Table of Contents for Supplemental Material

Supplementary Figure S1

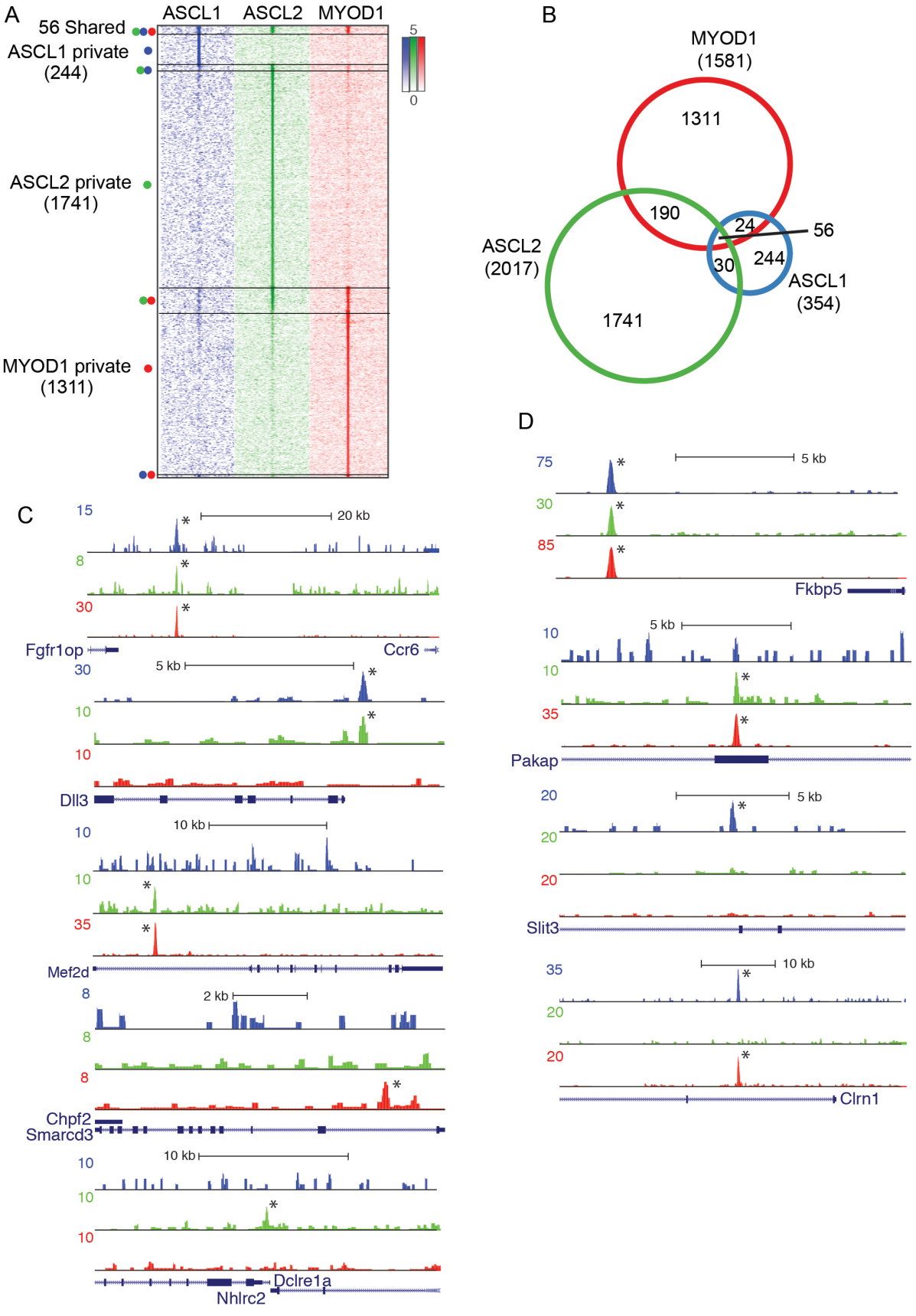

Figure S1. Validation of intrinsic binding specificity properties of bHLH factors in the ESC (supporting Figure 2). (A) Heatmap of ChIP-seq data sets from an independent experiment with ESCs induced to express ASCL1, ASCL2, or MYOD1. Each column shows the ChIP-seq signal for the factor indicated. Each row indicates a single interval +/-3 kb centered on the peak apex, with the ChIP-seq signal indicated by color density. Colored circles indicate which ChIP-seq data set(s) identified a significant peak. (B) Proportional area diagram of bHLH binding sites shown in the heatmap (A) with numbers representing distinct peaks identified in each subset. Peaks were considered overlapping if the peak apex was within 150 bp. (C,D) UCSC Genome browser tracks illustrating examples of bHLH ChIP-seq enrichment near gene targets with either shared or private peaks that confirm (C) those shown in Figure 2, and (D) additional examples to validate the conclusion that these bHLH factors bind distinct sites when presented with a similar genomic landscape. Asterisks indicate significant peaks. Note, ChIP enrichment in the second data set shown here was lower compared to the data shown in the main manuscript. To remove false positive peaks, in (A,B) we included a filter for the CAGSTG Ebox within 150 bp around the peak center.

Supplementary Figure S2

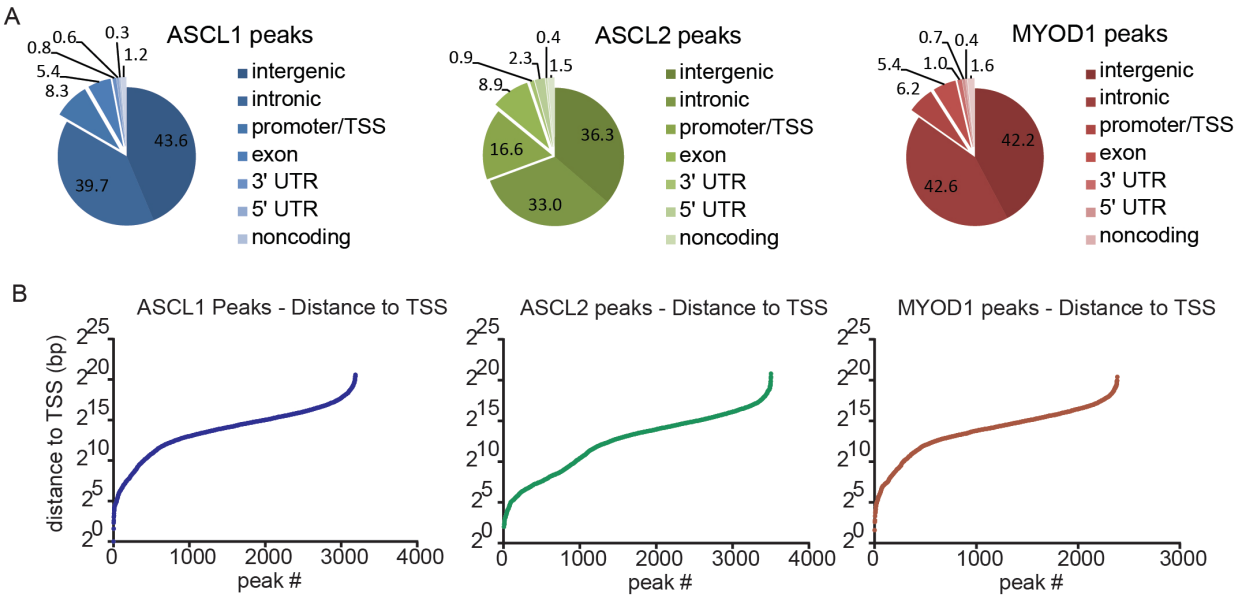

Figure S2: ASCL1, ASCL2, and MYOD1 binding sites are similarly distributed relative to gene features. (A) Genic feature annotation of ChIP-seq peak regions bound by bHLH factors in ESCs at 24 hours post-induction. Numbers represent percent of total peaks for each bHLH bound in each annotation category listed. Each peak is assigned to only one category shown. Annotation performed using HOMER. (B) Distribution of distances from bHLH binding sites to TSS. Plot represents absolute distance between peak center and nearest TSS for each peak identified within ChIP-seq data set.

Supplemental Figure S3

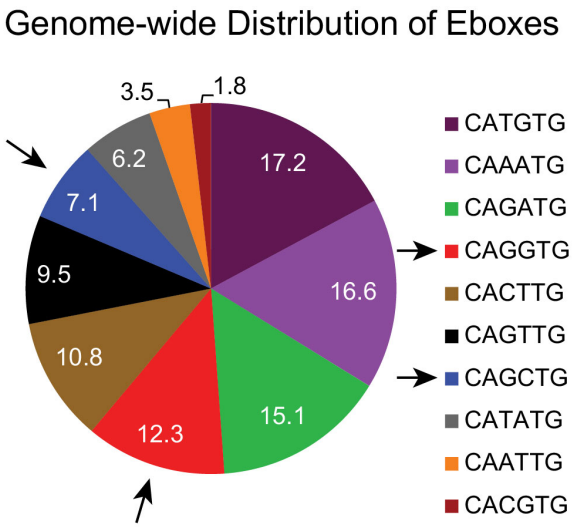

Figure S3. Genome-wide distribution of Eboxes demonstrating ASCL1, ASCL2, and MYOD1 Ebox preference is not based on genomic motif prevalence. Chart shows prevalence of each core dinucleotide Ebox sequence genome-wide in the mouse genome (mm10). Numbers reflect percentage of each Ebox dinucleotide as compared to total Ebox sequences. Arrows highlight enriched Ebox sequences in ChIP-seq for bHLH factors.

