (D) We packaged MPRA libraries into AAV for infection of mouse cortical neurons and quantified cDNA and control DNA barcode abundance by sequencing. We typically sequenced control DNA amplified from the plasmid pool into AAV for infection of mouse cortical neurons and quantified cDNA and control DNA barcode abundance by sequencing.

Because our primary goal was to enable comparison of promoter and enhancer activities regardless of the extent or polarity of neural activity regulation, we generally present for each sequence the condition with maximal activity (but also show each condition separately) (Supplemental Figs. S11–S14). We directly compared enhancer and promoter activities from MPRA with enhancer and promoter activities from luciferase assays and found them to be strongly correlated (Pearson’s R = 0.95, 0.99; P < 0.05) (Fig. 2C, D,G; Supplemental Fig. S11).

Enhancer and promoter activities colocalize in the genome but are quantitatively decoupled

We evaluated whether enhancer- and promoter-specific MPRA libraries, harvested RNA, and quantified barcode expression using high-throughput sequencing (Figs. 1D, 2A,B,E,F; Supplemental Figs. S2B, S3A,B, S4). After normalization, MPRA activity values represent a fold increase over negative control sequences, which were included in every MPRA library (Methods). Because CREBBP function is regulated by neural activity, we measured enhancer and promoter activities both before and after neural activation via potassium chloride (KCl)-mediated depolarization. We found examples of tiles that were positively regulated, negatively regulated, and unregulated by neural activity (Supplemental Table S1).
to the offset in sense and antisense activities that is predicted by a model in which sense and antisense initiation events require distinct core promoters separated by $\sim$180 bp (Duttke et al. 2015; Scruggs et al. 2015); it suggests instead that bidirectional initiation may occur at the flanks of activator-bound sequences with little requirement for specific core promoter elements.

Both positive and negative relationships between enhancer and promoter activities have been recently hypothesized (Kim et al. 2010; Li et al. 2012; Andersson et al. 2015), so we compared the enhancer and promoter activities of defined sequences. We observe a clear positive correlation between enhancer and promoter activity (Spearman’s $\rho = 0.45$ and $P < 10^{-16}$, asymptotic $t$-statistic) (Fig. 3D; Supplemental Fig. S12G,H). Interestingly, however, promoter-derived tiles have greater promoter activity than distal enhancer-derived tiles, despite having similar enhancer activity (Fig. 3A,D; Supplemental Fig. S7A,B). Correspondingly, the median ratio of promoter-to-enhancer activity of promoter-derived tiles is 3.1-fold higher than of distal enhancer-derived tiles ($P < 10^{-16}$, Wilcoxon rank-sum test) (Fig. 3E; Supplemental Fig. 14G,H). The lower promoter activity of distal enhancer-derived compared to promoter-derived sequences is unlikely due to a lack of directional annotation for distal enhancers because eRNAs are transcribed bidirectionally (Kim et al. 2010), and the promoter activity we observe for promoter-derived sequences is orientation-insensitive (Fig. 3C). Nor is the lower promoter activity of distal enhancer-derived sequences likely due to distal enhancers being longer and therefore less well captured by our short tiles because the distal enhancer loci have shorter DHS-accessible regions than the promoter loci (median 1252 versus 1755 bp, $P < 10^{-7}$ by Student’s $t$-test). We conclude that enhancer and

---

**Figure 2.** Replicability of enhancer and promoter activities from MPRAs. (A) Replicability of read counts across two biological replicates of an enhancer activity MPRA. (B) Replicability of enhancer activity from A. MPRA activity is defined in this and all subsequent figures as the cDNA to DNA ratio, normalized to the corresponding ratio for negative control sequences, which are included in each MPRA library. (C) Schematic of luciferase assay design. Enhancer activity was assessed with test sequences upstream of a minimal FOS promoter. Promoter activity was tested as in Figure 1C with a luciferase coding sequence in place of GFP. (D) Comparison of MPRA-based and luciferase-based enhancer activity measurements, with each shape representing a specific test sequence. Test sequences are artificial, with repeats of RFX (square), AP1 (circle), MYBL2 (triangle), and NFYSshort (diamond) motifs with spacers TTATTTTAAGA (RFX, MYBL2) and CCCGCGCTGCC (AP1, NFYSshort). (F–G) Same as A, B, D but for promoter activity. (H) Correlation of MPRA-based promoter activity with CAGE tag counts. Pearson’s $R = 0.27$, $P < 10^{-16}$ from $t$-statistic. Throughout this work, except where noted, enhancer and promoter activities are reported as the maximal activity of the unstimulated and KCl-depolarized conditions. In A, B, E, and F, the unstimulated condition is shown.

