SUPPLEMENTAL FIGURES

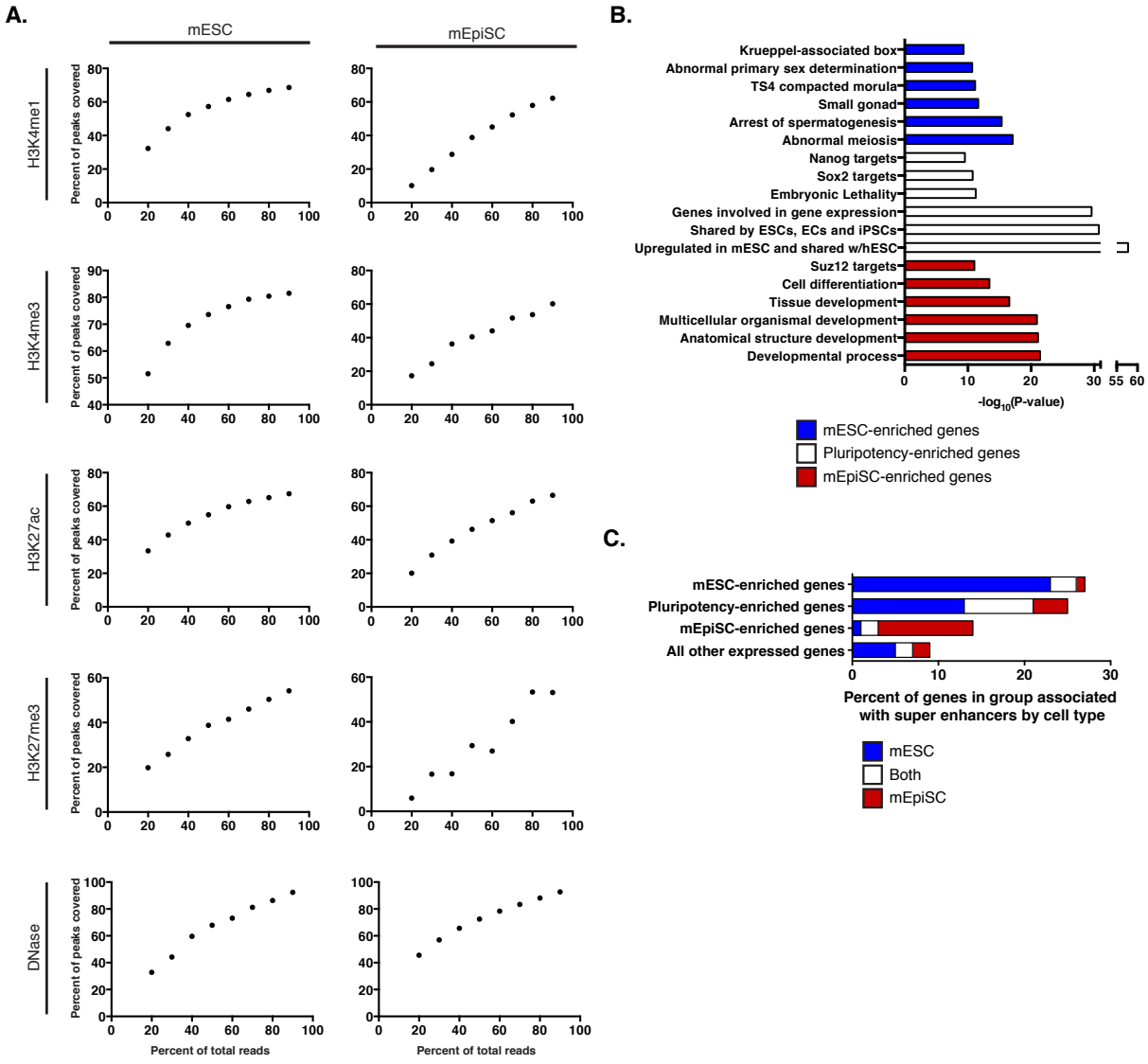

Figure S1. Saturation of chromatin state experiments and features of gene sets, related to Figure 1
(A) Saturation curves for mESC and mEpiSC histone modification ChIP-seq datasets and DNase-seq datasets. Plots indicate the percent of peaks between 0- and 20-fold enrichment that are covered when subsampling reads to a given percentage of total depth. (B) Selected gene ontology terms associated with mESC-enriched genes (blue), mEpiSC-enriched genes (red), and pluripotency enriched genes (white). (C) The percentage of each gene list associated with a super enhancer in either mESCs (blue), mEpiSCs (red), or both (white).

