EXTENDED EXPERIMENTAL PROCEDURES

Cell Culture

All in vitro derived cell types were derived from HUES64 (Chen et al., 2009). Human embryonic stem cells were expanded on murine embryonic fibroblasts (Global Stem) in KO-DMEM (Life Technologies) containing 20% Knockout serum replace (Life Technologies) and FGF2 (10 ng/ml) (Millipore). Cultures were passaged by enzymatic dissociation using Collagenase IV (1 mg/ml) (Life Technologies). Prior to differentiation, cells were plated on matrigel-coated plates (BD Biosciences) and cultured in mTeSR1 (Stem Cell Technologies) for 3 to 4 days. Endoderm differentiation was induced in Advanced RPMI (Invitrogen), 0.5% FBS (Hyclone), Activin A (100 ng/ml) (R&D) and WNT3A (50 ng/ml) (R&D). HUES64- derived hepatoblasts (dHep) were induced by culturing day 5 endoderm in RPMI media containing B27 (1X), FGF2 (10ng/ml) (Millipore) and BMP4 (20 ng/ml) (R&D) for five days, and collected after 10 days total of differentiation. Hepatocyte-like cells were derived by culturing the HUES64-derived hepatoblasts in Lonza hepatocyte culture media containing 10ng/ml of HGF (R&D) for 5 additional days, or 15 days total. Mesoderm differentiation was induced by the addition of media consisting of in DMEM/F12 (Life Technologies), 0.5% FBS (Hyclone), Activin A (100ng/ml) (R&D) (for the first 24 hr only), BMP4 (100 ng/ml) (R&D), VEGF (100 ng/ml) (R&D) and FGF2 (20 ng/ml) (Millipore). To induce osteoblast differentiation, the day 5 mesoderm population was dissociated with accutase and replacted on matrigel coated plates (BD) in EGM-2 media (Lonza) for 7 days, or 12 days total. Ectoderm differentiation was induced using A83-01 (2um) (Tocris), PNU 74654 (2um) (Tocris) and Dorsomorphin (2um) (Tocris), DMEM/F12 (Life Technologies) containing 15% Knock serum replacer (Life Technologies). Neurectoderm differentiation was induced by switching the day 5 ectoderm population to media containing 3 μM CHIR99021 (TOCRIS), 10 μM SU5402 (TOCRIS), and 10 μM DAPT (TOCRIS), and collected after 6 more days, or 11 days total. N2-supplement (Life Technologies) was added to cells in 25% increments every other day beginning four days after the initiation of ectoderm differentiation. For all cell types, media was changed daily.

Antibodies

ChIP was performed using the following antibodies: H3K4me3 (Millipore, 07-473, Lot DAM1623866), H3K27ac (Abcam, ab4729, Lot 509313), H3K27me3 (Millipore, 07-449, Lot DAM1514011), H3K36me3 (Abcam, ab9050, Lot 499302), H3K4me1 (Abcam, ab8895, Lot 659352), H3K9me3 (Abcam, ab8898, Lot 484088), POU5F1 (Abcam, ab19857), SOX2 (Santa Cruz, sc-17320X), NANOG (R&D, AF1997) and FOXA2 (R&D, AF2400).

For live cell FACS isolation, cells were stained for 30 min on ice with the following antibodies directed toward extracellular surface proteins: CD326-PerCP-Cy5 (clone EBA1) (BD Biosciences), CD56-PE (clone NCAM16.2) (BD Biosciences), and CD184-PE-Cy5 (clone 12G5) (BD Biosciences).

Immunostaining was done with the following primary antibodies: FOXA2 (R&D, AF2400), GATA2 (Santa Cruz, sc-16044) SOX17 (R&D, AF1924), PAX6 (Covance, PRB-278P) and HNF4α (Abcam, ab41989). Cells were fixed in 4% Formaldehyde, incubated in primary antibody overnight at 4°C, and then incubated in secondary antibody for 1 hr at room temperature. DNA was detected using Hoechst 33342 trihydrochloride trihydrate (Invitrogen).

FACS Analysis

FACS was done on a BD FACSAria II using linear FSC and SSC scaling, followed by height and width-based doublet discrimination. The viability of the populations was assessed by Propidium Iodide staining, with the positively stained populations being excluded from the sorting gates. Compensation was calculated using FACS Diva autocompensation algorithms, and supplemented by manual compensation to correct for autofluorescence.

Genomic DNA Isolation

Flash-frozen human tissues or cell pellets were lysed at 55°C overnight in 300–600 μl lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 10 mM NaCl and 0.5% wt/vol SDS) supplemented with 50 ng/μl DNase-free RNase (Roche) and 1 μg/μl proteinase K (NEB). After extraction with an equal volume of phenol:chlorofom:isopropanol alcohol (25:24:1; Invitrogen) and addition of 0.5 μl (20 μg/μl) glycogen (Roche) and 1/20 vol 5 M NaCl, DNA was precipitated with 2.5 vol ethanol, spun down (30 min/16,000 g) at 4°C and washed with 70% ethanol. DNA was re-suspended in 30–100 μl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and quantified using a Qubit fluorometer and a dsDNA BR Assay Kit (Life Technologies).