(Rho $= 0.29$, $P < 10^{-16}$) (Fig. 2H; Supplemental Fig. SSD–E; FANTOM Consortium and the RIKEN PMI and CLST [DGT] 2014). These results give us confidence that enhancer- and promoter-specific MPRAs provide relevant measurements of enhancer and promoter activities.

Our data provide a high-resolution survey of metazoan enhancer activity, enabling us to ask precisely where enhancer activities localize within the mouse genome. Interestingly, we find that mean enhancer activity is similar for promoter- and distal enhancer-derived tiles (Fig. 3A; Supplemental Figs. S7A,B, S14A,B). At promoters, enhancer activity is greatest for tiles that contain annotated mRNA TSSs (Fig. 3B; Supplemental Fig. S14C,E). Because gene promoters frequently interact with each other within genomic chromatin (Li et al. 2012), these results suggest that many gene promoters may also function to enhance transcription of other genes.

Our data also provide a high-resolution genomic survey of promoter activity. We find that promoter activity is maximal for tiles centered $\sim$40 bp upstream of mRNA TSSs, the same position with maximal enhancer activity (Fig. 3B; Supplemental Figs. S7C, S14C,E). Maximal promoter activity therefore colocalizes with upstream activating regions ($\sim$500 to $\sim$40 bp) that bind transcription-activating factors (Guarente 1984). It also colocalizes with traditional core promoter regions ($\sim$40 to $\sim$40) that contain elements such as the TATA box and the INR, which facilitate RNA polymerase II initiation (Kadonaga 2012). However, when we compared the promoter activities of tiles inserted into our MPRA reporters in sense versus antisense orientations relative to their endogenous mRNA TSSs, we found them to be indistinguishable ($P > 0.05$, KS test) (Fig. 3C; Supplemental Figs. S7D, S14D,F). This result is in contrast

---

Genome Research 3 www.genome.org
promoter activities can be encoded distinctly. In particular, the strongest promoter activity has additional sequence and mechanistic requirements, beyond those required for enhancer activity. The elevated promoter activity of promoter-derived tiles is not due to core promoter elements or 5′ splice signals.

Having determined that promoter activity has specific sequence determinants beyond those required for enhancer activity, we focused our efforts on identifying these determinants. Of all promoter-derived sequences, we found that promoter activity-only tiles are more likely to be situated from 0 to +100 relative to mRNA TSSs than enhancer activity-only tiles ($P < 0.003$, $\chi^2$ test) (Supplemental Fig. S8), suggesting that canonical core promoter elements might contribute to promoter but not enhancer activity. However, we found that the sequences that contained the TATA box (TTATTTGG) or the INR (YYANWYY or YYRRWYY) have no higher promoter activity than those that lack them, whether

![Figure 3](https://genomic.org/fig3.png)

**Figure 3.** Promoter and enhancer activities are positively correlated but quantitatively decoupled. (A) Mean promoter and enhancer activities across all promoter-derived and distal enhancer-derived tiles and the numbers of tiles with significant promoter or enhancer activities (FDR 0.1). (B) Mean promoter and enhancer activity as a function of the distance of the center of MPRA tiles to the nearest RefSeq TSS. Negative distances correspond to tile locations upstream of the TSS. (C) Mean promoter activity for tiles cloned into the MPRA promoter test in sense (566 tiles) versus antisense (467 tiles) orientations relative to their endogenous mRNA TSS. The genomic antisense TSS position depicts the approximate location of the endogenous antisense TSS, based on a median 180-bp distance between sense and antisense TSSs (Scruggs et al. 2015). (D) Enhancer activity versus promoter activity, with each dot corresponding to one tile. Each tile’s activities are shown once for unstimulated and once for KCl-depolarized neurons. The area in which both activities fall above an empirical FDR of 0.1 is shown in gray. (E) The ratio of promoter-to-enhancer activities of tiles derived from promoter and distal enhancer loci ($P$-value from two-tailed Student’s t-test). Error bars in A–C indicate SEM for $n = 2$ independent biological replications. Promoter and enhancer activities are defined as in Figure 2B.
computed across our entire experiment or for promoter activity-only versus enhancer activity-only tiles (P > 0.05, Student’s t-tests) (Supplemental Fig. 6E,F). These results suggest that canonical core promoter elements are not responsible for the greater promoter activity of promoter-derived compared to enhancer-derived tiles.

