SUPPLEMENTARY INFORMATION

Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes

Warren A. Whyte, David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young

CONTENTS

Supplementary Figures, Tables and Data

Supplementary Figures, Tables and Data

Figure S1: The *miR-290-295* enhancer and associated enhancer features, Related to Figure 1.

A) The *miR-290-295* enhancer and associated enhancer features**.** ChIP-Seq binding profiles (reads per million per base pair) for ESC transcription factors (Oct4, Sox2, Nanog (OSN)), the Mediator co-activator (Med1), histone modifications (H3K27ac, H3K4me1), and DNaseI hypersensitivity at the *miR-290-295* locus in ESCs. Gene models are depicted below the binding profiles. Enhancer and scale bar are depicted above the binding profiles.

B) Distribution of enhancer features across ESC enhancers, Related to Figure 1E. Normalized ChIP-Seq signal (total reads) for Med1, H3K27ac, H3K4me1, DNaseI hypersensitivity, and OSN across the 8,794 ESC enhancers. For each enhancer feature, the plot was normalized by dividing the ChIP-Seq signal at each ESC enhancer by the maximum ChIP-Seq signal. The enhancers on the x-axis are ranked according to ChIP-Seq signal for Mediator. The x- and y-axes have been adjusted so that the differences among the different enhancer features can be visualized.

Figure S2: Super-enhancers are not associated with housekeeping genes, Related to Figure 2.

Venn diagram of previously reported housekeeping genes (Dixon et al., 2012), and their overlap with super-enhancer-associated genes in ESCs. The overlap is not significant, based on a hyper-geometric test.

Figure S3: Additional functional attributes of super-enhancers, Related to Figure 3.

A) Super-enhancers drive higher levels of expression than multiple typical enhancers. Expression levels (reads per kilobase of exon per million (RPKM)) of genes associated with super-enhancers are shown in red. Expression levels of genes associated with one or more typical enhancers are shown in white.

B) Presence of Klf4 and Esrrb correlates with luciferase activity. Plot of Klf4 and Esrrb ChIP-Seq signal (reads per million per base pair) at each of the 8 constituent typical enhancers and 8 constituent super-enhancers that were cloned into the luciferase reporter vector.

C) Loss of Oct4 results in ESC differentiation. Phase contrast images of ESCs infected with shRNAs targeting GFP or Oct4 for 5 days. Cells were stained with Hoescht and Oct4. ESC differentiation was determined based on changes in cellular morphology (loss of ESC colony and increase in the number of single cells with a flattened, dark appearance).

Figure S4: PU.1 and Mediator occupancy are highly correlated in pro-B cells, Related to Figure 4.

Scatter plot of all PU.1 (x-axis) and Mediator (y-axis) ChIP-Seq signal (total reads) across the 13,814 enhancers in pro-B cells.

Figure S5: Super-enhancers are found in multiple cell types, and are largely cell type-specific, Related to Figure 5.

A) Distribution of ChIP-Seq density (total reads) for master TFs (MyoD, T-Bet or C/EBPα) across the enhancers in each of their respective cell types. TF occupancy is not evenly distributed across the enhancers, allowing for the identification of super-enhancers. There are 535 super-enhancers in myotubes, 436 super-enhancers in Th cells, and 961 super-enhancers in macrophages.

B) Comparison of super-enhancers and typical enhancers. Metagenes of ChIP-Seq signal (reads per million per base pair) of MyoD, T-Bet or C/EBPα across typical enhancers and super-enhancers identified in either myotubes, Th cells, or macrophages, respectively. The metagenes are centered on the enhancer regions, with 3kb surrounding each enhancer region. ChIP-Seq fold difference (total reads) for MyoD, T-Bet or C/EBPα at super-enhancers versus typical enhancers is displayed below the metagenes.