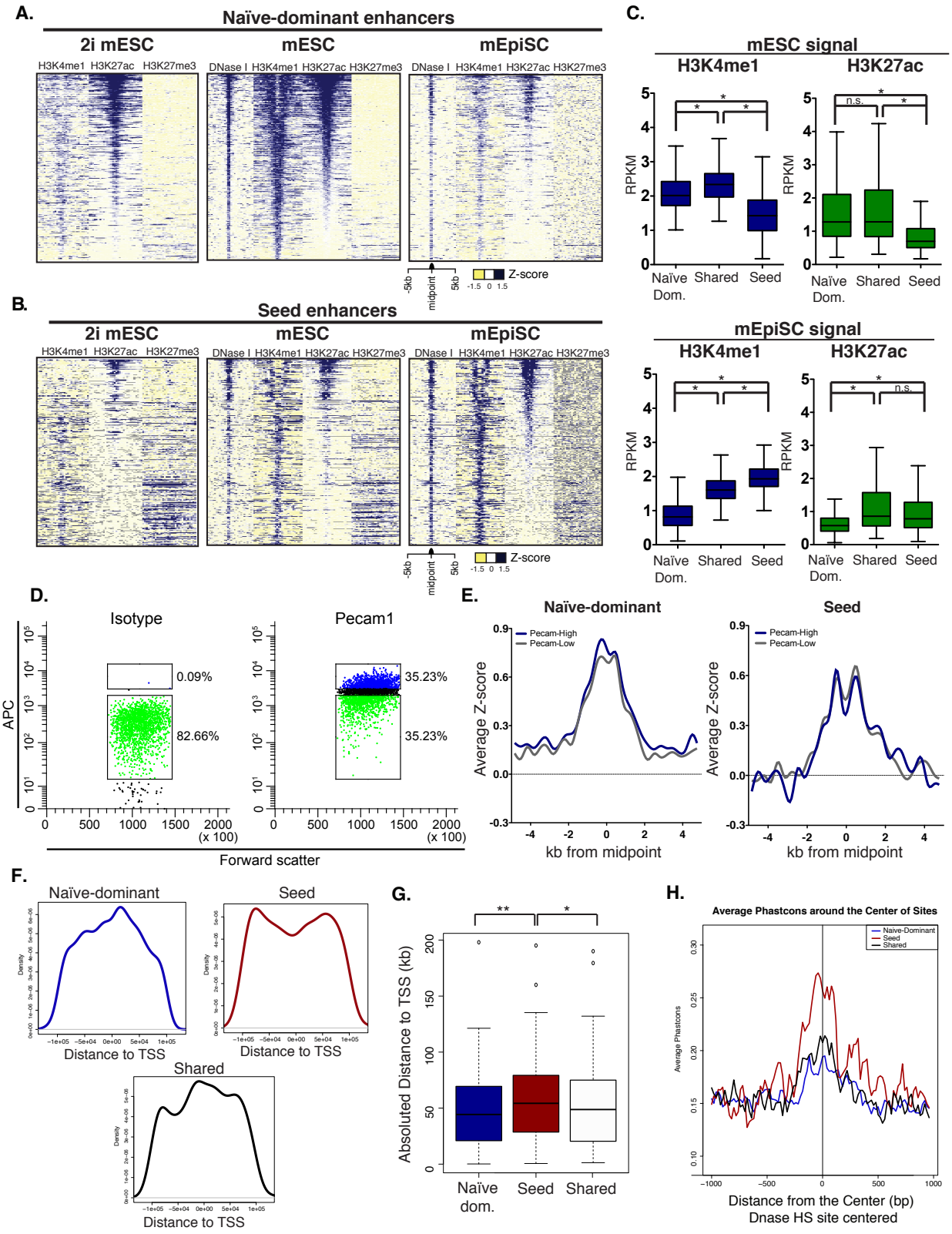

Figure S2. Characterization of enhancers of pluripotency-enriched genes, related to Figure 2

(A) Windowed heatmaps displaying Z-scored ChIP-seq signal intensity for H3K4me1, H3K27ac and H3K27me3 as well as DNase HS at naïve-dominant enhancers in mESCs grown in 2i conditions (left) mESCs grown in serum (center) and mEpiSCs (right). (B) Same as (A) for seed enhancers. (C) ChIP-seq signal intensity measured in RPKM for naïve-dominant, shared and seed enhancers for H3K4me1 (blue) and H3K27ac (green) in mESCs (top) and mEpiSCs (bottom) (Mann-Whitney-Wilcoxon test, *p-value<0.0001). (D) Representative flow cytometry plots following staining with isotype control antibody (left) and Pecam1 antibody (right). Sorted populations used for ChIP sequencing are indicated by black boxes (Pecam1-high and Pecam1-low). (E) Aggregate plots comparing H3K4me1 enrichment at naïve-dominant enhancers (left) and seed enhancers (right) between Pecam1 low and Pecam1 high populations of mESCs. (F) Distribution plots of distance between enhancers and the TSSs they are predicted to regulate using the PreSTIGEouse method, divided by class. (G) Distributions of absolute distance between TSSs and their predicted enhancers divided by enhancer class. (Wilcoxon-test, *p-value<0.003, **p-value<2.6E-9) (H) Conservation of enhancers by class, as measured by the PhastCons phylogenetic hidden Markov model in a 2-kb window surrounding underlying regions of DNase hypersensitivity.