We also considered 5’ splice signals (SpSs), whose presence at promoters but not distal enhancers could account for the unique ability of promoters to (1) initiate transcription that elongates beyond several kilobases; and (2) produce a spliced and polyadenylated mRNA instead of an eRNA (Andersson et al. 2015; Kim and Shiekhattar 2015). However, if SpSs were responsible for the elevated promoter activity of promoter-derived sequences, we would expect equivalent promoter activity from both promoter- and distal enhancer-derived sequences, because all tiles in our promoter activity test are placed upstream of the same SpSs-containing synthetic intron (Fig. 1B,C). Instead, as described above, we observe greater promoter activity from promoter-derived than from distal enhancer-derived tiles (Fig. 3A,D,E; Supplemental Fig. S7A). This elevated promoter activity is not due to the presence of additional SpSs in the promoter tiles themselves: Only 8% of promoter tiles overlap an annotated SpSS (Supplemental Table S1), and SpSS-containing tiles have no higher promoter activity than SpSS-lacking tiles (P = 0.78, Student’s t-test). We conclude that although a SpSS may be required to prevent premature termination of pre-mRNAs (Kaida et al. 2010; Almada et al. 2013; Nitini et al. 2013), it is insufficient to endow distal enhancer-derived tiles with promoter activity that is equivalent to that of promoter-derived tiles.

Promoter activity more than enhancer activity is associated with elevated CpG content and CpG-containing TF motifs

Compared to the genome at large, promoters but not most distal enhancers have elevated frequencies of CpG dinucleotides (Heintzman et al. 2007; Deaton and Bird 2011; Andersson et al. 2014), whose presence at promoters also correlates with activity in reporter assays (Landolin et al. 2010). We therefore asked whether CpG dinucleotides might contribute to the encoding of promoter activity. We find that the number of CpG dinucleotides in a sequence correlates with both promoter activity and the ratio of promoter activity. We find that the number of CpG dinucleotides positively correlates with promoter activity, when controlling for CpG content (Supplemental Table S2). Consistent with the idea that it is CpG dinucleotides rather than total G+C content that influence promoter activity, G+C content itself does not correlate positively with promoter activity, when controlling for CpG content. We then asked whether CpG dinucleotide content also correlates with promoter activity in human sequences. We performed MPRA with sequences from the human genome that are orthologous to a subset of our CREBBP-bound loci in mouse. We found that as with mouse tiles, the strongest promoter activity and the highest ratio of promoter-to-enhancer activity are encoded by CpG-rich sequences (Fig. 4B,C; Supplemental Figs. S9B, S12C,D; Supplemental Table S3). Consistent with previous observations and CREB, which are activity-regulated transcription factors with binding preferences for promoters (Knapska and Kaczmarek 2004; Greer and Greenberg 2008; Kim et al. 2010; Tang et al. 2010) and RFX, a developmental regulator (Baas et al. 2006). We also identified 20 motifs, including CREB and ETS, that are enriched in tiles with promoter activity compared to those without (Supplemental Fig. S10). Finally, we performed motif enrichment between promoter activity-only and enhancer activity-only tiles (Supplemental Fig. S8E). Enhancer activity-only tiles are enriched for AP1, a neuronal activity-regulated complex formed by a FOS/JUN heterodimer that binds neuronal activity-regulated enhancers (Malik et al. 2014). Promoter activity-only tiles are enriched for EGR1, SP1, and other GC/CpG-rich motifs. These results suggest that TF motifs could generate both enhancer and promoter activities, whereas AP1 may be biased toward the generation of enhancer activity.

Motifs differ in their intrinsic biases toward the generation of promoter versus enhancer activities

To investigate the ability of individual motifs to encode promoter and enhancer activities, we used MPRA to compare the promoter and enhancer activities of purely artificial sequences. For 32 of the most enriched motifs at CREBBP-bound genomic loci, we synthesized a sequence with tandem repeats of the motif. For each motif, we used MPRA to measure its promoter and enhancer activities (Fig. 5A; Supplemental Fig. S13A,B; Supplemental Table S4). Several motifs produce no enhancer or promoter activity at all, consistent with a previous observation that only a subset of TFs can generate transcriptional activity on their own (Smith et al. 2013). Motifs for EGR, CREB, and RFX families of TFs stand out as having the highest promoter activity, and they also generate enhancer activity. In contrast, motifs for the AP1 complex and NFY produced significant enhancer activity with little to no promoter activity. The ability of TF motifs to generate promoter activity is not simply a function of CpG content, because RFX motifs with and without CpGs both generate promoter activity. Nor is the generation of promoter activity by these artificial sequences due to the presence of
canonical core promoter elements, because neither of the two most salient core promoter motifs, the TATA box nor the INR (Frith et al. 2008), are present in any of the sequences. These results are consistent with the idea that only a subset of enhancer activity-generating motifs are also able to generate promoter activity.