Supplemental Figure S4

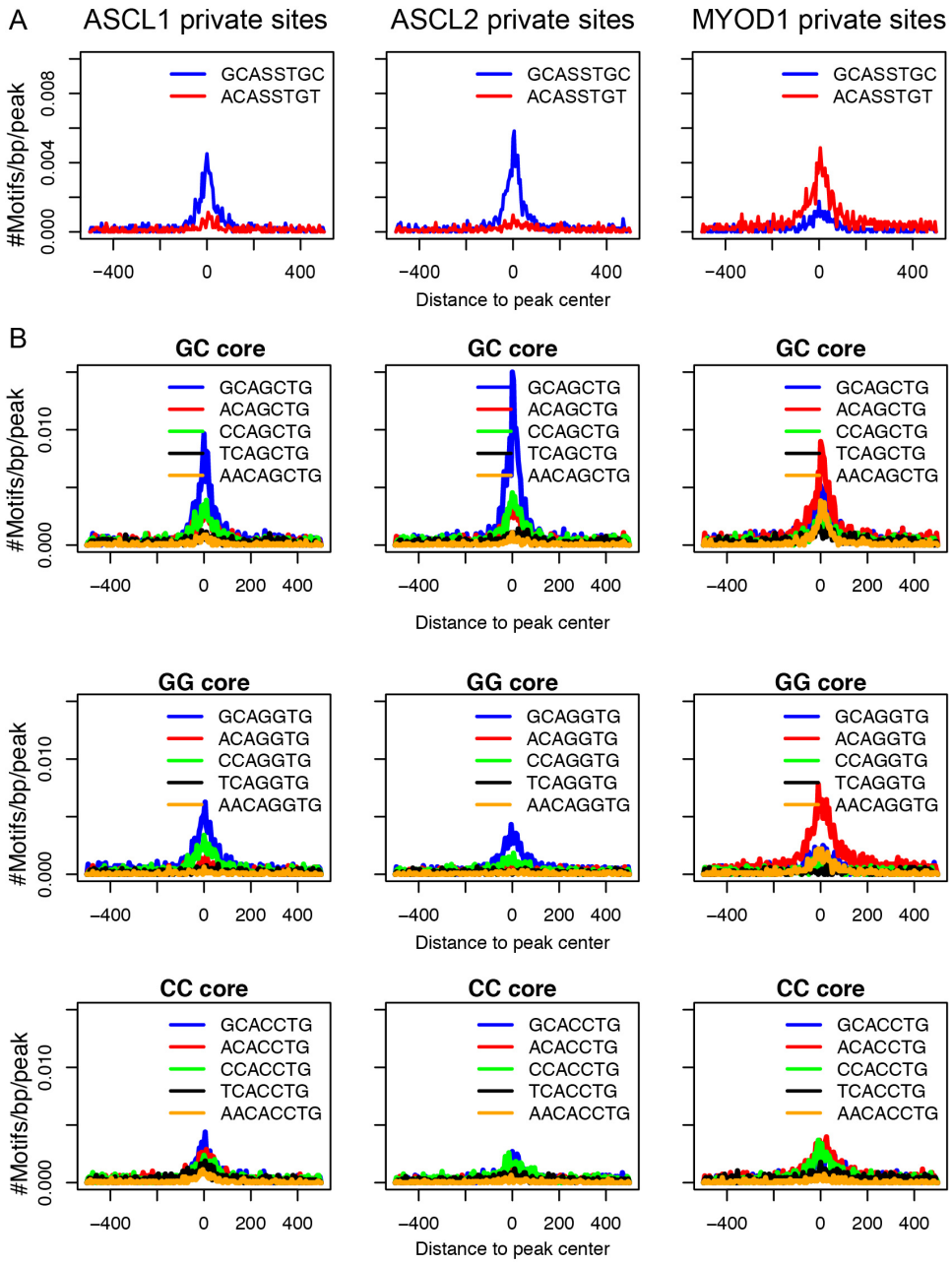

Figure S4: Extended Ebox flanking motifs are differentially enriched at ASCL1 and ASCL2 versus MYOD1 binding sites. Plots show distribution of indicated motifs on a 1 kb interval centered on ChIP-seq peak apex, as indicated. (A) Palindromic motifs. (B) Asymmentric motifs.

Supplemental Figure S5

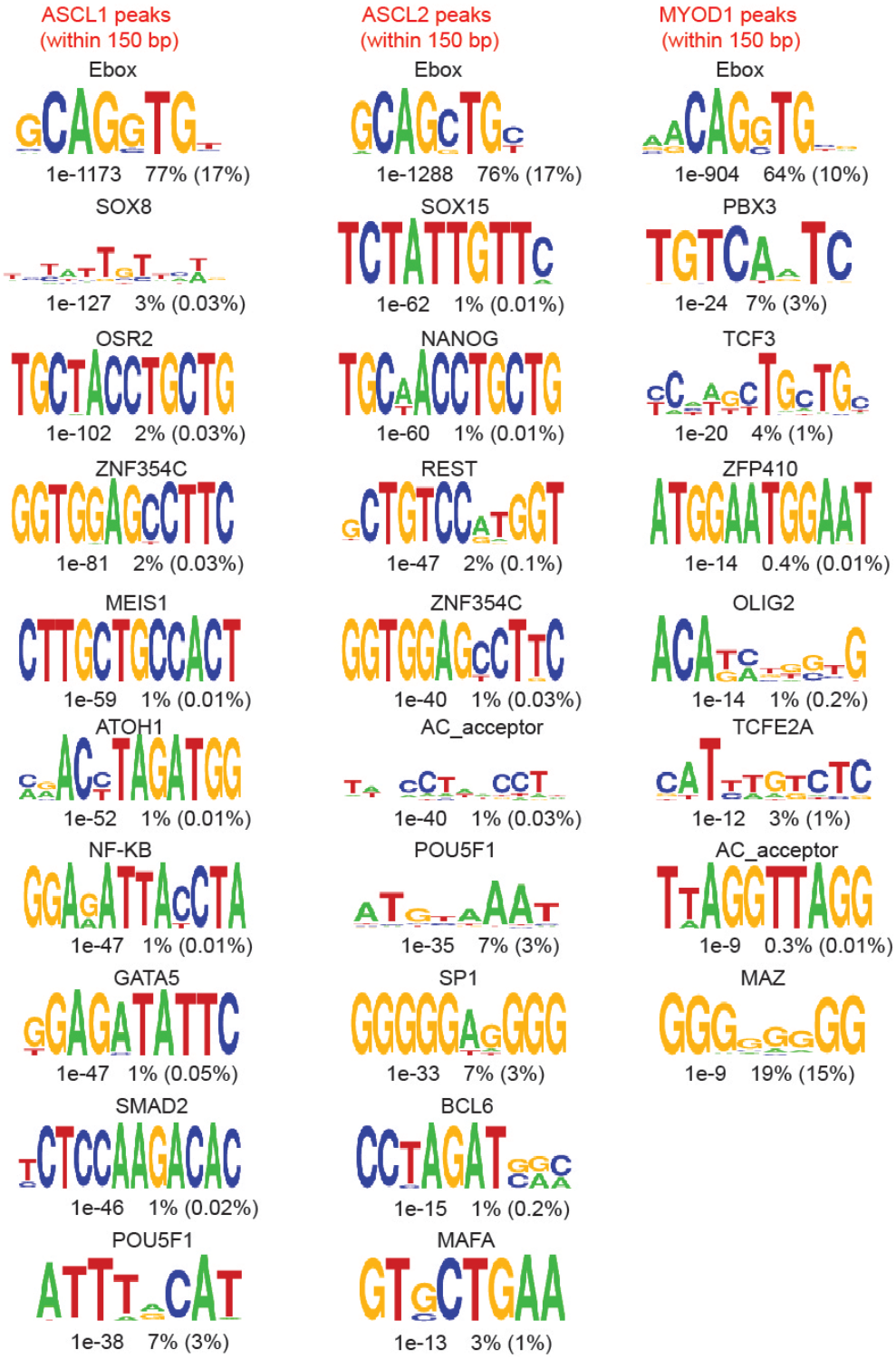

Figure S5. de novo motif analysis within 150 bp ChIP-seq peak regions for ASCL1, ASCL2, or MYOD1 induced in ESCs. HOMER was used for de novo motif identification from 150 bp sequences centered on the apex of the ChIP-seq peaks for each bHLH factor. The best, known TF match is shown with binomial enrichment p-value, percent of peaks containing the motif, and percent of sequence-normalized background regions containing the motif.