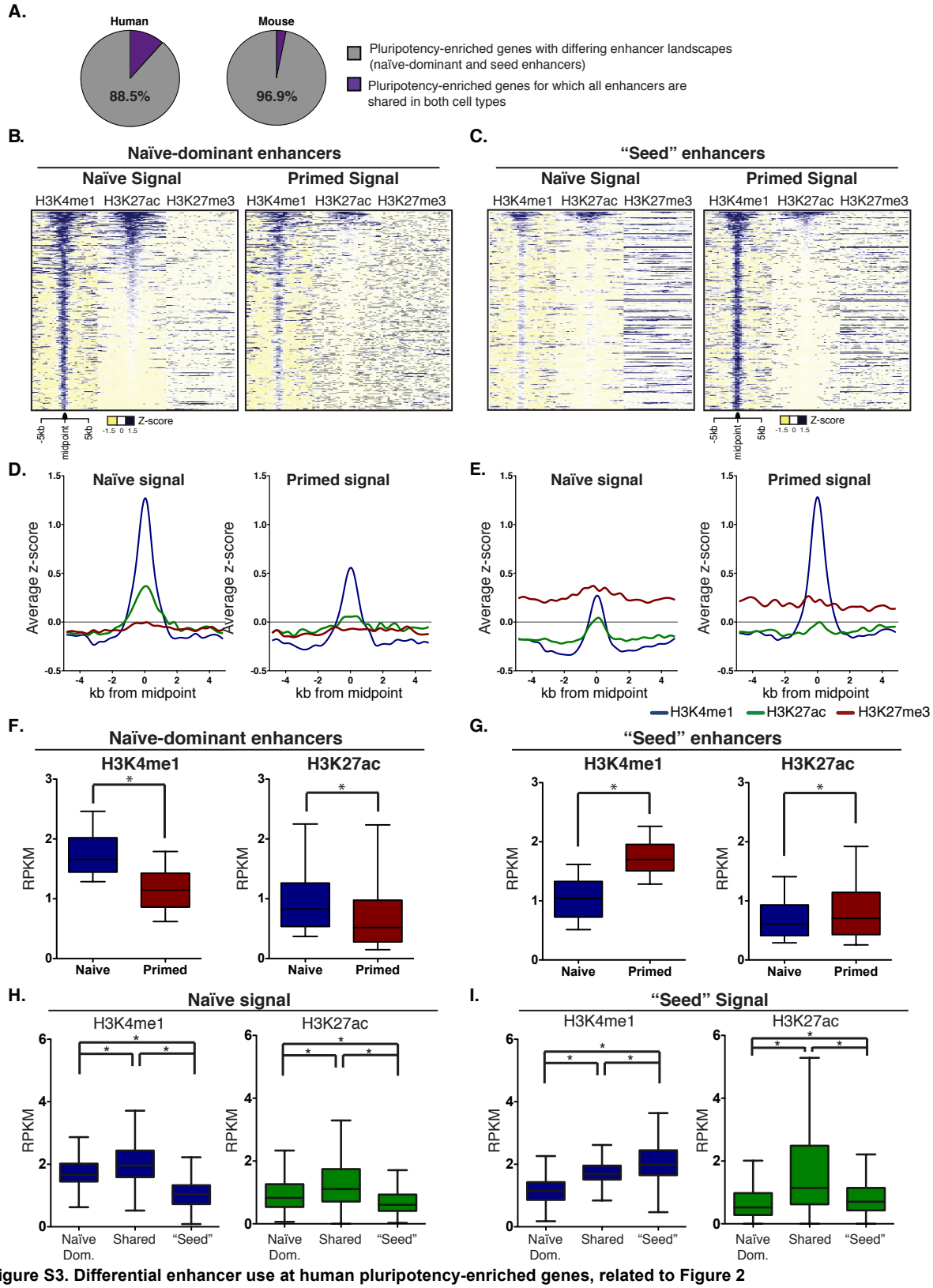

Figure S3. Differential enhancer use at human pluripotency-enriched genes, related to Figure 2
(A) Proportion of pluripotency-enriched genes for which the enhancer landscape has no differences between naïve and primed cell types (purple) compared to the proportion of genes with enhancers that differ between the two cell types (grey), displayed for both human (left) and mouse (right). (B) Windowed heatmaps displaying Z-scored ChIP-seq signal intensity for H3K4me1, H3K27ac and H3K27me3 for naïve-dominant enhancers in naïve and primed hESCs (C) Same as (B) for "seed"-like enhancers. (D) Aggregate plots depicting average chromatin modification signal at naïve-dominant enhancers in naïve hESCs (left) and primed hESCs (right). (E) Aggregate plots of chromatin modifications for "seed"-like enhancers in naïve cells (left) and primed cells (right). Boxplots depicting differences in active enhancer marks as measured by RPKM between the two cell types at naïve-dominant enhancers (F) and "seed"-like enhancers (G) (paired sample Wilcoxon signed rank test, *p-value<0.0001). ChIP-seq signal intensity in RPKM for naïve-dominant, shared and "seed"-like enhancers for H3K4me1 (blue) and H3K27ac (green) in naïve cells (H) and primed cells (I) (Mann-Whitney-Wilcoxon test, *p-value<0.0001).