C) Myotubes, Th cells, and macrophage super-enhancers neighbor genes critical for cell identity. Table of genes near super-enhancers previously shown to play important roles in either myotube, Th cell, or macrophage biology (Davis et al., 1987; Heinz et al., 2010; Lin et al., 2010; Szabo et al., 2000; Tapscott, 2005; Xie et al., 2004).
D) Super-enhancers are largely cell type-specific. *Leftmost panel*, Bar graph depicting the percentage of ESC super-enhancers that overlap a super-enhancer found in one or more cell types (ESC, pro-B, myotube, Th or macrophage cells). The red bar highlights the percentage of ESC super-enhancers found only in ESCs. The grey bars highlight the percentage of ESC super-enhancers that overlap (by at least one base pair) a super-enhancer found in at least one other cell type. *Remaining panels,* Bar graphs reflecting similar analyses performed for pro-B, myotube, Th and macrophage super-enhancers.

Table S1: Enhancers in ESCs, Related to Figure 1.
List of enhancers in ESCs, their locations, their associated genes, and background-subtracted total ChIP-Seq reads per million values for Oct4, Sox2, Nanog, Med1, H3K27Ac, Klf4 and Esrrb.

Table S2: Densities of ChIP-Seq reads in ESC super-enhancer and typical enhancer constituents, Related to Figure 1.
Densities of Oct4, Sox2, Nanog, Klf4, Esrrb, CTCF, c-Myc, Zfx, H3K27ac, H3K4me1, H3K4me3, Suz12, H3K9me3, Med1, Smc1, CDK8, CDK9, Brg1 and DNaseI hypersensitivity at OSN sites in ESCs were calculated as described in (Rahl et al., 2010). Densities were calculated as described below and corresponding WCE densities were subtracted, except for DNaseI hypersensitivity.

Table S3: Enhancers and corresponding Enhancer-Promoter Units (EPUs) and topological domains, Related to Figure 2.
List of all ESC enhancers, and their associated genes. Listed are the coordinates of the EPU containing the enhancer or gene (column 2,3) and whether the enhancer and gene fall into the same EPU (1=yes, column 4). Also listed are the coordinates of the topological domain wholly containing the enhancer or gene (column 5,6) and whether the enhancer and gene fall into the same topological domain (1=yes, column 7). Column 8 denotes if the enhancer is a super-enhancer.

Table S4: Gene expression from multiple experiments and treatments, Related to Figure 3.
RPKM and/or normalized Agilent expression data from wild-type ESCs, and 3, 4, and 5 days after Oct4 or Med12 shRNA knockdown.

Table S5: Enhancers in murine pro-B cells, associated genes, and RNA-Seq of associated genes, Related to Figure 4.
List of enhancers in pro-B cells, their locations, their associated genes, background-subtracted total ChIP-Seq reads per million values for PU.1 and Med1, as well as RNA-Seq RPKM for all associated genes.

Table S6: Enhancers in murine Th, myotube, and macrophage cells, associated genes, and RNA-Seq of associated genes, Related to Figure 5.
List of enhancers in myotubes, Th cells, and macrophages.

Table S7: Regions cloned into Luciferase expression constructs, Related to Figure 3.
List of all the primers used for cloning the *Oct4* promoter and the enhancer fragments. Endogenous sequences are shown capitalized.

Table S8: Datasets used in manuscript, Related to Figure 1.

The table below lists the datasets used in this work. All raw data can be downloaded from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/). Bold samples indicate data generated for this manuscript.

Data S1, Related to Figure 1

Data S1 contains .bed files for super-enhancers in ESCs, pro-B cells, myotubes, Th cells, and macrophages.

Extended Experimental Procedures

Cell Culture Conditions, and Oct4 and Mediator Knockdown Assays

Embryonic Stem Cells

V6.5 murine ESCs were grown on irradiated murine embryonic fibroblasts (MEFs), unless otherwise stated. Cells were grown under standard ESC conditions as described previously (Whyte et al., 2012). Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in ESC media; DMEM-KO (Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1000 U/ml LIF (ESGRO, ESG1106), 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/ml penicillin, 100 μg/ml streptomycin (Invitrogen, 15140-122), and 8 nL/ml of 2-mercaptoethanol (Sigma, M7522).

Lentiviral Production and Infection

Lentivirus was produced according to Open Biosystems Trans-lentiviral shRNA Packaging System (TLP4614). The shRNA constructs targeting Oct4 and Med12 are listed below and are available, including sequences, from Open Biosystems.

