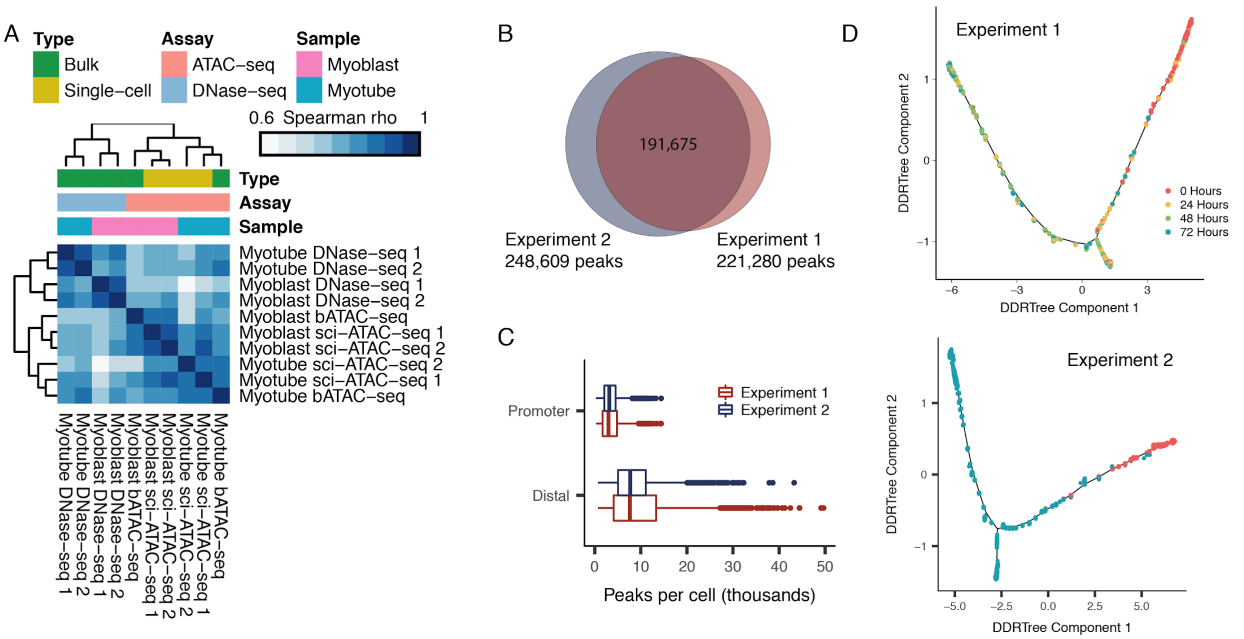
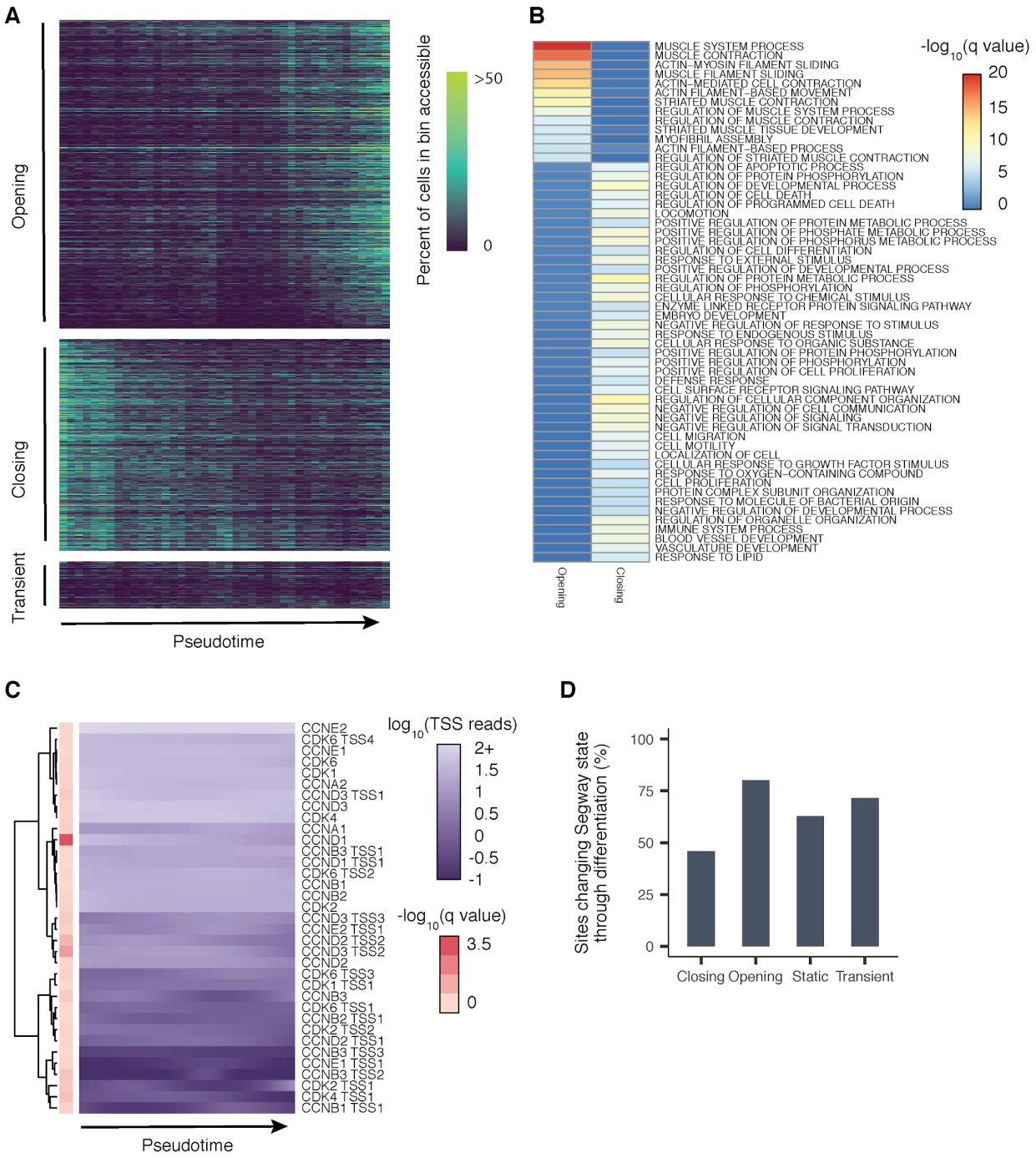
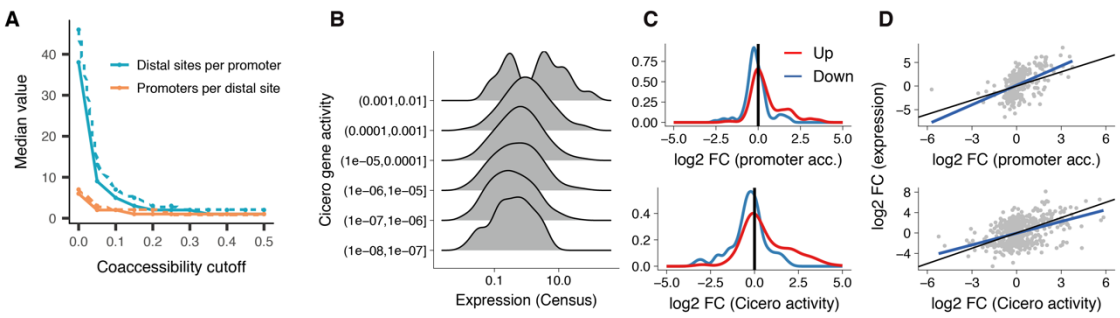