SUPPLEMENTAL TABLES

Table S1. Summary of publicly available and new datasets used in this study, related to Figure 1

Table S2 Genes classified as mESC-enriched, mEpiSC-enriched, and pluripotency-enriched in this study, related to Figure 1

Table S3 Motifs enriched at all naïve-dominant enhancers, all seed enhancers, and seed enhancers that are active in downstream tissues, related to Figure 2

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture. Pluripotent stem cells were cultured as described previously (Chenoweth and Tesar, 2010; Najm et al., 2011; Tesar et al., 2007). In brief, mESCs were grown in Knockout DMEM (Invitrogen) supplemented with 20% Knockout Serum Replacement (KSR) (Invitrogen), 1000 units/ml recombinant murine Leukemia Inhibitory Factor (LIF) (ESGRO, Millipore), 2mM L-alanyl-L-glutamine dipeptide (Glutamax; Invitrogen), 1x non-essential amino acids (Invitrogen), and 0.1mM 2-mercaptoethanol (Sigma). mESCs were passaged every third day as a single cell suspension using 0.25% trypsin/EDTA and seeded at $3.0x10^4$ cells/cm^2 for routine culture. mEpiSCs were cultured in Knockout DMEM supplemented with 20% KSR, 10ng/ml FGF2 (R&D Systems), 2mM Glutamax, 1x non-essential amino acids, and 0.1mM 2-mercaptoethanol and passaged every third day using 1.5 mg/ml collagenase type IV (Invitrogen) and trituration into small clumps of ~10-100 cells. Irradiated mouse embryonic fibroblasts (MEFs) served as a feeder layer for all pluripotent cell types and were maintained with DMEM supplemented with 10% fetal bovine serum, 2mM glutamax, 1x non-essential amino acids, and 0.1mM 2-mercaptoethanol. All cells were grown on Nunclon Δ-treated dishes or multiwell plates (Fisher Scientific) coated for 2 hours at 37°C with 0.1% (w/v) porcine gelatin (Sigma).

ChIP-seq. Histone modification ChIPs were performed from $5x10^6$ crosslinked mEpiSCs and sequencing libraries were prepared as previously described (Schmidt et al., 2009). The following antibodies were used for ChIP: rabbit anti-H3K4me1 (Abcam #8895), rabbit anti-H3K4me3 (Active Motif #39159), rabbit anti-H3K27ac (Abcam #4729), mouse anti-H3K27me3 (Abcam #6002). ChIP-seq libraries were sequenced on the Illumina GA II platform at the Case Western Reserve University Genomics Core Facility.

The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used to remove adapter sequences and trim read ends using a quality score cutoff of 20. ChIP-seq data were aligned to mm9 or hg19 genome assemblies (retrieved from http://cufflinks.cbcb.umd.edu/igenomes.html), using Bowtie v0.12.9 (Langmead et al., 2009), allowing reads with ≤ 2 mismatches and discarding reads with > 1 reportable alignment ("-m 1" parameter). PCR duplicates were removed using SAMtools (Li et al., 2009). Peaks were detected with MACS v1.4 (Zhang et al., 2008), using an aligned input DNA sample as control. Wiggle tracks stepped at 25 bp were

generated by MACS, normalized to the median signal by chromosome, converted to bigwig format using the UCSC Genome Browser's wigToBigWig utility (retrieved from http://hgdownload.cse.ucsc.edu/admin/exe/) and visualized on the UCSC Genome Browser. Quality control metrics for ChIP datasets are available in Table S1.

DNase-seq. $1x10^7$ cells were subjected to DNase sequencing as previously described (Song et al., 2011) with the exception of a 5' phosphate added to linker 1B to increase ligation efficiency. Libraries were sequenced on the Illumina HiSeq 2000 platform at the Duke IGSP Genome Sequencing and Analysis Core.

Adapters and low quality bases (Phred score <20) were trimmed from the end of reads using the FASTX-Toolkit. Reads were aligned to the mm9 genome assembly using Bowtie v0.12.9, allowing no more than 2 mismatches and randomly assigning multi-mapped reads to one alignment site ("-M 1" parameter). SAMtools was used to remove PCR duplicates. Peaks were identified using MACS v1.4. Tracks were visualized as described for ChIP-seq data.