Supplemental Figure S6

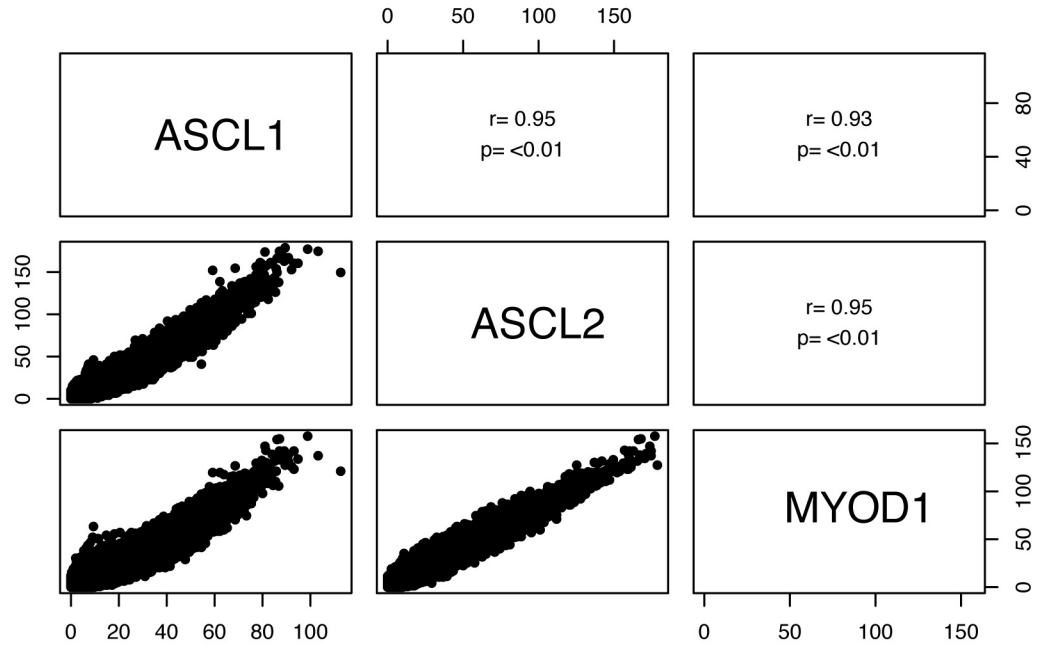

Figure S6. Scatter plots showing the similarity between ATAC-seq profiles across the different uninduced ESCs. The ATAC-seq signal profile was plotted in each accessible region identified in ESCs before the induction of ASCL1, ASCL2, or MYOD1. These values were normalized and Pearson correlation coefficient was computed to measure the similarity between ATAC-seq profiles. The panels below the diagonal show the actual signal comparison between each pair of factors. The panels above the diagonal show the Pearson correlation coefficient between each pair of factors.

Supplemental Figure S7

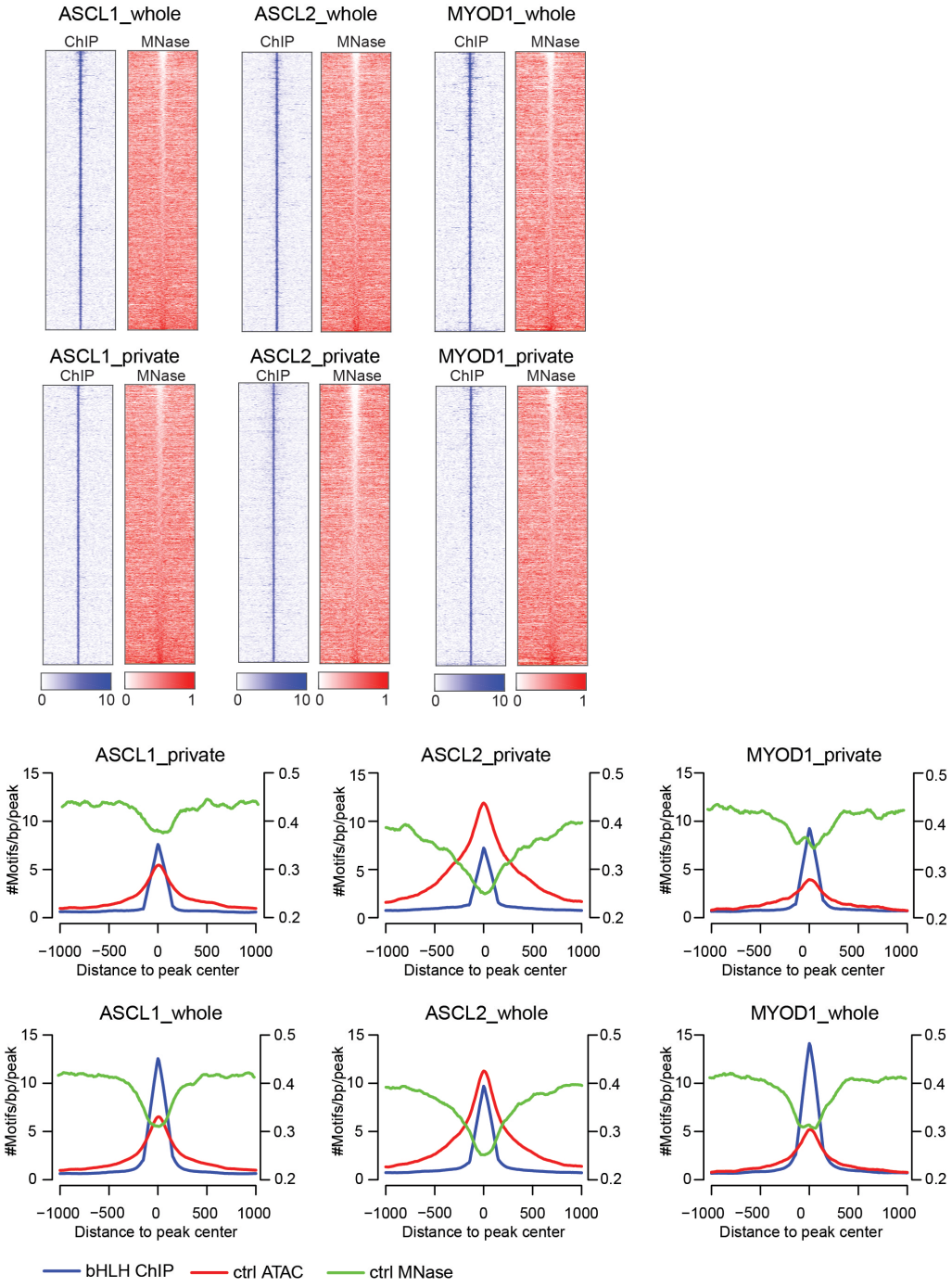

Figure S7: MNAse-seq demonstrates variable chromatin accessibility at bHLH binding sites identified by ChIP-seq in modified ESCs. (A) Heatmaps show ChIP-seq and MNase-seq signal on ~1 kb interval centered on each peak apex identified by ChIP-seq. MNase data from (GSM1400767) (Carone et al. 2014). Signal indicated for ChIP-seq and MNAse-seq as shown. (B) Histograms show mean ChIP-seq (blue), ATAC-seq (red), and MNase-seq (green) for total and private binding sites identified by ChIP-seq for each factor across a 2 kb interval centered on the peak apex. MNase-seq signal is plotted on right y-axis.