Oct4: TRCN0000009613
Med12: TRCN0000096467

The shRNA targeting GFP (TRCN0000072201, hairpin sequence: gtcgagctggacggcgacgta) was the negative control

ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and then plated in 6-well plates or 12-well plates. Cells were plated in 6-well plates at 1,000,000, 400,000 and 200,000 cells/well for the Oct4 knockdown microarray expression analyses. Cells were plated in 6-well plates at 200,000 cells/well for all Med12 knockdown microarray expression analyses. Cells were plated in 12-well plates at 80,000 cells/well for immunofluorescence. The following day cells were infected in ESC media containing 8μg/ml polybrene (Sigma, H9268- 10G) and plates were spun for 30 minutes at 2150 rpm. After 24 hours the media was removed and replaced with ESC media containing 3.5μg/mL puromycin (Sigma, P8833). ESC media with puromycin was changed daily. RNA was extracted for expression analysis on 3, 4 and 5 days post infection. Cells for imaging were crosslinked on day 5 post infection.

Immunofluorescence

Five days post infection cells were crosslinked for 15 minutes with 4% paraformaldehyde (EMS Diasum, 15710). Following crosslinking, the cells were washed once with PBS, twice with blocking buffer (PBS with 0.25% BSA, Sigma A3059-10G) and then permeabilized for 15 minutes with 0.2% Triton X-100 (Sigma, T8797-100ml). After two washes with blocking buffer cells were stained overnight at 4ºC for Oct4 (Santa Cruz Biotechnology, sc-5279; 1:100 dilution) and washed twice with blocking buffer. Cells were

incubated for 4 hours at room temperature with goat anti-mouse-conjugated Alexa Fluor 488 (Invitrogen; 1:200 dilution) and Hoechst 33342 (Invitrogen; 1:1000 dilution). Finally, cells were washed twice with blocking buffer and twice with PBS before imaging. Images were acquired on a Nikon Inverted TE300 with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition. Openlab and Photoshop CS3 Extended were used for image manipulation.

Cell Culture and RNA Isolation

For expression analysis, RNA was isolated from ESCs with TRIzol (Invitrogen, 15596-026), further purified with RNeasy columns (Qiagen, 74104) and DNase treated on column (Qiagen, 79254) following the manufacturer's protocols. RNA was then used for microarray expression analysis.

Chromatin Immunoprecipitation (ChIP)

ChIP Protocol

Protocols describing chromatin immunoprecipitation materials and methods have been previously described (Boyer et al., 2006). V6.5 ESCs were grown to a final count of 5-10 x 10^7 cells for each ChIP experiment. Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Cells were stored at -80^{o}C prior to use. Cells were resuspended, lysed in lysis buffers and sonicated to solubilize and shear crosslinked DNA. Sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

For Oct4, Sox2, Nanog and Mediator, the sonication buffer was 20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). Samples were kept on ice at all times. The resulting whole cell extract was incubated overnight at 4 degrees C with 100ul of Dynal Protein G magnetic beads that had been pre-incubated with approximately 10 ug of the appropriate antibody. Beads were washed 1X with the sonication buffer, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250mM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl.

Bound complexes were eluted from the beads (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65°C for 1 hour with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA reserved from the sonication step was also treated for crosslink reversal.

ChIP-Seq Sample Preparation and Analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203). A brief summary of the technique and minor protocol modifications are described below.

Sample Preparation

DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell.

After amplification, a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-350 bp (representing shear fragments between 50 and 250nt in length and ~100bp of primer sequence). The DNA was purified from the agarose and diluted to 10nM for loading on the flow cell.

Polony Generation and Sequencing

The DNA library (2-4 pM) was applied to the flow-cell using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1µm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing.

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Most primed flow-cells were loaded onto the HiSeq 2000. The whole cell extracts in myotubes and Pro-B cells, as well as CDK9 from ESCs were loaded onto the Genome Analyzer II. H3K4me3 and Suz12 from ESCs were loaded onto Genome Analyzer I. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced.

ChIP-Seq Data Analysis

All ChIP-Seq datasets were aligned using Bowtie (version 0.12.2) (Langmead et al., 2009) to build version NCBI37/MM9 of the murine genome. Alignments were performed with the following parameters: -n 2, -e 70, -m 1, -k 1, --best, with the additional parameter, -l, set to the read length of the data being aligned. Human ChIP-Seq reads were aligned to the hg18 genome using bowtie with options -n 2, -e 70, -m 2, -k 2, --best.