Supplemental Figure 1. Chromatin accessibility profiles of differentiating myoblasts are highly reproducible, related to Figure 1. A) Spearman correlation heatmap between pairs of chromatin accessibility profiles as measured by bulk ATAC-seq, DNase-seq, and aggregated sci-ATAC-seq from 0 and 72 hours. MACS was used to call new peaks on each dataset; these peaks were merged, and reads were counted in each peak from each dataset. These counts were used to calculate Spearman correlations. **B**) Venn diagram illustrating reproducibility in MACS2 peaks calls between independent sci-ATAC-seq experiments. Peaks in the intersection correspond to DNA elements called in both experiments. **C)** Boxplot of the number of MACS-called sci-ATAC-seq peaks per cell. Promoter-proximal peaks are peaks intersecting the first 500 base pairs upstream of a transcription start site (see Methods). Distal peaks are all other peaks. **D**) Monocle trajectories on each of the sci-ATAC-seq experiments. The top panel is identical to Figure 1C and included for comparison purposes.

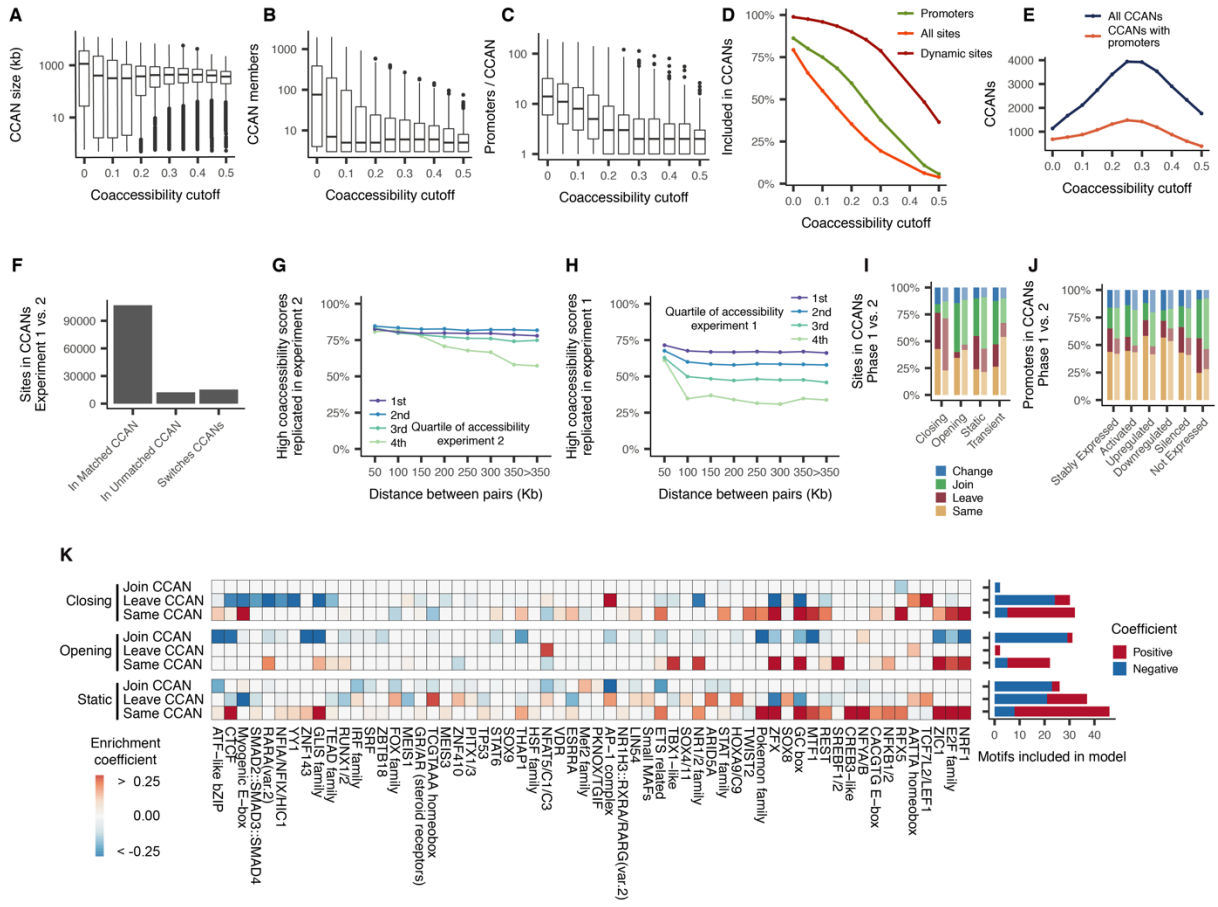


Supplemental Figure 2. DNA elements that open during differentiation are enriched for muscle related promoters, related to Figure 2. A) Heatmap of accessibility across pseudotime. Color represents the percent of cells per bin that are accessible at a given DNA element. Each row indicates a different DNA element, each column represents a bin of approximately even numbers of cells divided by pseudotime. Rows are in the same order as Figure 2A.. **B**) Gene set enrichment analysis of significantly opening and closing accessible sites. Adjusted p-values were