Supplemental Figure S8

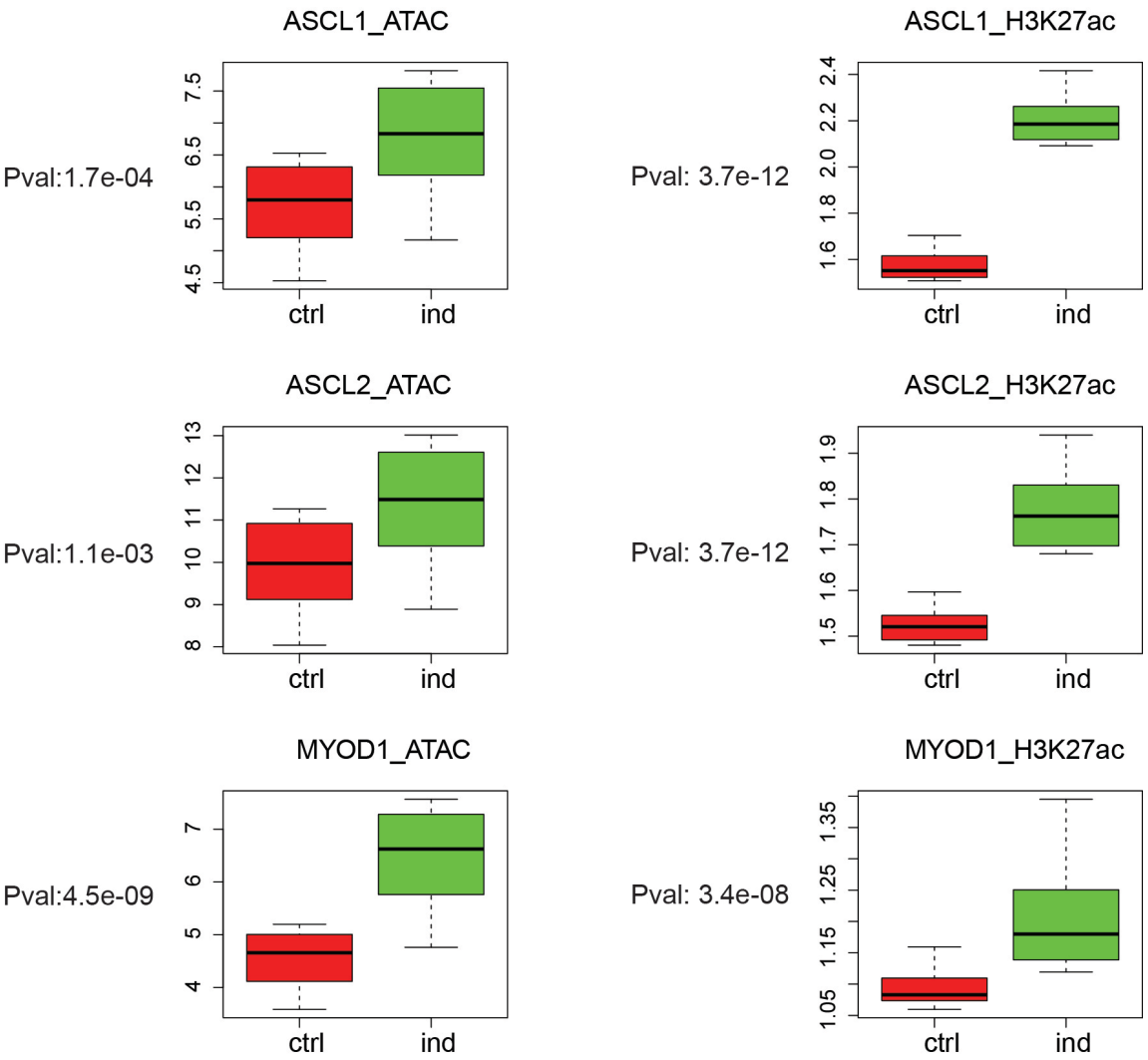

Figure S8: Boxplots showing the ATAC-seq and H3K27Ac signal change in bHLH binding sites before and after induction of bHLH factors. Cumulative ATAC-seq and H3K27AC signal was quantified before and after bHLH induction in 200 bp around the bHLH peak summits and visualized as boxplots. Y-axis shows the cumulative ATAC-seq or H3K27AC signal. P-values were computed using Wilcoxon signed-rank test.

Supplemental Figure S9

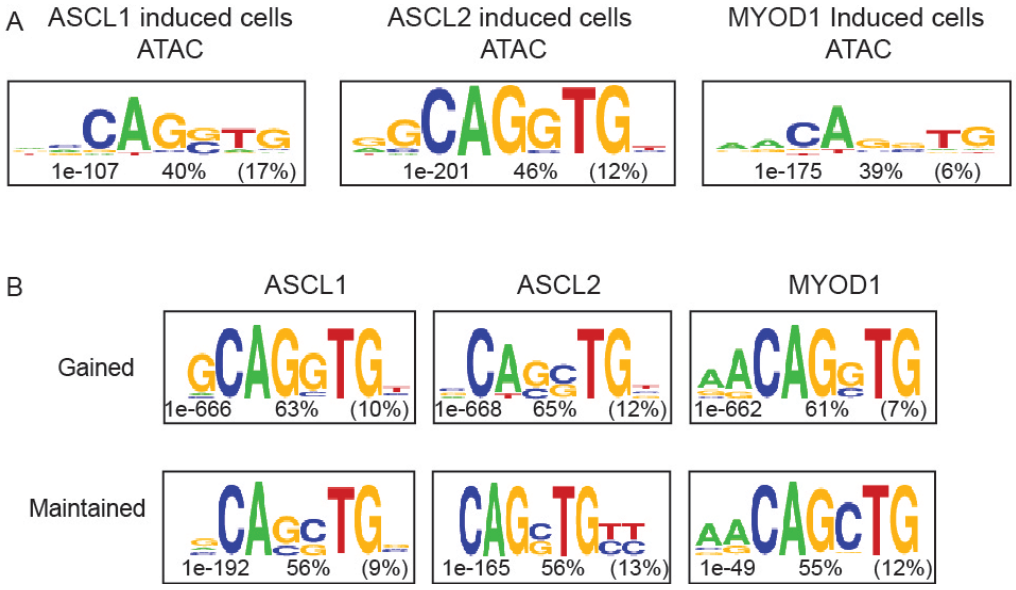

Figure S9: *de novo* motif analysis in ATAC-seq identified accessible regions. (A) Motifs identified in ESCs from ATAC-seq accessible regions identified in bHLH induced samples when compared to uninduced samples. (B). Motifs identified in ASCL1, ASCL2, and MYOD1 bound regions that were classified as ATAC-seq gained regions and maintained regions. For the analysis, ATAC-seq signal was quantified in each ASCL1, ASCL2, or MYOD1 bound region. The ATAC-seq gained/lost regions were defined if the ATAC-seq signal in respective bHLH induced ESCs was increased/decreased by at least 10% compared to uninduced ESCs. The peaks that do not fall in gained or lost peaks were classified as maintained regions. The binomial enrichment p-value, percent of peaks containing the motif, and percent of sequence-normalized background regions containing the motif are shown.

Supplemental Figure S10

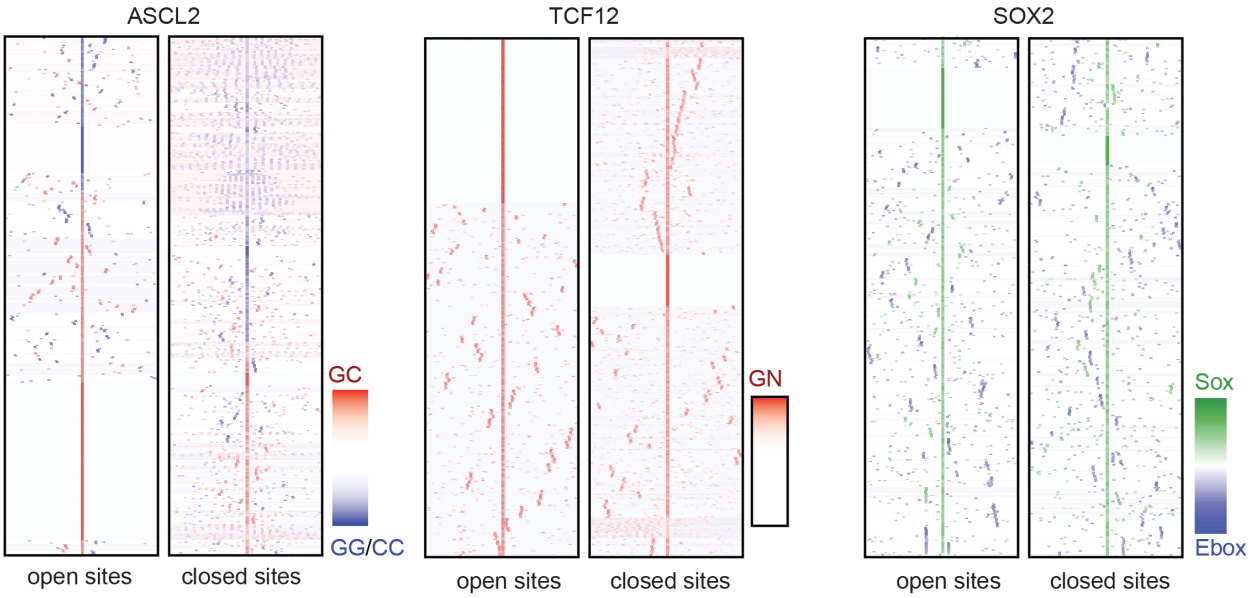

Figure S10. Reiterated Ebox patterning is not identified at TCF12 or SOX2 binding sites. Patterning of Ebox motifs identified in ASCL1, ASCL2, and MYOD1 bound sites is not identified at TCF12 (Class I bHLH) or SOX2 (SOX Family TF) in either open or closed chromatin (GSM1288311, GSM1288310, and GSM288347). Heatmap comparison of CAGSTG Ebox motifs identified in upper and lower fraction of binding sites identified by ChIP-seq for each factor. TCF12 and SOX2 sites represent most, and least accessible 5% of peaks, based on chromatin accessibility in uninduced ESCs, as assayed by ATAC-seq. ASCL2 bound sites shown for comparison, (plot parameters identical to Figure 6). Plots reflect cluster analysis of the distribution of each motif, with peak coordinates computationally centered on the CAGSTG Ebox motif closest to the peak apex, within +/-25bp of peak apex, and visualized across 300 bp centered on motif specified. Peak intervals lacking a CAGSTG motif within this interval are not shown in this comparison. TCF12 and SOX2 plotted based on de novo motifs identified for these factors (CAGNTG Ebox for TCF12, SOX motif for SOX2).

