High-throughput functional comparison of promoter and enhancer activities

Thomas A. Nguyen,1,5 Richard D. Jones,1,5 Andrew R. Snively,1 Andreas R. Pfenning,2 Rory Kirchner,3 Martin Hemberg,4 and Jesse M. Gray1

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

[Supplemental material is available for this article.]

Regulation of metazoan gene transcription is accomplished through at least two distinct biochemical activities encoded in DNA. DNA sequences with promoter activity initiate transcription at self-contained transcription start sites (TSSs) (Lenhard et al. 2012). DNA sequences with enhancer activity activate transcription initiation at promoters (Grosschedl and Birnstiel 1980; Banerji et al. 1981). Promoter and enhancer activities were once assumed to be encoded at distinct locations in metazoan genomes (Blackwood and Kadonaga 1998; Levine and Tjian 2003). Consistent with this view, gene promoters and the mRNA TSS-distal (“distal”) enhancers that regulate them are distinguishable based on their stereotypical chromatin signatures. Distal enhancers are distinguished from gene promoters based on enrichment of histone H3 monomethylated on Lys4 (H3K4me1), reduced enrichment of H3K4me3, and the absence of TSSs for spliced, polyadenylated RNAs within ∼1 kb (Heintzman et al. 2007; Andersson et al. 2014). Gene promoters are distinguished from distal enhancers by the presence of mRNA TSSs and enrichment of H3K4me3 (Heintzman et al. 2007).

Although gene promoters and distal enhancers are generally distinguishable based on chromatin signatures, they do not differ categorically in their ability to encode enhancer or promoter activities. Distal enhancers encode promoter activity, initiating synthesis of stable mRNAs but of enhancer RNAs (eRNAs) that are unstable, unspliced, and nonpolyadenylated (De Santa et al. 2010; Kim et al. 2010; Andersson et al. 2014). Conversely, gene and lncRNA promoter sequences also encode enhancer activity (Webster et al. 1988; Li et al. 2012; Zabidi et al. 2015; Paralkar et al. 2016); although, the frequency at which this occurs outside of Drosophila is unknown. These findings suggest that within metazoan genomes, sequences that encode enhancer and promoter activities may be overlapping or even identical (Weingarten-Gabgaly and Segal 2014; Andersson et al. 2015; Kim and Shiekhattar 2015; Li et al. 2016). However, because the promoter and enhancer activities of only a few defined sequences have been directly compared (Moreau et al. 1981; Serfling et al. 1985; Treisman and Maniatis 1985; Bienz and Pelham 1986), the fundamental relationships between enhancer activity, promoter activity, and the sequences that encode them are unknown.

Results

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 neuronal activation (Malik et al. 2014). We also required each locus to have a human
ortholog that is DNase accessible in fetal brain and contains one or more SNPs. The goal with these selection criteria was to identify a focused set of mouse genomic loci that are active in neurons, that can be compared to their human ortholog, and that are maximally relevant to human health. In mouse neurons, the distal enhancer loci express eRNAs and are enriched in H3K4me3 (but not H3K4me1) (Supplemental Table S1), and distal enhancer-derived sequences can be compared to their human orthologs, and that are maximally relevant to human health. 

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 MPRAs 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 MPRAs could detect the desired activities. After synthesis and cloning, we were able to assess enhancer and promoter activities for 1957 tiles (31%) in all samples. This total represents an average of 4.0 tiles per locus, with 924 distal enhancer-derived and 1033 promoter-derived tiles. As expected, promoter-derived sequences harbor promoter activity (19% of tiles, 37% of loci; FDR 0.1) (Supplemental Table S1), and distal enhancer-derived sequences harbor enhancer activity (13% of tiles, 22% of loci; FDR 0.1) (Supplemental Table S1). Because distal enhancers initiate synthesis of eRNAs, we also expected distal enhancer-derived sequences to harbor promoter activity, and they do (4% of tiles, 10% of loci; FDR 0.1) (Supplemental Table S1). The overall observation that only a subset of loci is transcriptionally active is unsurprising, given (1) previous high-throughput reporter results in which a similarly low fraction of sequences were active (20%–24% versus our 22%) (White et al. 2013); (2) the presence in our library of many “off-center” tiles that likely lack relevant regulatory elements (Supplemental Figs. S5, S6); and (3) a tile length that may in some cases be too short. We found reassuringly that the promoter activity of MPRAs tiles correlates with the abundance of RNA transcripts that initiate within corresponding genomic sequences, as detected by CAGE tag frequencies in neurons.
tissue-specific MPRA datasets. We compared promoter activity across these tissues with a 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 NFPshort (diamond) motifs with spacers TTATTTTAAGA (RFX, MYBL2) and CCCGCGCTGCC (AP1, NFPshort). (E–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.