RNA-seq. RNA was extracted from ~$2x10^6$ cells with TRIzol (Invitrogen), separated using Phase Lock Gel tubes (5 Prime), and purified using the RNAeasy Plus kit (Qiagen) according to the manufacturer's protocol. PolyA+ RNA was prepared for sequencing using the Illumina TruSeq RNA Sample Preparation Kit according to the manufacturer's protocol. RNA-seq libraries were sequenced on the Illumina HiSeq 2000 platform at the Oklahoma Medical Research Foundation's Next Generation Sequencing Core Facility.

For gene expression analysis, reads were aligned to mm9 or hg19 genome builds (retrieved from http://cufflinks.cbcb.umd.edu/igenomes.html), using Tophat v2.0.6 (Trapnell et al., 2009). The distribution of alignments was analyzed using the CollectRnaSeqMetrics module of Picard v1.89 (http://picard.sourceforge.net/). QC metrics for RNAseq datasets are available in Table S1.

FPKM values for known genes were calculated using Cufflinks v2.0.2 (Trapnell et al., 2010) provided with the GTF file via the -G (known genes only) option. FPKM values were tabled by converting background values (<0.25) to 0 and adding 0.25 to all values (Ramskold et al., 2009). Differential expression testing was performed using Cuffdiff v2.0.2; however, all FPKM values provided are those calculated by Cufflinks. In comparisons of mESC and mEpiSC expression, FPKMs were quantile normalized across all eight samples. In comparisons of mESC, mEpiSC and downstream tissues, FPKMs were quantile normalized across the eight pluripotent cell type samples, as well as all 19 downstream tissues.

Fluorescence-activated cell sorting. Single-cell suspensions of mESC were blocked in 10% normal donkey serum for 30 min., then stained for 30 min. with APC conjugated Pecam1 antibody (BD Bioscience, 551262, 1:100) or APC conjugated isotype control (BD Bioscience, 553932, 1:100). Cells were sorted on a FACSAria flow cytometer (BD Biosciences). Gates were set to isolate Pecam1-high and Pecam1-low populations. $1x10^6$ cells from each population were subjected to immediate fixation and ChIP sequencing was carried out according to the method described.

Defining gene sets. Genes were considered enriched in mESCs or mEpiSCs given Cuffdiff differential expression significance, FPKM > 1 in at least one replicate of the cell type, and higher expression in that cell type than the comparator. This produced lists of differentially expressed genes enriched in mESCs (n = 348) and mEpiSCs (n = 504).

Pluripotency-enriched genes were selected based on specificity of transcript levels to mESC and mEpiSC as compared to a panel of mouse tissues (shown in Figure 2A). Relative specificity was determined by calculating Shannon Entropy Q scores. Genes with Cuffdiff p-values <0.05 were removed from consideration. Finally, genes with >2-fold difference between mEpiSC and mESC average transcript levels were also removed. For human analyses, pluripotency genes were selected based on specificity of mRNA transcript to the naïve and primed cell types compared to 11 human cell types (GM12878, HeLa, HepG2, HMEC, HSMM HUVEC, K562, NHEK, NHLF, NPC, and MCF7). Additionally, genes with >1.3 fold difference between naïve and primed expression levels were removed.

Gene ontology. Gene ontology analysis of mESC-, mEpiSC-, and pluripotency-enriched gene TSSes was performed using GREAT (McLean et al., 2010).

Super enhancers. Super enhancers were identified using the ROSE software package v0.1 (retrieved from https://bitbucket.org/young_computation/rose) (Loven et al., 2013; Whyte et al., 2013). In each cell type, H3K27ac peaks identified by MACS were filtered to remove peaks within 2kb of RefSeq TSSes, and then peaks separated by less than 12.5kb were stitched together. All stitched peaks were then ranked by the density of H3K27ac minus input. Peaks higher than the inflection point on the density curve were designated super enhancers.