Identifying ChIP-Seq Enriched Regions

We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) (Zhang et al., 2008) peak finding algorithm to identify regions of ChIP-Seq enrichment over background. A p-value threshold of enrichment of 10^{-9} was used for all datasets. All other parameters were set as their defaults except for --keep-dup which was set to "auto". For murine ESCs, we calculated the strict overlap of Oct4, Sox2 and Nanog enriched regions to create Oct4/Sox2/Nanog co-bound enriched regions. For human ESCs, we calculated the enriched regions using Oct4 alone. The GEO accession number and background used for each dataset can be found in the "*Datasets Used in Manuscript*".

WIG files were created using MACS with options –w –S –space=50 to count reads in 50bp bins and then divided by the number of treatment reads after filtering to normalize to mapped-reads-per-million.

Definition of Enhancers

Constituent enhancers were defined as the enriched regions of master transcription factor(s) of a given cell type (See "Identifying ChIP-Seq enriched regions"). The enriched regions used to call constituent enhancers were Oct4/Sox2/Nanog co-bound, Oct4, PU.1,

MyoD, T-Bet, and C/EBPα for murine ESCs, human ESCs, pro-B, myotubes, Th cells, and macrophages, respectively. We noted that there were often closely spaced enriched regions with very high signal, and we wished to capture that whole span as a single region. Thus, we further combined the constituent enhancers that occurred within 12.5kb of each other into a single larger enhancer domain. We used a distance of 12.5kb based on an analysis of the enriched regions in murine ESCs. In that dataset we found that a distance of 12.5kb was optimal for stitching together the closely spaced enriched regions with very high signal while not being so large as to stitch together the more widely spaced regions with lower signal.

The coordinates of each set of stitched enhancers can be found in Table S1 for ESCs, Table S5 for pro-B cells, and Table S6 for myotubes, Th cells, and macrophages.

Identifying Super-enhancers

To identify super-enhancers, we first ranked all enhancers in a cell type by increasing total background-subtracted ChIP-Seq occupancy of Med1, and plotted the total background-subtracted ChIP-Seq occupancy of Med1 in units of total rpm/bp (reads per million per base pair)(Figure 1C, 4B). In cases where Med1 ChIP-Seq data was not available, we used the total background subtracted ChIP-Seq occupancy of the master regulator instead (Figure S8 and Figure S11). These plots revealed a clear point in the distribution of enhancers where the occupancy signal began increasing rapidly. To geometrically define this point, we first scaled the data such that the x and y axis were from 0-1. We then found the x-axis point for which a line with a slope of 1 was tangent to the curve. We define enhancers above this point to be super-enhancers, and enhancers below that point to be typical enhancers. The classification of enhancers in each cell type as a super-enhancer or typical enhancer can be found in Table S1 for ESCs, Table S5 for pro-B cells, and Table S6 for myotubes, Th cells, and macrophages.

We calculated background subtracted total reads per million of OSN, DNaseI, H3K27ac, Med1, and H3K4me1 at the 8,794 ESC enhancers. We then normalized the signal such that the maximum was 1.0 for each factor. We then sorted the enhancers and visualized the distribution, zooming in on the bottom right corner of the plot for greater clarity (Figure 1E, Figure S2). Med1 has the sharpest transition between the two populations and was therefore considered “optimal”.

We also investigated whether the enhancer feature H3K27ac alone can be used to identify super-enhancers. We found enriched regions of H3K27ac ChIP-Seq data, and clustered the binding peaks as described in the *Definition of Enhancers* section. We then ranked these domains by H3K27ac ChIP-Seq signal, and used the tangent of the curve to define two enhancer populations, as described above. This analysis identifies 725 candidate super-enhancers. Of these, 155 were previously identified using OSN and Mediator (Figure 1C). Hence, the enhancer feature H3K27ac alone cannot directly substitute for the master transcription factors and Mediator in our analysis pipeline.