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 $p = 0.45$ and $P < 10^{-16}$, asymptotic $t$ approximation) (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. S14G,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 tiles 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 promoter activities are coregulated in a tissue-dependent manner. In many cases, 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 ∼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.

Our data also provide a high-resolution genomic 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 ∼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 (∼500 to ∼40 bp) that bind transcriptional activators (Gurante 1984). It also colocalizes with traditional core promoter regions (∼40 to ∼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 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 ∼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 $p = 0.45$ and $P < 10^{-16}$, asymptotic $t$ approximation) (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. S14G,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 tiles 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 promoter activities are coregulated in a tissue-dependent manner. In many cases, 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 ∼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.
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 (TATAWAWR) or the INR (YYANWYY or YYRRWYY) have no higher promoter activity than those that lack them, whether...
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 $S$' splice signals (SpSSs), 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 SpSSs 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 SpSSs 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-to-enhancer activity, even after controlling for total G+C nucleotide content using partial correlation (partial $R = 0.33$ and $R = 0.21$, $P < 10^{-20}$, Student's $t$-tests) (Fig. 4A; Supplemental Figs. S9A, S12C,D; Supplemental Table S3). 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 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, S12E,F). These results suggest that CpG dinucleotides positively influence promoter activity more than enhancer activity.

Elevated CpG content is sufficient to establish promoter-associated chromatin states (Blackledge et al. 2010; Wu et al. 2011; Clouaire et al. 2012; Fenouil et al. 2012; Wachter et al. 2014). We therefore asked whether CpG richness alone is also sufficient to encode promoter activity. For this purpose, we created a set of human promoter tiles whose sequences were reversed without complementation (i.e., TGACGTCA → ACTGCAGT). This manipulation preserves G+C content but destroys transcription factor (TF) binding motifs. It also destroys individual CpGs and GpCs via inter-conversion. However, average CpG content is actually increased because there are more GpCs than CpGs in our genomic sequences (0.08% CpGs versus 0.05% GpCs, consistent with CpG depletion in the genome). We find that sequence reversal categorically abolishes promoter activity (Fig. 4D; Supplemental Fig. S11E,F). In a complementary experiment, we also used MPRA to ask whether artificial dinucleotide repeats in the absence of motifs could generate promoter activity. We compared 20 ApG, CpA, or CpG repeats on a fixed background (Supplemental Fig. S9C). None of the sequences generated significant promoter activity ($P > 0.7$, one-tailed Student's $t$-test). Thus, high CpG content per se is insufficient to encode promoter activity but may be required together with TF motifs.

To identify TF motifs that might function as specific determinants of promoter activity, we performed motif enrichment analyses. First, we identified 28 motifs enriched in the 149 tiles with >8.76% CpGs per dinucleotide (most of which produce promoter activity) (Supplemental Fig. S9D), relative to reversed control sequences (Bonferroni-adjusted $P < 0.05$, binomial test) (Fig. 4E; Supplemental Table S2). Enriched motifs include those for EGR and CREB, which are activity-regulated transcription factors with binding preferences for promoters (Knapiska and Kazmarek 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 the possibility that RFX, CREB, EGR, and other 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 API 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 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.
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.
Motifs and their corresponding transcription factors exhibit specific preferences for promoters or distal enhancers in the genome. (A) Motifs with at least twofold enrichment at mouse gene promoters or distal CREBBP-bound enhancers, relative to corresponding flanking sequences. Motifs corresponding to CREB, RFX, and AP1 are labeled for the 10 most enriched motifs at gene promoters or distal enhancers. For colored motifs, enrichment at both classes of loci is significant ($P < 10^{-9}$, Bonferroni-corrected binomial test). (B) Percentages of cortical neuron CREB, RFX, and AP1 ChIP-seq peaks located at gene promoters (top) and distal enhancers (bottom). Antibodies used were anti-CREB, anti-FOS (AP1), and anti-MYC (with MYC-tagged RFX dominant negative): (∗) $P < 10^{-14}$ by a hypergeometric test.