Supplemental Methods

Culturing ESCs and induction of bHLH factors

These studies were performed in three inducible murine ES cell (mESC) lines derived from 129S6/SvEvTac that express *Ascl1*, *Ascl2*, or *Myod1,* under the control of a tetracycline-repressive promoter system (Nishiyama et al. 2009). The engineered ESCs contain a tetracycline regulated transgene coding for each bHLH factor as a His6-FLAG fusion followed by an internal ribosome entry site (IRES) then Venus all in the *ROSA* locus (*R26R).* Frozen aliquots of these cells are provided by Coriell Institute for Medical Research.

ES culture media contained Dulbecco's Modified Eagle's Medium (Millipore, SLM-120-B), 20% v/v Fetal Bovine Serum (Gemini, 100-525), 0.1mM β-mercaptoethanol (Millipore, ES-007-E), 1.93mg/ml L-glutamine (Fisher Scientific, BP379-100), 1% v/v penicillin/streptomycin (Gibco, 15070-063), 100 μM non-essential amino acids (Millipore, TMS-001-C), 1% v/v nucleosides (Millipore, ES-008D), 1mM sodium pyruvate (Sigma-Aldrich, P5280-25G), 1000U/ml Leukemia Inhibitory Factor (Gemini, 400-495), 1 μg/ml puromycin (1μg/ml), and 0.2μg/ml doxycycline (Sigma-Aldrich, D9891). Mitomycin-C treated SNLP (puromycin resistant murine embryonic fibroblast feeder cells) were plated at 10^6 cells per 10cm plate, and cultures were allowed to grow for at least 24 hours prior to plating ESCs. Cells were maintained in a water-filled 37C incubator with 5% CO_2, and passaged at 48h on to new feeder cultures. After initial recovery from frozen stocks, ESCs were plated at a density of 10^7 cells to 10cm plates in 5-10ml of media, and maintained in culture at below 80% confluence to avoid spontaneous differentiation. ESC media was changed at least every 24 hours, and cells were passaged every 48 hours. Cell pluripotency was assessed by staining formaldehyde-fixed cells with alkaline phosphatase (Millipore, SCR004). All cell lines tested showed strong positive alkaline phosphatase staining, indicating that these cells are effectively maintained in a pluripotent state at passage numbers beyond those used for experiments in our study.

ESCs were grown on SNLP feeder cells for expansion prior to experiments. As murine SNLP cells are also present in these cultures, they represent a potential source of experimental bias. To reduce the influence of these cells in our genomic studies, ESCs were passaged at equivalent density to gelatinized plates without feeder cells for the last two passages prior to induction. As mitomycin-C treated cells are non-proliferative, this is sufficient to effectively remove these cells from culture prior to harvest for experiments. Prior to induction, cells were plated $1.5x10^7$ cells/10cm plate. Induction of their respective transgenes by removal of doxycycline was accomplished by serial washes and replacement of the media with no doxycycline. Three rounds of washes, using 37C Ca(-), Mg(-) phosphate buffered saline (PBS), with three hour delays between these washes, proved effective for inducing robust expression of the VENUS reporter. To minimize the potential effect of the additional media changes in induced cells, uninduced control cells were removed from the incubator, and the media replaced in

parallel. At the last wash, induced and uninduced cells were passaged to new plates without feeder cultures.

RNA preparation

RNA was purified from ESCs in parallel with chromatin preparation. Cells were observed for VENUS fluorescence to confirm induction of transgene prior to harvest. Prior to sample collection, laminar flow hood, instruments, benchtop surfaces, and centrifuge interiors were treated with RNAseAway (Fisher Scientific, 10328011) to remove potential contaminants. 10cm plates containing samples of induced and uninduced control ESCs (~$1.5x10^7$ cells) were removed from 37C incubator, and washed once in 15mL ice-cold PBS. The PBS was decanted away, and 1 mL of RNA lysis buffer (Zymo Research, R1054) was immediately added to each plate. Disposable nuclease-free cell lifters were used to detach cells from the plate surface. Lysates were transferred to 1.5mL microfuge tubes via pipette. Samples were stored at -80C as stabilized lysates, and were purified for RT-qPCR analysis and sequencing. RNA purification was performed using a small volume column elution as per the *Zymo Research* provided protocol, including 15 minute DNAse I treatment to remove residual trace DNA prior to column elution (Zymo Research, R1054). All samples were eluted into nuclease-free water, and quantified using a NanoDrop benchtop spectrophotometer (Thermo Scientific, ND1000 or ND2000).

Based on the result of spectrophotometric analysis, 5ug samples of purified RNA in nuclease-free water were prepared for sequencing, and stored at -80C pending analysis of sample quality. Transgene expression was evaluated by reverse-transcription quantitative polymerase chain reaction analysis (RT-qPCR). cDNA preparation was performed from 1µg purified RNA using Invitrogen SuperScript III (Invitrogen, 18080-044), using a Bio-Rad C1000 thermocycler. Expression of *Ascl1, Ascl2* and *Myod1* transcripts was evaluated using primers directed against endogenous and transgenic transcripts for these factors. All RT comparisons validated to be RNA-transcript specific by comparison of non-reverse-transcribed control reactions. Quality of RNA samples was assessed by bioanalysis, and all samples used for sequencing demonstrated RNA Integrity Number (RIN) ≥ 9. 5 µg of purified RNA was submitted to the UT Southwestern Microarray Core facility for sequencing library preparation, and single-end 50bp sequencing on an Illumina HiSeq 2500 line. Analysis of the resulting data sets is described in a later section.

Primers used to evaluate bHLH expression		
Gene Symbol	**Forward Sequence**	**Reverse Sequence**
Gene specific PCR primers for total transcript containing ORF		
Venus	CAACAGCCACAACGTCTATATCACCG	CTTTACTTGTACAGCTCGTCCATGCC
Ascl1	CACCATCTCCCCCAACTACTCCAAC	GAACCAGTTGGTAAAGTCCAGCAGC
Ascl2	ATGGAAGCACACCTTGACTGGTACG	TTTGCACCTTCACGGGCCTC
Myod1	GTGGCGACTCAGATGCATCCAG	GTCGTAGCCATTCTGCCGCC
Gene specific PCR primers for endogenous transcript		
Ascl1	TTAGCCCAGAGGAACAAGAGCTGC	TGCTTCCAAAGTCCATTCCCAGG
Ascl2	AGGAGCTGCTTGACTTTTCCAGTTG	TTGGGCTAGAAGCAGGTAGGTCCAC
Myod1	ATCCAGCCCCAAAGAAAGGACATAG	TGGCCACTCAAGGATCAGCTCTG
H2A specific PCR primer used in normalization		
H2afz	TTGCAGCTTGCTATACGTGGAGATG	TGTTGTCCTTTCTTCCCGATCAGC

Chromatin immunoprecipitation

Fixed whole ESCs were transferred to conical tubes, washed briefly in 5ml ice-cold modified RIPA buffer solution (50mM HEPES-KOH, 140mM NaCl, 1mM EDTA, 10% v/v glycerol, 0.5% v/v Nonidet P-40 substitute IGEPAL 630, 0.25% v/v Triton X-100). Cells were then centrifuged in a refrigerated centrifuge for 5 minutes at 400 x g, and supernatant was discarded. Pellets were resuspended in 5ml ice-cold saline solution (1mM Tris pH 8.0, 200mM sodium chloride, 1mM EDTA pH 8.0, 0.5mM EGTA pH 8.0), and centrifugation repeated. Supernatant was removed, and nuclear pellets resuspended in 275 µl ice-cold lysis buffer (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% w/v sodium deoxycholate, 0.5% w/v N-laurylsarcosine), transferred to siliconized low-adhesion microfuge tubes, and incubated on ice for 15 minutes.

Samples were sonicated using an ice-chilled Diagenode Bioruptor Standard Sonicator, for a total of 35 minutes, using a 15:15 second on/off cycle, in 7 bouts of 5 minutes each. Bath temperature was regulated by regular replacement of ice water. Samples were diluted in chromatin immunoprecipitation buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% Triton X-100, and 2mM EDTA), and centrifuged 30 minutes at max speed (~30,000 x g) to remove cellular debris. The clear supernatant was transferred to new siliconized microfuge tubes, and incubated overnight with the antibodies described below. Each reaction was performed using 5µg of mouse anti-MASH1 (BD

Pharmingen 556604) for ASCL1 ChIP or mouse anti-FLAG (Sigma-Aldrich F1804) antibody for ASCL2 and MYOD1 ChIP. Antibody/lysate were added to new tubes containing 25µg Protein G Dynabeads (Life Technologies 10003D) and incubated for 4-6 hours at 4C on a benchtop rotator to immunoprecipitate bound fragments.