Enrichment of Factors at Super-Enhancers

For the analysis of enrichment signal in super-enhancers constituents versus typical enhancer constituents (Figure 1H), densities were calculated in constituents as described in (Rahl et al., 2010). Briefly, ChIP-Seq reads aligning to the region were extended by 200 base pairs and the density of reads per base pair was calculated. In order to eliminate PCR bias, multiple reads of the exact same sequence aligning to a single position were collapsed into a single read. Only positions with at least two overlapping extended reads contributed to the overall region density. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of

reads per million mapped reads per base pair (rpm/bp). Densities were then background subtracted and the resulting values are shown in Table S2.

Assigning Enhancers to Genes

We assigned enhancers to genes defined in the RefSeq (NCBI37/MM9)(Pruitt et al., 2007) gene annotations. To assign each enhancer to a gene, we calculated the distance from the center of the enhancer to the TSS of each gene. The enhancer was then assigned to the closest gene. The assignment of enhancers to genes in each cell-type can be found in Table S1 for ESCs, Table S5 for pro-B cells, and Table S6 for myotubes, Th cells, and macrophages.

To determine agreement of proximity assignment of super-enhancer to genes with other methods, we examined topological domains (Dixon et al., 2012) and enhancer-promoter units (EPUs)(Shen et al., 2012). For comparison to EPU's, we calculated the percent of super-enhancers where the super-enhancer and the TSS of its nearest gene overlapped the same EPU. For the super-enhancers that overlap an EPU, 95% of those EPUs also overlap its proximal gene (Table S3). We also determined which topological domains contained the super-enhancer and which contained their proximal gene. 93% of super-enhancers were found in the same topological domain as their proximally assigned gene (Table S3).

Gene Set Enrichment Analysis (GSEA)

We used Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) to determine whether the set of super-enhancer associated genes was statistically enriched for genes that were important for the maintenance of ES cell state. Our gene set consisted of genes associated with super-enhancers. The Z-score for a gene was calculated as the average Z-score of all shRNAs associated to that gene from Kagey et al. (Kagey et al., 2010).

Motif Analysis

To find sequence motifs enriched in super-enhancers in murine ESCs,human ESCs and pro-B cells, we analyzed the genomic sequence under the MACS-defined master regulator enriched regions that were located within super-enhancers. We extracted their sequence from either the mm9 or hg18 genome and used this as input for TRAP using TRANSFAC vertebrates as the comparison library, mouse or human promoters as the control, and Benjamini-Hochberg as the correction (Thomas-Chollier et al., 2011). P-values displayed in Figure 1I and 4D correspond to the corrected p in the output.

To calculate presence of motifs in a given constituent enhancer, we used FIMO with a custom library of all Transfac (and Jaspar) motifs at a p-value threshold of 10^{-4} (Grant et al., 2011; Matys et al., 2006). The number of occurrences of the different groups of motifs was summed for each region and enrichment differences were calculated using a two-tailed t-test.

Matrices used:
Oct4: M01124
Sox2: M01272
Nanog: M01123
Klf4: M01588
Esrrb: M01589
CTCF: M01259
c-Myc: M01154
Ebf1: M00977

Foxo1: M00474
PU.1: M01203
E2A: M00973
Zfx: M01593

ChIP-Seq Metagenes

Genome-wide average "meta" representations of Med1 (Figure 1D, 4B, the master regulators for Figure S8) ChIP-Seq occupancy at typical enhancers and super-enhancers were created by mapping Med1 ChIP-Seq read density to the enhancer regions and their corresponding flanking regions. Each enhancer or flanking region was split into 100 equally sized bins. This split all enhancer regions, regardless of their size, into 300 bins. All typical enhancer or super-enhancer regions were then aligned and the average Med1 ChIP-Seq density in each bin was calculated to create a meta genome-wide average in units of reads per million per base pair. In order to visualize the length disparity between typical and super-enhancer regions, the enhancer region (between its actual start and end) was scaled relative to its median length.

Metagenes shown in Figure 1G were created in a similar fashion. Constituent enhancers of super and typical enhancers, as well as 3kb upstream and downstream regions were each broken into 50 bins. The ChIP-Seq density in these regions was calculated and combined together to get 150 bins spanning 3kb upstream region, the constituent enhancer, and 3kb downstream region. The average combined profiles for the super- or typical enhancer constituents is shown.