computed using a hypergeometric test. Terms shown are all sites with an adjusted p-value < 1e-6 in either the opening set or the closing set. Color represent the $-\log_{10}$ adjusted p-value. Sites are ordered by the $-\log_{10}$ adjusted p-value of the opening set. **C**) Smoothed pseudotime-dependent accessibility curves, generated by a negative binomial regression of each for a set of selected cell cycle relevant genes. Each row indicates a different DNA element. Annotation column represents the -log10 adjusted p-value for the test of differential accessibility across pseudotime. For visualization, fitted curve range was capped at 100. **D**) Percent of dynamic and static sites with changing Segway state assignment from myoblast to myotube.

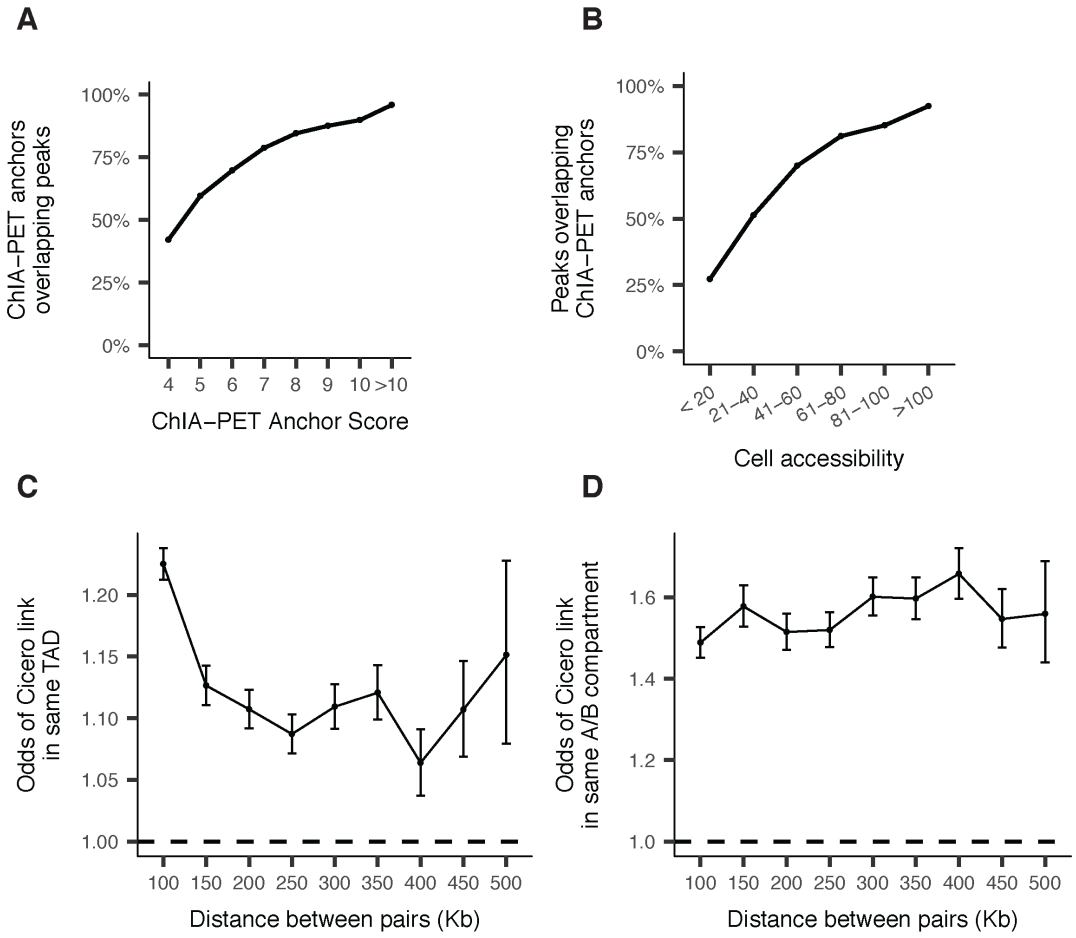


Supplemental Figure 3. Cicero gene activity scores correlate with gene expression, related to Figure 3. A) The median number of linked distal sites per promoter and promoters per distal site as a function of the co-accessibility threshold of the links considered. Dashed lines indicate experiment 2. Solid lines indicate experiment 1. **B**) Average Cicero gene activity scores across cells in phase 1 compared to their average expression from sc-RNA-seq libraries from myoblasts. Cicero gene activity scores were computed by summing the reads falling in distal sites linked to a gene's promoter (see Methods for details). **C**) Top panel: Log_2 fold-changes in mean accessibility at gene promoters for genes that are significantly up- (red) or down-regulated (blue) between 0 and 72 hours as measured by sc-RNA-seq. Bottom panel: corresponding changes in Cicero gene activity scores. **D**) Top panel: comparison of log_2 fold changes between expression and promoter accessibility. Bottom panel: fold changes in expression versus changes in Cicero gene activity scores. Black lines indicate perfect concordance between log_2 fold changes, while blue lines indicate linear regressions between Cicero activity or promoter accessibility and expression.