Samples were washed on a benchtop rotator in a series of 4 minute washes in 1 ml volumes, in ice cold solutions. First wash in a low-salt buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA, 0.1% w/v SDS, 1% v/v Triton X-100), and a second wash in a high-salt buffer (20mM Tris-HCl pH 8.0, 400mM NaCl, 2mM EDTA, 0.1% w/v SDS, 1% v/v Triton X-100). 5 washes were performed in lithium chloride Wash buffer (250mM LiCl, 1% v/v NP-40 substitute, 1% w/v sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl pH 8.0). Beads were washed once in Tris-EDTA to remove trace detergents. Elution was performed at 70C in a heated robotic shaker, using two sequential elutions with heated buffer (10mM Tris-HCl pH 8.0, 1% w/v lithium dodecyl sulfate, and 1mM EDTA) to a total of 250 µl. Samples were treated with Proteinase K solution (11µl 5M NaCl, 5µl Proteinase K 10mg/ml) for 4 hours shaking at 55C, and incubated overnight at 65C in a heated robotic shaker to reverse crosslinks. ChIP and input samples were purified using Qiagen QIAquick miniature affinity purification columns, and stored in the provided elution buffer. Samples were evaluated by qPCR and quantified using the Qubit DNA high sensitivity kit, and duplicates combined prior to library preparation.

ChIP purification for H3K27ac was performed using the same protocol as for the bHLH factors. Importantly, unlike ChIP for bHLH factors, ChIP for H3K27ac utilized 0.1% sodium dodecyl sulfate (SDS) in the lysis buffer to facilitate fragmentation. ChIP purification was performed on fixed, flash frozen aliquots of whole cells prepared as described for ESC harvests. ChIPs for histone markers were performed from aliquots of $1.0x10^7$ cells. Each reaction was performed using 5µg of anti-H3K27ac antibody (Abcam, ab4729). 25µg Protein G Dynabeads (Life Technologies, 10003D) were used to immunoprecipitate bound fragments. Elution was performed at 70C in a heated robotic shaker using a solution of lithium dodecyl sulfate. Samples were treated with Proteinase K solution, and incubated overnight at 65C in a heated robotic shaker to reverse crosslinks. ChIP and input samples were then purified using Qiagen QIAquick miniature affinity purification columns. Samples were evaluated by qPCR, and duplicates combined prior to library preparation.

Preparation of ChIP sequencing libraries

Illumina DNA sequencing library preparation was performed as per NEBNext ChIP-Seq Library Preparation protocol using Illumina-compatible multiplexing primers. Libraries were generated using 2-4ng of ChIP purified chromatin as well as purified 10ng input controls from the same samples. Library amplification was performed using multiplexing primer pairs (New England Biolabs, E7735S). Size selection was performed using Ampure XP bead purification. The resulting libraries were sequenced by the UT Southwestern Microarray facility, using single-end 50bp sequencing, on an Illumina HiSeq 2500.

The resulting reads were demultiplexed and aligned to the mouse *mm10* genome (GRCm38), using Bowtie2 (Langmead et al. 2009). Peak calling, intersection, annotation, and motif analysis were performed using HOMER v4.7 (Heinz et al. 2010). Peak to gene calling was performed using HOMER v4.7 (Heinz et al. 2010) and GREAT (McLean et al. 2010).

Transposase-Accessible Chromatin (ATAC-seq)

ATAC-seq from ESCs was performed as per the previously published protocol outlined in Buenrostro et al., 2013 and Buenrostro et al., 2015. Cells were harvested by dissociation with warm trypsin-EDTA, quenched by the addition of ice-cold serum-containing media, and diluted in 4C PBS. 50,000 cells were used in the preparation of these libraries. Traditional polymerase chain reaction amplification (PCR) was utilized for amplification of transposed elements in a Bio-Rad C1000 thermocycler, using the programs specified in Buenrostro et al., 2015. Multiplexed single-end 50bp sequencing was performed using an Illumina HiSeq 2500 using Nextera-compatible amplification primers. Due to differences in multiplexing primers, backwards-compatible Nextera sequencing primers were used to sequence these samples. Outputs of this sequencing were demultiplexed using sample-specific Nextera primer sequences, and fastQC (Andrews et al. 2010) was used to filter and score the resulting sequencing runs. This sequencing provided high read depth and complexity, as expected for these samples. The sequencing results of each sample were aligned to the mouse mm10 genome (Kent et al. 2002; Kent et al. 2010), and processed for downstream analysis using Bowtie2 (Langmead et al. 2009), and HOMER (Heinz et al. 2010).

Bioinformatics and computational analysis

The study presented here makes extensive use of a number of previously developed open source computational algorithms and software packages, which are introduced here in brief. The value of these resources in completing this study cannot be overstated. In particular, *Hypergeometric Optimization of Motif EnRichment* (Li et al. 2009), a comprehensive genomic analytics package (Heinz et al. 2010) has been extensively utilized.

Data handling and software used for analysis

Genome-wide DNA sequencing data (ChIP-seq and ATAC-seq) from multiple lanes were demultiplexed using sample-specific Illumina primer sequences. The resulting data sets were aligned to the mm10 genome using Bowtie2 v2.2.6 (Langmead et al. 2009). Reads with a Bowtie2 quality score less than 10 were removed using SAMtools v1.3 (Li et al. 2009) with parameters (*-bh -F 0x04 -q 10*). Duplicate reads were removed using picardtools v1.119, and the remaining reads were normalized to 10M reads using HOMER v4.7 (Heinz et al. 2010). All UCSC Genome Browser plots shown (Kent et al. 2002; Kent et al. 2010; Raney et al. 2014; Rosenbloom et al. 2015) reflect these normalized tag counts.

Sequencing of RNA samples was aligned to the mouse mm10 genome using TopHat 2.1.0 (Langmead et al. 2009; Trapnell et al. 2009; Trapnell and Salzberg 2009). Default settings were used, with the exception of *–G*, specifying assembly to the mm10 genome, *--library-type fr -first strand*, and *–no-novel-juncs*, which disregards noncanonical splice junctions when defining alignments. *edgeR* (Robinson et al. 2010; Nikolayeva and Robinson 2014) was used to incorporate RNA-seq data from three biological replicates for each factor tested, and identify genes which were differentially expressed between samples, using the default parameters.

Use of previously published genomic data sets

In addition to the experiments described here, the results of these analyses make use of a number of previously available data sets. These sets were downloaded from the public *Gene Expression Omnibus* (Edgar et al. 2002) from the accession numbers provided in the works cited, and processed for analysis using the same approach as our ChIP-seq data sets (as described below) for unbiased comparison. In each instance, the original source of these data is indicated in the text accompanying its use.

Identification of bHLH binding sites from sequencing data (peak calling)

The sequencing data sets generated from ChIP-seq from the ESCs were used to identify putative binding sites (ChIP-enriched peak regions) genome-wide based on the distribution of the aligned reads. ChIP-seq data sets for transcription factors were normalized to 10M reads. Peaks for each sample were called based on respective input control samples created during immunoprecipitation. This approach addresses potential bias from variation in immunoprecipitated fragment length and sequence bias in sequencing library preparation. Peak calling was performed using a sequencing-depth independent approach to correct for variation between data sets used; the *findPeaks* library of HOMER 4.7 (Heinz et al. 2010) was used to call significantly enriched peak regions from each data set. Significance is evaluated as a FDR ≤0.0010 (0.1%). Parameters used specify for selection of focal peak regions (*-factor*), which uses the autocorrelated predicted fragment length (derived from aligned read distribution) to identify changes, modeled on a Poisson distribution. These peaks are subjected to local filtering of fourfold compared to the surrounding interval (*-L 4*), and discards regions not meeting significance by Poisson p-value threshold of ≤1.00e-04. Sequencing data sets from ASCL2 and MYOD1 were subjected to additional filtering based on the uninduced control sample from the same experiments. This additional filtering was added to address the presence of a small number of non-specific peak regions in an unbiased manner. Peak calling was performed comparing ASCL2 and MYOD1 24h induced data sets to their respective input controls, as described for ASCL1. Peak calling was then performed using ASCL2 and MYOD1 induced samples versus an uninduced control sample at lower stringency (*-LP 1E-2*), and the merged peak lists from above two comparison were annotated for the presence of a canonical REST/NRSF motif (*JASPAR MA0138.2*)

(Mathelier et al. 2016). This was observed in ChIP-seq from both induced, and uninduced ESCs of *tTA-Ascl2-FLAG*, and *tTA-Myod1-FLAG* cell lines, and thus was nonspecific. Regions which showed the presence of this motif within 100bp of the empirically determined peak center were removed from the ASCL2 and MYOD1 peak lists.