Gene Ontology (GO) Analysis

To find the GO terms enriched in super-enhancer-associated genes we used the DAVID web-tool with super-enhancer-associated genes. (Huang da et al., 2009). For Figure 2E, Gene Ontology Molecular Function terms were analyzed and terms associated with transcription factor activity were the top terms returned in the analysis. For Figure 5C and Figure S10, we report the first 10 most significant Gene Ontology Biological Process terms and their associated Bonferonni p-values.

Gene Expression Sample Preparation and Analysis

RNA-Seq Data Analysis

The ESC RNA-Seq data in Table S3, the pre-computed values were downloaded from http://chromosome.sdsc.edu/mouse/download.html. For the pro-B RNA-Seq, sequences were aligned using Bowtie (version 0.12.2) to build version NCBI37/MM9 of the murine genome. The RPKM (reads per kilobase of exon per million mapped reads) was then computed for each gene (Table S5).

Agilent Sample Preparation and Data Analysis

For microarray analysis, Cy3 and Cy5 labeled cRNA samples were prepared using Agilent's QuickAmp sample labeling kit starting with 1ugtotal RNA. Briefly, double-stranded cDNA was generated using MMLV-RT enzyme and an oligo-dT based primer. In vitro transcription was performed using T7 RNA polymerase and either Cy3-CTP or Cy5-CTP, directly incorporating dye into the cRNA. Agilent mouse 4x44k expression arrays were hybridized according to our laboratory's standard method, which differs slightly from the standard protocol provided by Agilent. The hybridization cocktail consisted of 825 ng cy-dye labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization cocktails were fragmented at 60°C for 30 minutes, and then Agilent 2X hybridization buffer was added to the cocktail prior to application to the

array. The arrays were hybridized for 16 hours at 60°C in an Agilent rotor oven set to maximum speed. The arrays were treated with Wash Buffer #1 (6X SSPE / 0.005% n-laurylsarcosine) on a shaking platform at room temperature for 2 minutes, and then Wash Buffer #2 (0.06X SSPE) for 2 minutes at room temperature. The arrays were then dipped briefly in acetonitrile before a final 30 second wash in Agilent Wash 3 Stabilization and Drying Solution, using a stir plate and stir bar at room temperature.

Arrays were scanned using an Agilent DNA microarray scanner. Array images were quantified and statistical significance of differential expression for each hybridization was calculated using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol. To calculate an average dataset from the biological replicates the log2 ratio values for each Agilent Feature were averaged (Supplementary Table 4). For each gene we selected the Agilent Feature with the best average p-value that was annotated to that gene. Genes with no annotated features were reported as N/A.

Generation and Expression of Luciferase Expression Constructs in ESCs

A minimal Oct4 promoter was amplified from mouse genomic DNA and cloned into the XhoI and HindIII sites of the pGL3 basic vector (Promega). Enhancer fragments were amplified from mouse genomic DNA, and cloned into the BamHI and SalI sites of the pGL3-pOct4 vector. All primers used are listed in Table S3. 500μg of the plasmids were used to transfect $2x10^5$ murine ESCs in 24-well plates using Lipofectamine 2000 (Invitrogen) according to the manufacturer´s instructions. The amount of transfected plasmid was adjusted according to its size whenever necessary. 10μg of the pRL-SV40 plasmid was co-transfected in each well as a normalization control. Cells were incubated for 24 hours, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). Luciferase activity was normalized to the activity measured in cells transfected with a construct containing only the Oct4 promoter. Experiments were performed in triplicates.

Other Data Analysis

Significance Calculations in Box Plots

The P-value for the difference between samples was compute using a two-sided t-test for all panels.

Supplementary References

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

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

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature.

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

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J.*, et al.* (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat Immunol *11*, 635-643.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K.*, et al.* (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic acids research *34*, D108-110.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E.*, et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 267-273.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Szabo, S.J., Kim, S.T., Costa, G.L., Zhang, X., Fathman, C.G., and Glimcher, L.H. (2000). A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell *100*, 655-669.

Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nature protocols *6*, 1860-1869.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.