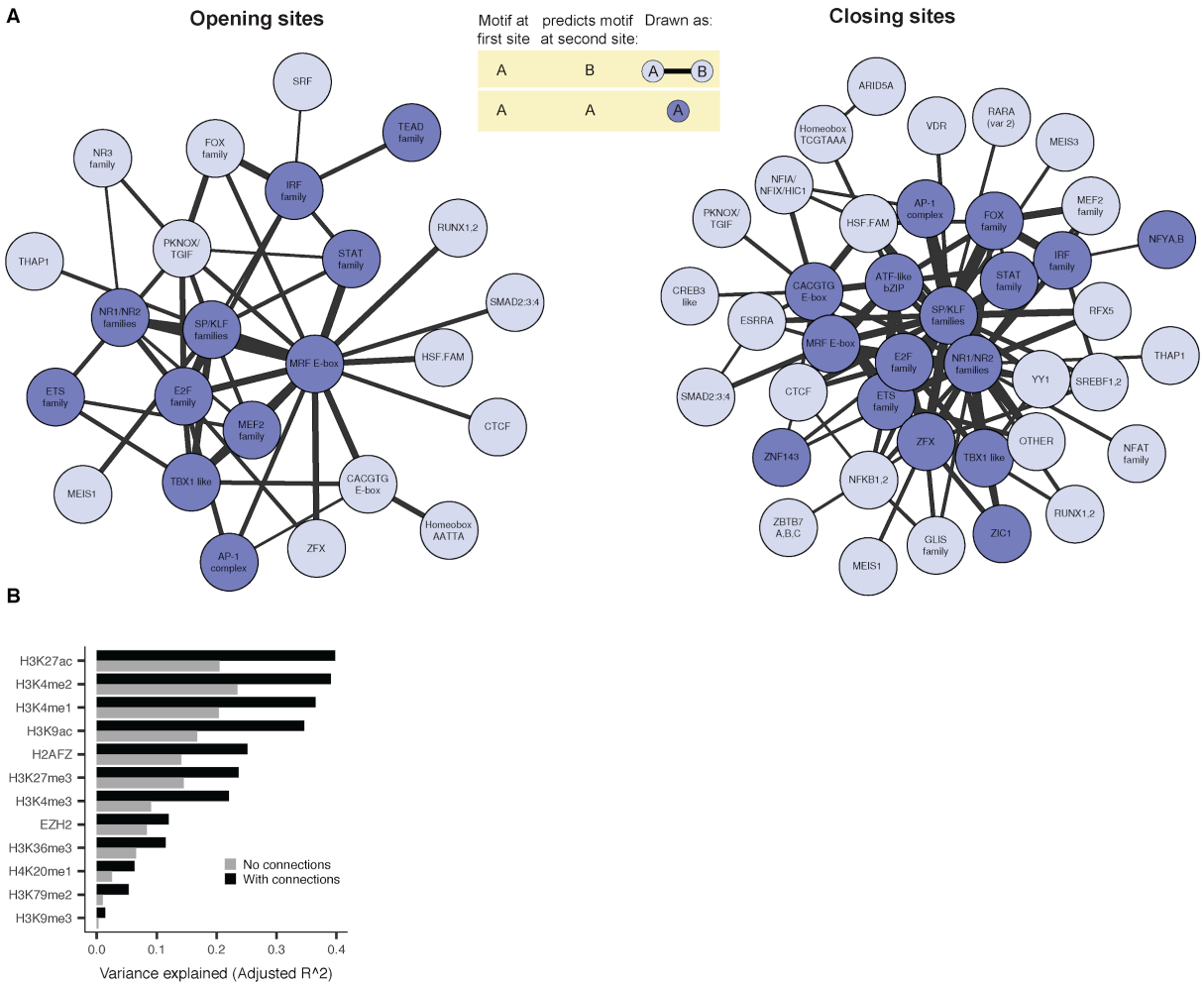


Supplemental Figure 4. Cis-coaccessibility networks (CCANs) maintain properties at varying cutoffs, related to Figure 3. **A**) Boxplots of the length in the linear genome (in kilobases) of CCANs formed at varying thresholds of co-accessibility in experiment 1. CCANs are defined as groups with 3 or more co-accessible DNA elements identified with Louvain community detection. Prior to running Louvain, connections below the indicated Cicero co-accessibility score are excluded (see Methods for details). **B**) Boxplots of the number of sites in CCANs formed at varying thresholds of co-accessibility. **C**) Boxplots of the number of expressed gene (at a level of 10 transcripts / cell on average in scRNA-seq) promoters per CCAN at increasing co-accessibility score cutoffs. **D**) Percent of sites recruited into a CCAN at increasing co-accessibility score cutoff. Colors represent subsets of sites: green represents promoters for genes that are expressed; orange and red represent sites that are accessible and differentially accessible across pseudotime, respectively. **E**) Number of CCANs identified with varying co-accessibility cutoffs. Blue series shows the total number of CCANs, orange shows the number of CCANs that include a promoter of at least 1 detectably expressed gene. **F**) Number of sites that linked into CCANs that are matched in experiment 1 and experiment 2 by a maximum weighted bipartite matching method (see Methods). Also shown are sites that are linked into CCANs that are not matched at all and sites that are linked into CCANs that are matched in experiment 1 and experiment 2, but not to one another. **G**) Fraction of pairs of sites linked in experiment 1 at co-accessibility > 0.25 that are also linked in experiment 2 at co-accessibility > 0. Colors indicate the quartile of accessibility in experiment 2. **H**) Reciprocal plot to panel I, examining sites linked at co-accessibility > 0.25 in experiment 2 and > 0 in experiment 1. Colors indicate the quartile of accessibility in experiment 1. **I**) Sites linked into CCANs found in both phase 1 and phase 2 by