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 S’ splicing signal protects against premature termination (Kaid et al. 2010; Almada et al. 2013; Nitini 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 endogenous 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; Lublinier 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
at least two (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 control promoter activity using MPRAs, with median and 25th–75th percentiles shown as bars and boxes. A ratio of one (no change in promoter activity) is indicated with a vertical red line. Promoter activity is defined as in Figure 2B. The 20 control tiles include 11 distal enhancer tiles, shown separately in Supplemental Figure S5F. (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.

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 pTAN01 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 pTAN01 were excised and cloned into pTAN08 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...
is similar to a recently published one (Supplemental Methods; Supplemental Fig. S2B; Shen et al. 2015).

Data access
The sequencing data from this study are described in Supplemental Table 36 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 Meiniok, and Connie Cepko for sharing reagents and offering advice; Sarah Boswell, Mike Springer, Nicholas DeStefano, Kelsey Tyssowski, Jenn Harrow, Michal Kosicki, Ilia Georgakopoulous-Soares, Daniel Gaffney, Leopold Parts, and Tae-Kyun Gum 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 MH101528-01).

Author contributions: Conceptualization: J.M.G. and M.H.; methodology: T.A.N., A.R.S., A.R.P., and M.H.; investigation: T.A.N., A.R.S., R.D.J., and M.H.; resources: R.K.; supervision: J.M.G.; review and editing: R.D.J. and J.M.G.; funding acquisition: R.K.; writing (original draft): J.M.G.; review and editing: R.D.J. and J.M.G.; funding acquisition: J.M.G. and M.H.; resources: R.K.; supervision: J.M.G.

References
Almadne AE, Wu X, Kritz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature 499: 360–363.
Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging interactions provide a topological basis for transcription regulation. Trends Genet 28: 162–171.
Bartfeld A, Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022.
Deaton AM, Bird A. 2013. Genes Dev 27: 1311–1320.
Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022.
Dutcke HSC, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U. 2015. Human promoters are intrinsically directional. Mol Cell 57: 674–684.
FANTOM Consortium and the RIKEN PMI and CLST (DGT). 2014. A promoter-level mammalian expression atlas. Nature 507: 462–470.
Fenouil R, Cauchi P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Skalnik D, Bird A. 2012. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res 22: 2399–2408.
Fisch TM, Pywres R, Roeder RG. 1987. C-fos sequence necessary for basal expression and induction by epidermal growth factor, 12-O-tetradecanoyl phorbol-13-acetate and the calcium ionophore. Mol Cell Biol 7: 3490–3502.
Frith MC, Valen E, Krogh A, Hayashizaki Y, Campani P, Sandelin A. 2008. A code for transcription initiation in mammalian genomes. Genome Res 18: 384–390.
Gree PL, Greenberg ME. 2008. From synapse to nucleus: calcium-dependent gene transcription in the control of synapse development and function. Neuron 59: 846–860.
Hede A, Birnstiel ML. 1980. Spacer DNA sequences upstream of the proximal gene are necessary for efficient mammalian transcription. Nature 280: 274–277.
Heintz ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature 447: 391–396.
Heintz ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee IK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108–112.
Hsieh CI, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee GS, Chen S, et al. 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci USA 111: 7319–7324.
Holm C, Vartanian F, Lagnado S, Purcell AM, Poirier V, Shen Y, Cavadini P, Illingworth R, Heintz S, Kadonaga JT, Ohler U. 2015. A deficiency in RFX3 causes hydrocephalus associated with abnormal differentiation of ependymal cells. Eur J Neurosci 42: 112–123.
Jolma A, Yang Y, Whitington T, Toivonen J, Nitta RR, Rast P, Morganova E, Enge M, Tailpale M, Wei G, et al. 2013. DNA-binding specificities of human transcription factors. Cell 152: 327–339.
Jolma A, Yang Y, Whitington T, Toivonen J, Nitta RR, Rast P, Morganova E, Enge M, Tailpale M, Wei G, et al. 2013. DNA-binding specificities of human transcription factors. Cell 152: 327–339.
Kadonaga JT. 2012. Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol 1: 40–51.
Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyszus G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468: 664–668.
Kikkonen MU, Spinn NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, et al. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer activation. Mol Cell 51: 310–323.
Kim TK, Shiekhattar R. 2013. Architectural and functional commonalities between enhancers and promoters. Cell 152: 948–959.
Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptevicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 182–187.
Knopkins E, Kaczmarek L. 2004. A gene for neuronal plasticity in the mammalian brain. Zbl Physiol 190: 148–151.
Knopkins E, Kaczmarek L. 2004. A gene for neuronal plasticity in the mammalian brain. Zbl Physiol 190: 148–151.
Landolin JM, Johnson DS, Trinklein ND, Aldred SF, Medina C, Shulha H, Weng Z, Myers RM. 2010. Sequence features that drive human promoter function and tissue specificity. Genome Res 20: 890–898.
Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet 13: 233–245.
Levine M, Tjian R. 2003. Transcription regulation and animal diversity. Nature 424: 147–151.
Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148: 84–98.
Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Meinkeur D, Zhang J, Ogki K, Song X, et al. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498: 516–520.
Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet 17: 207–223.
Lubliner S, Regev I, Lotan-Popman M, Edelstein S, Weinberger A, Segal E. 2015. Core promoter sequence in yeast is a major determinant of expression level. Genome Res 25: 1008–1017.
Maligren S, Sur I, Nilsson M, Toftgård R. 2004. Involvement of RFX proteins in transcriptional activation from a Ras-responsive enhancer element. *Arch Dermatol Res* **296**: 482–489.