We sought to reevaluate, using independent sequence contexts, our finding that promoter activity is encoded by a subset of the TF motifs that generate enhancer activity. This time we repeated each motif with three different spacers taken from inactive sequences (“spaced configuration”) (Fig. 5B; Supplemental Fig. 13C,D). We find that whereas many (15/18) of the spaced motifs produce enhancer activity, only two (RFX and ELK) produce significant promoter activity, and both of these motifs also produce promoter activity in the tandem repeat configuration (Fig. 5A). In contrast, AP1 generates enhancer activity but only minimal promoter activity in both spaced and tandem configurations. For RFX and AP1 motifs, we tested and confirmed these MPRA results using luciferase reporter assays (Supplemental Fig. S11). We conclude that some TF motifs (e.g., AP1) have a bias toward the generation of only enhancer activity, whereas others (e.g., RFX) generate both activities.

We asked whether neuronal depolarization could differentially affect enhancer versus promoter activity. To our surprise, across many artificial and genomic sequences, mean promoter activity decreases, while mean enhancer activity increases with depolarization. In genomic sequences, these changes are driven by a decrease in the promoter activities of promoter-derived loci and an increase in the enhancer activities of distal enhancer-derived loci (Supplemental Fig. S12A,B). However, in artificial sequences, some individual motifs (e.g., EGR_long, FOXN1) exhibit simultaneous increases in enhancer activity and decreases in promoter activity with depolarization. In addition, a PAX motif exhibits constitutive promoter activity but highly inducible enhancer activity (Supplemental Fig. S13). These results indicate that the relative extent of enhancer versus promoter activity generated by a DNA sequence can be regulated dynamically.

Figure 4. Promoter activity more than enhancer activity is associated with CpG content. (A) The enhancer and promoter activities of mouse genomic DNA tiles ranked by the CpG dinucleotide frequency in each tile. (B) We tiled human promoter and distal enhancer loci that are orthologous to selected mouse loci. (C) The enhancer and promoter activities of 725 human genomic tiles ranked by CpG frequency. (D) The enhancer and promoter activities of 764 reversed human genomic control tiles ranked by CpG frequency. (E) The 28 most-enriched TF motifs in tiles with a CpG fraction >0.087, compared to reversed controls. The highest-scoring matches in the JASPAR database were chosen as the TF motif names, and six additional motifs without clear TF matches are in Supplemental Table S2. The number of CpG dinucleotides in each motif is indicated in red. In A, C, and D, each value is a mean from n = 2 biological replicates. Promoter and enhancer activities are defined as in Figure 2B.
Motifs that generate promoter activity in MPRAs are present both at gene promoters and distal enhancers

We asked whether the TF motifs with intrinsic biases toward the generation of enhancer or promoter activities are correspondingly enriched at distal enhancers or gene promoters in the genome.

Among the top 10 most-enriched motifs at promoters or distal enhancers were those for CREB, AP1, and RFX (Fig. 6A; Supplemental Table S5), each of which is also transcriptionally active in MPRAs. CREB motifs are similarly enriched at distal enhancers and promoters, and RFX and AP1 motifs are biased toward distal enhancers. For these three factors, we therefore also examined TF binding at