Prediction of gene targets of enhancer elements using PreSTIGEouse. Enhancer-gene assignments were made as described in (Corradin et al., 2014). Briefly, predictions were made using comparative analysis across a panel of 13 tissues. For an interaction to be predicted the normalized H3K4me1-enhancer signal intensity had to be above background and highly specific to the cell line of interest compared to the remaining 12 cell lines. Additionally, the gene must be within 100-kb of the enhancer and must show relatively cell type-specific transcript levels. PreSTIGEouse predictions for mESC and mEpiSC were made using independent comparisons to a panel that included bone marrow, cerebellum, embryonic heart, intestine, kidney, liver, lung, MEF, olfactory bulb, placenta, testis and thymus (selected for tissue diversity). Predictions were made at both low and high stringency (based on relative specificity scores); predictions used in the manuscript are high stringency unless otherwise noted.

Determination of naïve-dominant and seed enhancers. Naïve-dominant, seed, and shared enhancers were determined by comparing predicted enhancer elements of pluripotency-enriched genes. Naïve-dominant enhancers were defined as high stringency PreSTIGEouse enhancer predictions of pluripotency-enriched genes in mESCs that were not predicted to target the pluripotency-enriched genes in mEpiSC (at high or low

stringencies). Seed enhancers were determined by identifying high stringency enhancer predictions in mEpiSC, which were not predicted in mESCs. Shared enhancers are those that are predicted to regulate pluripotency-enriched genes in both mEpiSC and mESC.

Motif analysis. To identify transcription factor motifs enriched in enhancer sets, enhancers were centered on DNase hypersensitivity sites, and the SeqPos module of the Cistrome tool was used to scan a 1kb window for enriched curated and *de novo* motifs (Liu et al., 2011).

Physical interaction of promoters of pluripotency-enriched genes with enhancer classes. To functionally evaluate the likelihood of the promoters of pluripotency-enriched genes to physically interact with naïve-dominant and seed enhancers we utilized high-depth Hi-C datasets generated in mESC. Experiments were obtained from (Selvaraj et al., 2013). Paired reads were aligned separately to mm9 using bowtie2 (Langmead and Salzberg, 2012). The Homer software Hi-C analysis package (including makeTagDirectory.pl and findHiCInteractionsByChr.pl) was utilized to identify significantly interacting genomic loci (Heinz et al., 2010). To identify intrachromosomal interactions the number of read pairs that associated two distinct 3-kb windows was compared to a background model of read-pair alignments. We next identified significant interactions for which the TSS of a pluripotency-enriched gene was contained within one of the paired windows and then determined how often these genes were physically associated with naïve-dominant enhancers and how often they were associated with seed enhancers.

Activity of enhancer classes in downstream tissues. We utilized H3K4me1 and H3K27ac ChIP-seq experiments for 15 mouse tissues (Figure 3B). For every enhancer of each enhancer class we calculated the normalized maximum read depth for H3K4me1 and H3K27ac in the downstream tissues. We classified activity by robust H3K27ac signal (normalized read depth>20) for each enhancer window in each downstream tissue. The same analysis was repeated using H3K4me1 above threshold (normalized read depth >20) as well as both H3K4me1 and H3K27ac above threshold to classify activity (data not shown). Enhancer clusters were defined by identifying 100-kb windows that contained four or more enhancer elements (as defined by PreSTIGEouse) in a given tissue.

SUPPLEMENTAL REFERENCES

Chenoweth, J.G., and Tesar, P.J. (2010). Isolation and maintenance of mouse epiblast stem cells. Methods Mol Biol *636*, 25-44.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., and Scacheri, P.C. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome research *24*, 1-13.

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. Molecular cell *38*, 576-589.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

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 Biology *10*, R25.

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

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome biology *12*, R83.

Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320-334.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. Cell stem cell *8*, 318-325.

Ramskold, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol *5*, e1000598.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240-248.

Selvaraj, S., J, R.D., Bansal, V., and Ren, B. (2013). Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nature biotechnology *31*, 1111-1118.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., 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 research *21*, 1757-1767.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

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

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology *28*, 511-515.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

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 Biology *9*, R137.