Identification of potential regulatory targets by peak-to-gene association

To identify potential regulatory targets of these factors, peak-to-gene calling was performed, which observes the location of transcriptional start sites (TSS) present in genomic intervals surrounding the putative binding sites identified from ChIP-seq data. HOMER 4.7 (Heinz et al. 2010) and *Gene Region Enrichment Association Tool* (GREAT) v3.0 (McLean et al. 2010) were used to perform this analysis in two distinct ways, addressing two distinct aims. Peak-to-gene calling HOMER utilizes a straightforward approach to gene calling, comparing the distance from the peak to the RefSeq-curated catalog of transcription start sites (TSS) of nearby genes, and selecting the closest TSS as the gene identified. HOMER also includes specific non-genic features present in the RefSeq catalog, such as miRNAs, ncRNAs, and pseudogenes (Pruitt et al. 2014), and identifies a single feature for each genomic position. This allows for objective comparison of distances to nearby start sites, and identifies non-genic features which represent potential targets. However, as class II bHLH transcription factors are known to function primarily at distal enhancer regions, this approach is expected to underestimate the number and identity of potential regulatory targets.

To address this, GREAT v3.0 (McLean et al. 2010) was used to perform peak-to-gene calling when surveying potential gene targets. GREAT observes the location of multiple TSS for each putative regulatory region, and reports the genes identified, based on the canonical isoform for each gene, and reports gene ontology associations (GO) from the list of genes identified. For all analysis performed in these studies, the association rules are 5kb 5' of the TSS, 1kb3' of the TSS, and an extension to the next gene of up to 1mb. Thus, this approach can associate a single peak with multiple genes, and a single gene with multiple peaks. In practice, this increases the total number of genes associated with each putative binding site, allowing for association with multiple genes. While neither approach identifies every potential regulatory target of a given binding site, GREAT generally identifies more genes associated with binding sites identified in ChIP, generating larger lists of genes for further comparisons. HOMER identifies a single genomic feature, and reports non-genic features as well as RefSeq genes. The algorithm used in each analysis is noted where relevant.

De novo motif discovery

This study makes considerable use of *de novo* motif discovery to observe the revealed preference for specific DNA sequence of bHLH factors; HOMER v4.7 was used for all motif discovery presented here, but analysis with MEME (Bailey et al. 2009) and GEM (Guo et al. 2012) gave similar

results. To compare bHLH factor binding sites, a narrow observational window of 50bp centered on the peak apex identified by HOMER was used. This narrow window was selected specifically to avoid the influence of adjacent regions when identifying the primary motif bound by these factors. To screen for potential DNA-binding co-factor motifs, a broader window of 150bp centered on the peak apex was used to identify enriched motifs adjacent to the primary binding site. The interval used for each analysis is noted in the text. Parameters *–S 10 –bits* were used unless otherwise noted. In *de novo* motif analysis, HOMER searches for motifs with 8, 10, or 12 bps on both strands such that flanking site preferences are not degraded. The statistical comparisons shown reflect the significance identified by HOMER for the specific *de novo* motifs identified from ChIP-seq. The *de novo* motifs shown for each analysis represent the most significantly enriched motif identified from the specified set of intervals. In instances where non-specific or low-information motifs are identified, the next best match is shown. Statistics reflect the motif shown in each instance. For genome-wide motif discovery, the binomial distribution is used for statistical comparisons, whereas for promoter motifs, the hypergeometric distribution is used to observe significance as compared to promoter-specific background regions.

Scatterplot, histogram, and heatmap generation from genome-wide sequencing data

To visualize genomic density of sequencing reads, HOMER's *annotatePeaks.pl* library was utilized to generate incidence matrices from the normalized sequencing reads generated for each data set (Heinz et al. 2010). The resulting matrices were sorted using either the Linux command line or *Microsoft Excel*, based on criteria described in the text accompanying each analysis. Heatmap plots were created using *Java TreeView 1.6* (Saldanha 2004), and *MatLab v. R2016b*®. Histogram representations of sequencing reads and motif distribution at binding sites were generated in HOMER (Heinz et al. 2010), and were plotted using *GraphPad Prism*® *7.0*. Scatterplot comparisons were created using HOMER (Heinz et al. 2010), and plotted in *RStudio* (Team 2015) using *ggPlot2* (Wickham 2009; Wickham 2016).

Identification of chromatin signatures using ChromHMM

To test whether bHLH binding sites identified by ChIP-seq in ESCs were associated with an identifiable signature, ChromHMM (Ernst and Kellis 2012) was used to apply this technique to our study. To build a model of potential chromatin states, we utilized ATAC-seq and H3K27ac ChIP-seq data sets from the uninduced ESCs, and combined them with previously published ChIP-seq data sets for 26 histone marks (Table S4). The information about the public data sets was obtained from cistrome project (Song et al. 2011). The raw datasets were downloaded from NCBI GEO database and processed same as our samples. Using these datasets, the 18-state model was selected as it appeared to allow sufficient complexity to observe apparent differences in histone marker distribution, and additional states did not demonstrate additional informative complexity. This model describes the chromatin states identified

from these histone data sets genome-wide, and allows for observation of these states at genomic intervals of interest, such as bHLH binding sites.

Enrichment of ChromHMM states in ChIP-seq peak regions

To identify enrichment of each state at ChIP-seq peaks, we used ASCL1, ASCL2, MYOD1, NEUROD1, GATA4, and SOX2 binding regions in ESCs, using genomic segmentation defined by ChromHMM identified state segments in ESCs as input for *NeighborhoodEnrichment* module in ChromHMM. Enrichment of chromatin states in 2000bp around peak apex was calculated and plotted.

Identification of open and closed peaks

To identify the differences between bHLH binding in DNA regions that are open and closed in uninduced samples, we used ATAC-seq data from *Ascl1* uninduced cells and calculated the ATAC-seq signal in the bHLH binding regions. The cumulative ATAC-seq signal in 100bp around the bHLH peak center was sorted and selected the peaks with ATAC-seq signal greater than or equal to third quartile values were considered as open peaks and ATAC-seq signal less than or equal to first quartile values were considered as closed peaks.

Determination of Ebox spacing in open versus closed bHLH peaks

To determine the spacing patterns in open versus closed bHLH peaks, we excluded open or closed bHLH peaks that do not contain CAGSTG Ebox in 50bp around the peak center and annotated these peaks with *de novo* motifs identified from corresponding open and closed peaks set. We identified motif instances of all these motifs in 300bp around the peak center.

In-house informatics for motif spacing used to generate data shown in Figure 6

Homer command to center peaks on specified motif: annotatePeaks <input peaks> <genome> -size < cutoff size in bp to have motif presence > -center <motif file> > <centered peak list

Homer command to identify ebox motif instances location in a peak list: annotatePeaks.pl <centered peak list> <genome> -size 300 -m <Motif files separated by space character> -strand + -norevopp –noann

Perl script to identify motif instances and their spacing information within 300bp window around specific motif centered peaks:

```
#!/usr/bin/perl
use strict;
my @listing_genes=();
#specify the directory containing Homer annotatepeaks output files with .txt suffix
opendir (DIR,"ebox_center") || die "Can't :$!";
@listing_genes =readdir DIR;
close DIR;
```

```
for(my $i=1;$i<=$#listing_genes;$i++)
{
	if($listing_genes[$i]=~/^(.*).txt/)
	{
		#specify output file name suffix
		my $op=$1."_ebox_center.txt";
		open (FILE, "ebox_center/$listing_genes[$i]") || die "can't: $!";
		open (OUTPUT,">ebox_center/$op") || die "can't: $!";
		while(<FILE>)
		{
			my $line = $_;
			chomp($line);
			my %gc=();
			my %gg=();
			my %cc=();
			my $j=-150;
			my $k=0;
			if($line=~/^PeakID.*/)
			{
				next;
			}
			else
			{
			#specify the column index values to get GC, GG, & CC core ebox instances
				my @array=split(/\t/,"$line");
				my @gc1=split(/\(/,"$array[21]");
				my @gg1=split(/\(/,"$array[22]");
				my @cc1=split(/\(/,"$array[23]");
				foreach my $i(0 .. $#gc1-1)
				{
					my @gc2=split(/\,/,"$gc1[$i]");
					if(!exists $gc{$gc2[-1]})
					{
						$gc{$gc2[-1]}=1;
					}
				}
```

```
		foreach my $i(0 .. $#gg1-1)
		{
			my @gg2=split(/\,/,"$gg1[$i]");
			if(!exists $gg{$gg2[-1]})
			{
				$gg{$gg2[-1]}=1;
			}
		}
		foreach my $i(0 .. $#cc1-1)
		{
			my @cc2=split(/\,/,"$cc1[$i]");
			if(!exists $cc{$cc2[-1]})
			{
				$cc{$cc2[-1]}=1;
			}
		}
	}#end of else
	while ($j<=149)
	{
		if(exists $gc{$j})
		{
			#Denoting GC motif instances with 10
			print OUTPUT "10\t10\t10\t10\t10\t10\t";
			$j=$j+6;
			$k++;
		}
		elsif(exists $gg{$j})
		{
			#Denoting GG motif instances with -10
			print OUTPUT "-10\t-10\t-10\t-10\t-10\t-10\t";
			$j=$j+6;
			$k++;
		}
		elsif(exists $cc{$j})
		{
			#Denoting CC motif instances with -10
```

```
                                print OUTPUT "-10\t-10\t-10\t-10\t-10\t-10\t";
                                $j=$j+6;
                                $k++;
                        }
                        else
                        {
                                print OUTPUT "0\t";
                                $j++;
                        }
                }
                #'k' variable is used to quantify number of ebox motif instances per peak
                if($j==150)
                {
                        print OUTPUT "0\t$k\n";
                }
                if($j>150)
                {
                        print OUTPUT "$k\n";
                }
        }
    }
}
```