Figure 5. Motifs differ in their intrinsic biases toward the generation of promoter versus enhancer activities. (A, top) Sequence design for tandem repeats of 32 motifs, each totaling 87 bp in length; (bottom) enhancer and promoter activities of motif tandem repeats. (B, top) Sequence design for spaced motif repeats, with 18 motifs and three spacer sequences, each totaling 87 bp in length; (bottom) enhancer and promoter activities of spaced motif sequences. The log2 of enhancer activity is shown. Stars indicate at least twofold activity with significance in a one-sided Student’s t-test (P < 0.05). Error bars are SEM of n = 3 biological replicates. Promoter and enhancer activities are defined as in Figure 2B.
promoters and distal enhancers in neurons. To assess RFX binding, we performed ChIP-seq using a MYC-tagged RFX dominant negative (RFX-DN) protein (Fig. 6B; Majigren et al. 2004; Bae et al. 2014). To assess CREB and AP1 binding in neurons, we analyzed previously published ChIP-seq data (Fig. 6B; Kim et al. 2010; Malik et al. 2014). We found CREB to be bound preferentially at gene promoters and AP1 preferentially at distal enhancers. These preferences are consistent with the idea that promoters in the genome are distinguished from distal enhancers in part by the presence of specific complements of bound TFs because CREB motifs generate more promoter activity than AP1 in MPRA (Fig. 5A). Interestingly, however, RFX motifs are preferentially enriched at enhancers, despite the potency of RFX in generating promoter activity, suggesting that RFX may be required for the promoter activity of distal enhancers.

**Discussion**

We describe here a high-throughput comparison of the promoter and enhancer activities of short, defined genomic sequence fragments. Sequences that co-encode both promoter and enhancer activities are the rule rather than the exception, suggesting that enhancer activity shares with promoter activity a common underlying mechanism. One possibility is that enhancer function requires eRNA synthesis (Ørom et al. 2010; Kaikkonen et al. 2013; Li et al. 2013; Mousavi et al. 2013; Hsieh et al. 2014; Schaukowitch et al. 2014). Consistent with this idea, eRNA synthesis at distal enhancers is correlated with transcription of nearby genes (De Santa et al. 2010; Kim et al. 2010; Rada-Iglesias et al. 2010; Amer et al. 2015). However, we believe that defined sequences do not produce enhancer and promoter activities at a fixed ratio. This decoupling of enhancer and promoter activity suggests that enhancers can activate promoters without initiating transcription.

Our work builds on recent findings that distal enhancers and gene promoters have remarkably similar genomic architectures, each with divergent transcription from nucleosome-depleted cores (Core et al. 2014; Duttke et al. 2015; Scruggs et al. 2015). At gene promoters, the presence of a 5’ splicing signal protects against premature termination (Kaida et al. 2010; Almada et al. 2013; Ntini et al. 2013). Its presence may be the sole reason that pre-mRNA transcription, but not upstream antisense promoter transcription or eRNA transcription, extends beyond a few kilobases from TSSs (Anderson et al. 2015; Kim and Shiekhattar 2015). Consistent with the idea that this promoter directionality is conferred by a splicing signal, the promoter activity we measure here in the absence of exogenous splicing signals is intrinsically bidirectional. However, our results also make clear that independently of splicing signals, gene promoters generate more promoter activity than distal enhancers.

Our results suggest that promoter activity requires little in the way of nucleotide structure because it can be generated by endogenous and artificial sequences that lack canonical core promoter elements. This flexibility contrasts with the demonstrated requirement for core promoter elements at some promoters (Smale and Kadonaga 2003; Mogno et al. 2010; Kadonaga 2012; Lubliner et al. 2015). Our results suggest instead that promoter activity at CREBBP-bound loci is best understood as a function of the specific complement of transcriptional activators that are bound (Fig. 7E), although we note that alternative mechanisms might predominate at non-CREBBP-bound loci or in other cell types. The greater promoter activity generated by some TFs (e.g., RFX) compared to others (e.g., AP1) may depend on recruitment of
distinct coactivator or preinitiation complexes that are predisposed toward the generation of promoter activity. This idea fits also with the observation that TFs form complementation groups within which they may substitute for each other in the generation of transcriptional activity (Stampfel et al. 2015). The comparison of complexes recruited by RFX and AP1 could identify a generalized mechanism that distinguishes promoter from enhancer activity.

Our conclusions pertain to a specific experimental context—CREBBP-bound sequences in cortical neurons—outside of which they may differ. Generally, transcription regulation is context dependent. The same regulatory sequences likely function differently in other cell types (Heintzman et al. 2009) or with a different minimal promoter (Zabidi et al. 2015). The enhancer and promoter activities that we observe may be influenced by cryptic enhancers present in our specific MPRA plasmids, by the short tile lengths, or by the viral transduction used to introduce libraries into cells. We speculate that the ability of a specific transcriptional activator or coactivator to drive promoter activity may be context dependent, as is the ability to encode transcriptional activation generally (Smith et al. 2013).