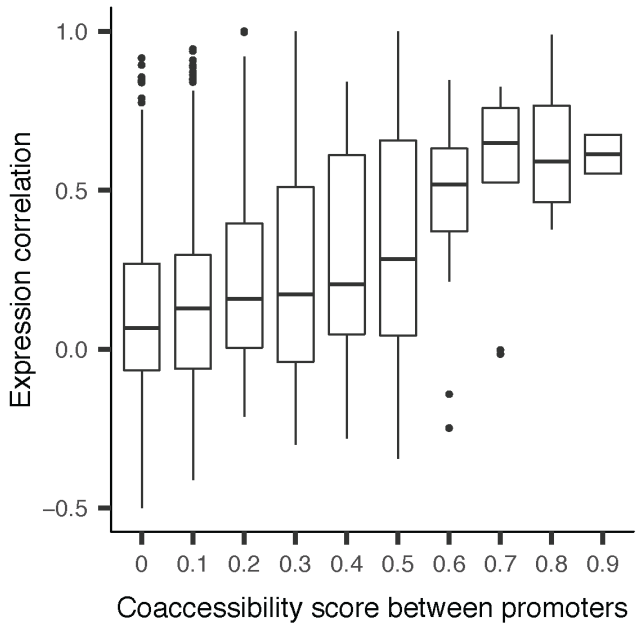
maximum matching, subdivided by those that were linked into the CCAN in both phases (yellow), those linked in phase 1 but unlinked in phase 2 (red), those unlinked in phase 1 but linked in phase 2 (green), and those linked into different (i.e. non-matched) CCANs in the two phases. The four groups of sites from Figure 2b-d are considered. The left bar in each group corresponds to experiment 1, while the right bar corresponds to experiment 2. **J)** Similar to panel J but considering only promoters: groups are promoters of genes that are "stably" expressed at an unchanged level, those that are silent in myoblasts but expressed in myotubes ("activated"), and those expressed in both myoblasts and myotubes, but higher in myotubes ("upregulated"). Similarly, we show promoters of genes that are downregulated or fully silenced in myotubes, as well as those that are not detectably expressed (at a level of 10 transcripts / cell on average) in either cell type. **K**) A heatmap of regression coefficients from three multinomial elastic net regression analyses that predict whether a site will join, leave, or remain linked into its CCAN during differentiation on the basis of varying sequence motifs. Coefficients were capped at -0.25 and 0.25 for visualization. Only sites with consistent CCAN dynamics across both experiments were included in the models. The number of positive and negative coefficients surviving regularization in each model are shown in the barplot to the right. Regression was performed using the glmnet package in R and the regularization parameter was chosen that produced the minimum mean cross-validated error after 10-fold cross validation.