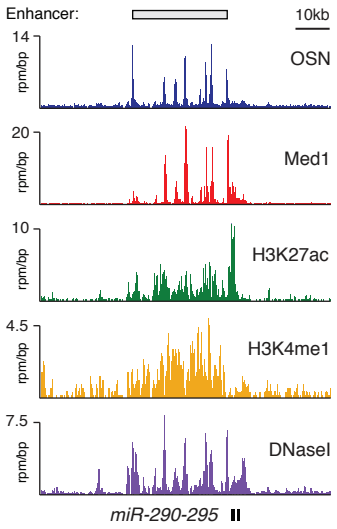
Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Supplemental Figure 1

Figure S1

A



B

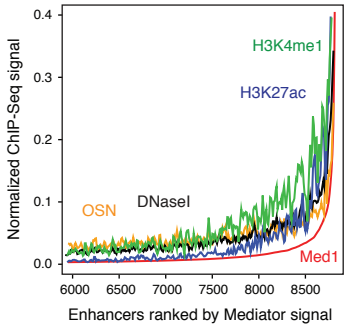


Figure S2

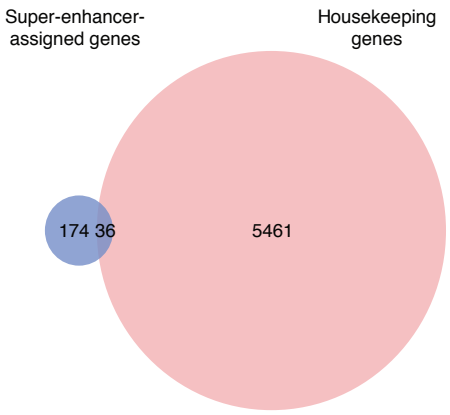
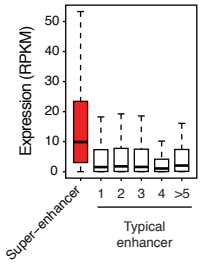
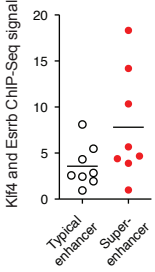