Methods

Genomic tiles synthesized (TN03 library)

We analyzed tiles from 487 mouse genomic loci. For each distal enhancer or promoter locus, we synthesized 11 tiles centered on the called CREBBP peak, whose center was defined as the midpoint of the peak (Kim et al. 2010). For the mouse sequences, we inserted all tiles whose nearest annotated TSS is on the positive strand into our reporter assay in the “sense” orientation. We inserted all mouse tiles whose nearest annotated TSS is on the negative strand into our reporter assay in the “antisense” orientation. We also analyzed tiles from 217 human genomic loci whose nearest RefSeq TSS is transcribed on the positive genomic strand, each inserted into reporter plasmids in the sense orientation. Our nonsense controls were 498 sequences derived from human genomic loci whose sequences were reversed without complementation. Design and analysis was performed using custom scripts in R (Supplemental Methods) (R Core Team 2014).

Processing of MPRA libraries

Genomic (TN03 library), mutated genomic (TN05 library), and artificial (TN04 library) sequences were synthesized by CustomArray, Inc., and amplified by emulsion PCR (Supplemental Methods). Each oligo library was designed to contain 12,472 oligonucleotides. For TN04, sequence barcodes were added during synthesis, with 9–11 distinct barcodes per test sequence (Melnikov et al. 2012). For TN03 and TN05, random 18-bp barcodes were added during amplification (Supplemental Fig. S1). Amplified sequences were then cloned into vector pTA011 using SfiI sites. DNA was prepared from ~200,000 colonies for TN03 and TN05, in a deliberate bottleneck designed to limit the number of total barcodes in the plasmid library. For TN04, 5 × 10^6 to 10 × 10^6 colonies were prepared. For the promoter test, a synthetic intron-GFP-KAN ORF containing fragment was cloned into the isolated libraries (Supplemental Fig S1). For the enhancer test, barcoded sequences from pTA011 were excised and cloned into pTA008 using SfiI sites (Supplemental Fig. S1). After each of these ligations, DNA was prepared from 5 × 10^6 to 10 × 10^6 colonies for TN03, TN04, and TN05 to maintain library complexity. Our AAV MPRA strategy

Figure 7. The promoter activity of distal enhancers requires RFX motifs. (A) Mean promoter activity of distal enhancer tiles that have zero (“−”) or at least two (“+”) occurrences of RFX, CREB, and AP1 motifs; (*) P < 0.05 by Student’s t-test. Promoter activity is defined and normalized as in Figure 2B, and each tile’s mean promoter activity from two biological replicates is averaged across all “+” and “−” tiles. The apparent lack of error bars in “−” tiles is due to the mean and SEM being computed across a large number of tiles, because most tiles lack these motifs. (B) Mutations of RFX motif occurrences within distal enhancer sequences. Each motif occurrence was altered in four different ways. A gray portion indicates a nucleotide that does not match the corresponding motif represented at right. Motifs are from HT SELEX position frequency matrices for RFX5 (Jolma et al. 2013). (C) Ratios of mutant to endogenous promoter activity. (D) Same as C but for RFX dominant negative-transduced neurons. (E) Model: Transcriptional activators and coactivators determine the relative extent of enhancer and promoter activities.
is similar to a recently published one (Supplemental Methods; Supplemental Fig. S28; Shen et al. 2015).

Data access

The sequencing data from this study are described in Supplemental Table S6 and have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE77213.

Acknowledgments

We thank Tarjei Mikkelsen, Alex Melnikov, and Connie Cepko for sharing reagents and offering advice; Sarah Boswell, Mike Springer, Nicole DeStefano, Kelsey Tyssowski, Jen Harrow, Michal Kosicki, Ilias Georgopolous-Soares, Daniel Gaffney, Leopold Parts, and Tae-Kyung Kim for helpful discussions and critical reading of the manuscript; Manolis Kellis for support and advice; Daniel Wiegand and Brian Turczyk for assistance with oligonucleotide synthesis; and Clarissa Scholes for making the model in Figure 7. This work was funded by the National Institutes of Health, National Institute of Mental Health (R01 MH1101528-01).


References


Received January 25, 2016; accepted in revised form June 14, 2016.
High-throughput functional comparison of promoter and enhancer activities


Genome Res. published online June 16, 2016
Access the most recent version at doi:10.1101/gr.204834.116

Supplemental Material
http://genome.cshlp.org/content/suppl/2016/07/04/gr.204834.116.DC1.html

Published online June 16, 2016 in advance of the print journal.

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

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

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2016 Nguyen et al.; Published by Cold Spring Harbor Laboratory Press