Malik AN, Viebuchen T, Hemberg M, Rubin AA, Ling E, Couch CH, Stroud H, Spiegel I, Fair KK, Harmin DA, et al. 2014. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat Neurosci* **17**: 1330–1339.

Melnikov A, Murugan A, Zhang X, Tesleanu T, Wang L, Rogov P, Feizi S, Gurtik A, Callan CG, Kinney JB, et al. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**: 271–277.

Mogno I, Vallania F, Mitra RD, Cohen BA. 2010. TATA is a modular component of synthetic promoters. *Genome Res* **20**: 1391–1397.

Moreau P, Hen R, Wasylyk B, Everett R, Gaub MF, Chambron P. 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* **9**: 6947–6968.

Mousavi K, Zare H, Dell’Orso S, Gronvold L, Gutierrez-Cruz G, Derfoul A, Hager GI, Sartorelli V. 2013. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* **51**: 606–617.

Ntini E, Jarvelin Al, Bornholdt J, Chen Y, Boyd M, Jørgensen M, Andersson R, Hoof I, Schein A, Andersen PR, et al. 2013. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* **20**: 923–928.

Ørom UA, Derrien T, Beranger M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytynicki M, Notredame C, Huang Q, et al. 2010. Long noncoding RNAs from its associated cis element. *Mol Cell* **62**: 104–110.

R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rada-Iglesias A, Baijai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2010. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–285.

Schaukowitch K, Foo JY, Liu X, Watts JK, Martinez C, Kim TK. 2014. Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell* **56**: 29–42.

Scruggs RS, Gilchrist DA, Nechaev S, Muse GW, BurkhOLDER A, Fargo DC, Adelman K. 2015. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol Cell* **58**: 1101–1112.

Serfling E, Lübke A, Dorsch-Häuser K, Schaffner W. 1985. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO J* **4**: 3851–3859.

Shen SQ, Myers CA, Hughes AE, Byrne LC, Flannery JC, Corbo JC. 2015. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res* **25**: 238–255.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449–479.

Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovscharenko I, Abitou N. 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* **45**: 1021–1028.

Stampfle G, Kazmar T, Frank O, Wienerroither S, Reiter F, Stark A. 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**: 147–151.

Tang C, Shi X, Wang W, Zhou D, Tu J, Xie X, Ge Q, Xiao PF, Sun X, Lu Z. 2010. Global analysis of in vivo EGR1-binding sites in erythroleukemia cell using chromatin immunoprecipitation and massively parallel sequencing. *Electrophoresis* **31**: 2936–2943.

Treisman R, Maniatis T. 1985. Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. *Nature* **315**: 73–75.

Wachter E, Quante T, Merusi C, Arcezewski A, Stewart F, Webb S, Bird A. 2014. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife* **3**: e03397.

Webster N, Jin JR, Green S, Hollis M, Chambron P. 1988. The yeast UASy is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell* **52**: 169–178.

Weingarten-Gabbay S, Segal E. 2014. A shared architecture for promoters and enhancers. *Nat Genet* **46**: 1253–1254.

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* **110**: 11952–11957.

Wu H, D’Alesio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **472**: 389–393.

Zahid MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**: 556–559.

Received January 25, 2016; accepted in revised form June 14, 2016.