Super-enhancer-assigned genes
Housekeeping genes
174
36
5461

Figure S3

A



B



C

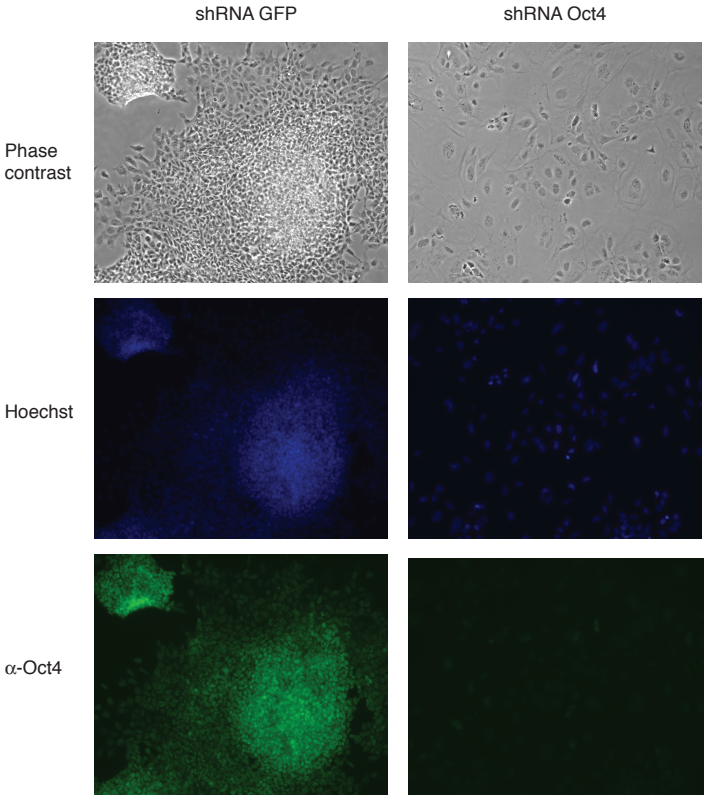


Figure S4

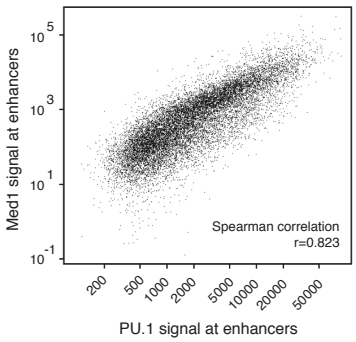
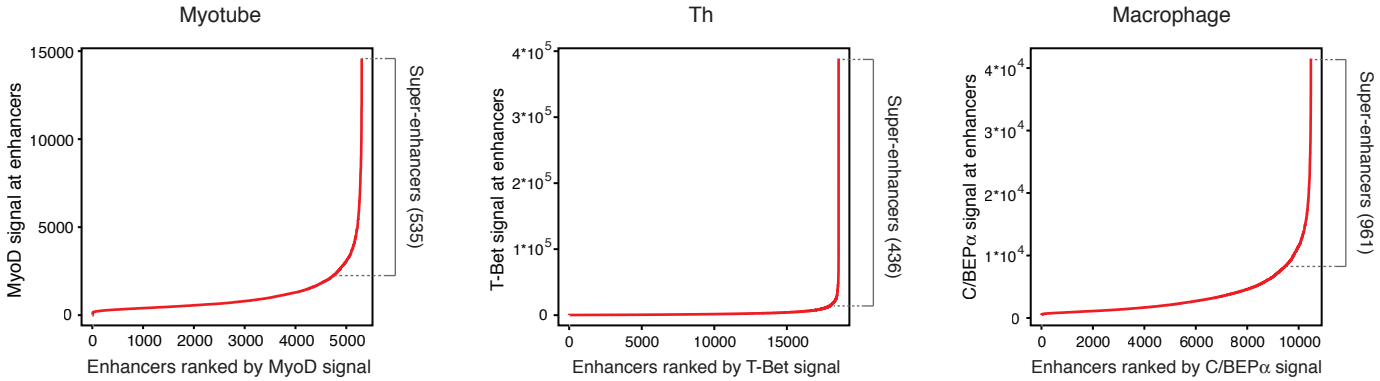
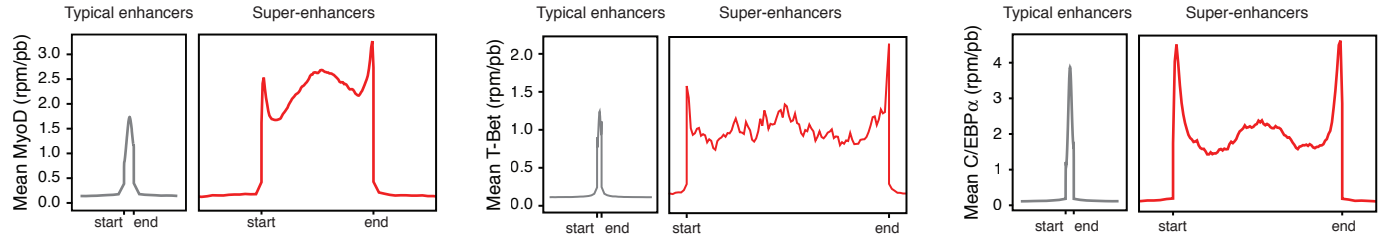

Med1 signal at enhancers
10^{5}
10^{3}
10^{1}
10^{-1}
Spearman correlation
r=0.823
200
500
1000
2000
5000
10000
20000
50000
PU.1 signal at enhancers

Figure S5

A



B



	Typical enhancers	Super-enhancers
Number:	4769	535
Median size:	362 bp	4169 bp
MyoD:	1x	5x

	Typical enhancers	Super-enhancers
Number:	18129	436
Median size:	976 bp	43900 bp
T-Bet:	1x	12x

	Typical enhancers	Super-enhancers
Number:	9514	961
Median size:	643 bp	13400 bp
C/EBPα:	1 unit	5x

C

Selected genes associated with super-enhancers

Myotube	Th	Macrophage
MyoD, MyoG	*T-Bet, IRF8*	*CD68, CD31, Itgb2, IRF8*

D