Supplemental Figure 5. ChIA-PET anchors are concordant with sci-ATAC-seq peaks, related to Figure 4. **A**) Percent of pol II ChIA-PET anchors within 1 kb of an sci-ATAC-seq peak as a function of ChIA-PET anchor score provided by Tang et. al. (2015). **B**) Percent of sci-ATAC-seq peaks within 1 kb of pol II ChIA-PET anchors as a function of overall cell accessibility (number of cells where the peak is accessible). **C**) Odds ratio that both members of Cicero linked pairs (co-accessibility > 0) are in the same TAD, compared with unlinked pairs at the same distance. GM12878 TAD calls are from (Rao et al., 2014). **D**) Odds ratio that both members of Cicero linked pairs (co-accessibility > 0) are in the same A/B compartment, compared with unlinked pairs at the same distance. A/B compartment calls are from (Fortin and Hansen, 2015).



Supplemental Figure 6. DNA motifs predict motifs in Cicero-linked sites, related to Figure 5. A) Motifs in accessible sites predict motif content of Cicero-linked sites. The network summarizes a graphical model that captures how occurrences of motifs in pairs of sites predict whether they are connected. Each motif is connected to the motifs it suggests will exist in one or more connected sites. A motif that predicts itself in a connected site is shown in dark blue. If motif "A" at a distal site predicts that "B" will be found at a promoter, and symmetrically "B" at a distal site suggests "A" will be found at a promoter, they are connected with a black line, with a width proportional to the strength of the co-accessibility. Asymmetric motif relationships are not shown. **B)** Variance explained by a linear model that aims to predict $\log_2$-transformed fold changes in the listed ChIP-seq read counts between myoblasts and myotubes. Two models are considered. The first, with performance indicated as gray bars, uses a site's accessibility and MYOD binding status. The second, indicated as black bars, augments the first with accessibility and MYOD at linked sites. The predictor for MYOD at linked sites was significant (p-value < .05) for all augmented models. See Methods for more details.



Supplemental Figure 7. Expression correlation increases with increasing coaccessibility, related to Figure 6. Correlation in expression among linked differentially expressed genes. Boxplots of the cell-wise correlation between gene expression among pairs of differentially expressed genes whose promoters have different Cicero co-accessibility scores.

Supplemental Table 1. sgRNA sequences targeting *Meis1*, *PBX1* and non-targeting controls, related to STAR Methods

sgRNA	Gene targeted	sgRNA sequence
Meis1_1	*Meis1*	TACTTGTACCCCCCGCGAGC
Meis1_2	*Meis1*	CACAGCTCATACCAACGCCA
Meis1_3	*Meis1*	GGTGGCCACACGTCACACAG
Meis1_4	*Meis1*	ACTCGTTCAGGAGGAACCCC
PBX1_1	*PBX1*	GATCCTGCGTTCCCGATTTC
PBX1_2	*PBX1*	TGGTCCGGCTTTGCTCTCGC
PBX1_3	*PBX1*	CCTGCGCCTCATCCAAACTC
PBX1_4	*PBX1*	CGGCCATCCCGACCCCAGCA
PBX1_5	*PBX1*	TGTGAAATCAAAGAAAAAAC
NTC_1	Non-targeting control	ACGGAGGCTAAGCGTCGCAA
NTC_2	Non-targeting control	CGCTTCCGCGGCCCGTTCAA
NTC_3	Non-targeting control	ATCGTTTCCGCTTAACGGCG
NTC_4	Non-targeting control	GTAGGCGCGCCGCTCTCTAC
NTC_5	Non-targeting control	CCATATCGGGGCGAGACATG