Compendium of Accession Numbers

Data Set Label	Accession #	Original Publication
H4K16ac	GSM1156617	(Taylor et al. 2013)
H4K20me3	GSM656527	Rahl P, Bilodeau S, Young RA, unpublished
H3K27me1	GSM985054	(Ferrari et al. 2014)
H3K9-K14ac	GSM1294872	(Clouaire et al. 2014)
H3K14ac	GSM775314	(Karmodiya et al. 2012)
H2A-X	GSM1037482	(Wu et al. 2014)
H3K4me2	GSM686995	(Lienert et al. 2011)
H3.1	GSM487542	(Goldberg et al. 2010)
H2A-Zac	GSM958502	(Ku et al. 2012)
H2B	GSM1692827	(Ishii et al. 2015)
H3K64ac	GSM1668879	(Cayrou et al. 2015)
H3K9ac	GSM1000123	(Yue et al. 2014)
H3K56ac	GSM1148636	(Tan et al. 2013)
H3K36me2	GSM1019770	(Ballare et al. 2012)
H3K27me2	GSM1234538	(Ferrari et al. 2014)
H3K4me1	GSM1359829	(Org et al. 2015)
H4R3me2	GSM1045500	(Girardot et al. 2014)
H2A-Ub1	GSM850471	(Brookes et al. 2012)
H3K9me1	GSM1543627	(Liu et al. 2015)
H3K9me2	GSM1314605	(Liu et al. 2015)
H2A-Z	GSM984544	(Subramanian et al. 2013)
H3K4me3	GSM1526288	(Ji et al. 2015)
H3K9me3	GSM1526292	(Ji et al. 2015)

H3K27me3	GSM1526291	(Ji et al. 2015)
H3K36me3	GSM1526290	(Ji et al. 2015)
H3K79me2	GSM1526289	(Ji et al. 2015)
SOX2	GSM288347	(Chen et al. 2008)
NEUROD1	GSM1586779	(Pataskar et al. 2016)
NEUROD1	GSM1586781	(Pataskar et al. 2016)
ASCL1 (Emb NT)	GSM1347006	(Borromeo et al. 2014)
ASCL1 (Emb NT)	GSM1347007	(Borromeo et al. 2014)
ASCL2 (T-cells)	GSM1276938	(Liu et al. 2014)
ASCL2 (T-cells)	GSM1276937	(Liu et al. 2014)
MYOD1 (myotubes)	SRR035036, SRR035037, SRR035038	(Cao et al. 2010)
MYOD1 (myotubes)	SRR035056	(Cao et al. 2010)
SOX17	GSM1059856	(Aksoy et al. 2013)
SOX17	GSM1059866	(Aksoy et al. 2013)
GATA4	GSM1015512	(Oda et al. 2013)
GATA4	GSM1015514	(Oda et al. 2013)
TCF12	GSM1288311	(Yoon et al. 2015)
TCF12	GSM1288310	(Yoon et al. 2015)
MNase	GSM1400767	(Carone et al. 2014)

References

Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S et al. 2013. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* **32**: 938-953.

Andrews JL, Zhang X, McCarthy JJ, McDearmon EL, Hornberger TA, Russell B, Campbell KS, Arbogast S, Reid MB, Walker JR et al. 2010. CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. *Proc Natl Acad Sci U S A* **107**: 19090-19095.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.

Ballare C, Lange M, Lapinaite A, Martin GM, Morey L, Pascual G, Liefke R, Simon B, Shi Y, Gozani O et al. 2012. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**: 1257-1265.

Borromeo MD, Meredith DM, Castro DS, Chang JC, Tung KC, Guillemot F, Johnson JE. 2014. A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. *Development* **141**: 2803-2812.

Brookes E, de Santiago I, Hebenstreit D, Morris KJ, Carroll T, Xie SQ, Stock JK, Heidemann M, Eick D, Nozaki N et al. 2012. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* **10**: 157-170.

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL et al. 2010. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18**: 662-674.

Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, Fazzio TG, Rando OJ. 2014. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* **30**: 11-22.

Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Mechali M. 2015. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res* **25**: 1873-1885.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.

Clouaire T, Webb S, Bird A. 2014. Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biol* **15**: 451.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207-210.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215-216.

Ferrari KJ, Scelfo A, Jammula S, Cuomo A, Barozzi I, Stutzer A, Fischle W, Bonaldi T, Pasini D. 2014. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* **53**: 49-62.

Girardot M, Hirasawa R, Kacem S, Fritsch L, Pontis J, Kota SK, Filipponi D, Fabbrizio E, Sardet C, Lohmann F et al. 2014. PRMT5-mediated histone H4 arginine-3 symmetrical dimethylation marks chromatin at G + C-rich regions of the mouse genome. *Nucleic Acids Res* **42**: 235-248.

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X et al. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**: 678-691.

Guo Y, Mahony S, Gifford DK. 2012. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput Biol* **8**: e1002638.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Ishii H, Kadonaga JT, Ren B. 2015. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc Natl Acad Sci U S A* **112**: E3457-3465.

Ji X, Dadon DB, Abraham BJ, Lee TI, Jaenisch R, Bradner JE, Young RA. 2015. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc Natl Acad Sci U S A* **112**: 3841-3846.

Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**: 424.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**: 2204-2207.

Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H, Carr SA, Bernstein BE. 2012. H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol* **13**: R85.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schubeler D. 2011. Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet* **7**: e1002090.

Liu N, Zhang Z, Wu H, Jiang Y, Meng L, Xiong J, Zhao Z, Zhou X, Li J, Li H et al. 2015. Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. *Genes & development* **29**: 379-393.

Liu X, Chen X, Zhong B, Wang A, Wang X, Chu F, Nurieva RI, Yan X, Chen P, van der Flier LG et al. 2014. Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nature* **507**: 513-518.

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R et al. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**: D110-115.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495-501.

Nikolayeva O, Robinson MD. 2014. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol* **1150**: 45-79.

Nishiyama A, Xin L, Sharov AA, Thomas M, Mowrer G, Meyers E, Piao Y, Mehta S, Yee S, Nakatake Y et al. 2009. Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. *Cell Stem Cell* **5**: 420-433.

Oda M, Kumaki Y, Shigeta M, Jakt LM, Matsuoka C, Yamagiwa A, Niwa H, Okano M. 2013. DNA methylation restricts lineage-specific functions of transcription factor Gata4 during embryonic stem cell differentiation. *PLoS Genet* **9**: e1003574.

Org T, Duan D, Ferrari R, Montel-Hagen A, Van Handel B, Kerenyi MA, Sasidharan R, Rubbi L, Fujiwara Y, Pellegrini M et al. 2015. Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. *EMBO J* **34**: 759-777.

Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T, Tiwari VK. 2016. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J* **35**: 24-45.

Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM et al. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**: D756-763.

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D et al. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**: 1003-1005.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M et al. 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43**: D670-681.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D et al. 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**: 1757-1767.

Subramanian V, Mazumder A, Surface LE, Butty VL, Fields PA, Alwan A, Torrey L, Thai KK, Levine SS, Bathe M et al. 2013. H2A.Z acidic patch couples chromatin dynamics to regulation of gene expression programs during ESC differentiation. *PLoS Genet* **9**: e1003725.

Tan Y, Xue Y, Song C, Grunstein M. 2013. Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. *Proc Natl Acad Sci U S A* **110**: 11493-11498.

Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA. 2013. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res* **23**: 2053-2065.

Team R. 2015. RStudio: Integrated Development for R. RStudio, Inc.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Trapnell C, Salzberg SL. 2009. How to map billions of short reads onto genomes. *Nat Biotechnol* **27**: 455-457.

Wickham H. 2009. *Ggplot2 : elegant graphics for data analysis*. Springer, New York.

Wickham H. 2016. *Ggplot2*. Springer Science+Business Media, LLC, New York, NY.

Wu T, Liu Y, Wen D, Tseng Z, Tahmasian M, Zhong M, Rafii S, Stadtfeld M, Hochedlinger K, Xiao A. 2014. Histone variant H2A.X deposition pattern serves as a functional epigenetic mark for distinguishing the developmental potentials of iPSCs. *Cell Stem Cell* **15**: 281-294.

Yoon SJ, Foley JW, Baker JC. 2015. HEB associates with PRC2 and SMAD2/3 to regulate developmental fates. *Nat Commun* **6**: 6546.

Yue F Cheng Y Breschi A Vierstra J Wu W Ryba T Sandstrom R Ma Z Davis C Pope BD et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355-364.