WGBS Library Construction

Genomic DNA (1–5 μg) was fragmented to 100–500 bp using a Covaris S2 sonicator 9 times for 60 s at duty cycle 20%, intensity 5 and 200 cycles per burst. DNA fragments were cleaned up using a QIAGEN PCR purification kit. End-repair reactions (100 μl) contained 1x T4 DNA ligase buffer (NEB), ATP, 0.4 mM dNTPs, 15 units T4 DNA polymerase, 5 unites Klenow DNA polymerase, 50 units T4 polynucleotide kinase (all NEB) and were incubated for 30 min. at 19°C and 15 min. For some libraries we used a dCTP-free dNTP mix instead of all four dNTPs during for the end-repair to avoid artificially unmethylated sites. Adenylation was performed for 30 min. at 37°C in 50 μl 1x Klenow buffer containing 0.2 mM dATP and 15 units Klenow exo$^-$ (NEB). Adenylated DNA fragments and methylated paired-end adapters (purchased from ATDBio) were incubated overnight at 16°C in a 50 μl reaction containing 5,000

units concentrated T4 DNA ligase (NEB) and 3 μM of adapters. Each enzymatic reaction was terminated and cleaned-up by phenol/chloroform extraction and ethanol precipitation as described above.

To determine unmethylated cytosine conversion rates and methylated cytosine over-conversion rates by sodium bisulfite treatment, adaptor-ligated fully methylated and fully unmethylated internal control DNA fragments (Table S7), were spiked into WGBS library preparation at a molar ratio (spike-in to WGBS library) of 1:16,000 each. Adaptor-ligated DNA of 270-370 bp, corresponding to DNA insert sizes of 150-250 bp, was size-selected on a 2.5% Nusieve (3:1) agarose gel (Lonza). Two consecutive bisulfite conversions were performed with an EpiTect Bisulfite Kit (QIAGEN) following the protocol specified for DNA isolated from FFPE tissue samples. One of 40 μl bisulfite-converted DNA was used in each of four 10-μl reactions to determine the minimal PCR cycle number for library amplification. PCR reactions contained 0.5 U of *PfuTurboCx* Hotstart DNA polymerase (Agilent technologies), 1 μl of 10x PCR buffer, 250 μM dNTPs, 1.5 μM of Primer 1.0 and 2.0 (Illumina). The thermocycling profile was 2 min. at 95°C followed by 5-15 cycles of 30 s at 95°C, 30 s at 65°C, 1 min. at 72°C, and a final 7 min. extension at 72°C. Preparative library amplification using the empirically determined number of PCR cycles was performed in eight 25-μl aliquots, each containing 3 μl of bisulfite-converted DNA, 1.25 U of *PfuTurboCx* Hotstart DNA polymerase, 2.5 μl of 10x PCR buffer, 250 μM of dNTP, 1.5 μM of Primer 1.0 and 2.0. PCR products were pooled and purified twice using Agencourt *AMPure XP SPRI Beads* (Beckman Coulter) as per the manufacturer's instructions. The final library DNA was quantified using a Qubit fluorometer and a Quant-iT dsDNA HS Kit (Invitrogen). The insert size was checked on a 4%–20% non-denaturing polyacrylamide gel (Bio-Rad). Paired-end sequencing with 100 base reads was performed on an Illumina Hiseq 2000 followed the manufacturer's guidelines.

WGBS Data Processing and Analysis

WGBS raw sequencing reads were aligned using maq in bisulfite mode against human genome version hg19/GRCh37, discarding duplicate reads. DNA methylation calling was performed based on an extended custom software pipeline published previously for RRBS (Bock et al., 2011). To ensure comparability of region DNA methylation levels across all samples, only CpGs covered by $\geq$ 5x in 85% of the samples qualified for the computation of region DNA methylation levels. To assess the DNA methylation state of various genomic regions, we resorted to our previously published protocol estimating a genomic region's methylation state as the coverage weighted average across all CpGs within each region. Subsequently, we averaged a region's DNA methylation level over replicates. Differentially methylated regions (DMRs) were defined as exhibiting significantly ($p \leq 0.05$, Fisher's exact test) different DNA methylation levels of at least 0.1.

Many gene regulatory elements (GREs) are marked by spatially highly constrained reduced DNA methylation levels. It has recently been suggested that besides CpG islands, which are mostly unmethylated (UMR) a second class of GRE is marked by low to intermediate DNA methylation (IMR) (Stadler et al., 2011). We reasoned that these regions might be of particular regulatory importance in our system and might be missed by looking at histone modification enrichments alone. Therefore we adopted a similar Hidden Markov model approach as proposed in Stadler et al. (2011) to identify regions of reduced DNA methylation level. Briefly, we utilized a three state Hidden Markov Model operating on the methylation levels of each CpG in the human genome. Each state's emission probabilities for the DNA methylation levels were modeled by a normal distribution. The model was trained on all CpGs of chromosome 19 in the HUES64 data set using an adaption of the well known Baum Welch algorithm to incorporate the normal distribution (Press, 2007). After initial parameter estimation, we utilized the approach reported by Stadler et al. (2011) to determine the FDR for IMR regions and adapted the initial parameter estimates for the IMR and HMR states to finally 0.01(UMR), 28.8 (IMR), 81.6 (highly methylated, HMR), yielding an FDR of 2%. This parameter set was subsequently used to segment all WGBS data sets. Finally, we used the Viterbi algorithm to compute the most probable path through each chromosome separately and assigned the CpG states accordingly to either unmethylated, intermediate or highly methylated. Subsequently, we merged neighboring CpGs residing in the same state and being less than 200bp apart into unmethylated, intermediate, or highly methylated regions. Only regions harboring more than 3 CpGs were retained for subsequent analysis. The resulting region set is more likely to pick up DMRs due to the highly spatially constrained nature of the marked GRE (often 200-400bp) which easily gets masked by a coarse grained tiling based approach. The HMM inference framework was implemented as custom software in python (http://python.org/) and extended to incorporate other state distribution types. To determine differentially methylated regions between two samples, we followed our previously established protocol (Bock et al., 2011).

ChIP and ChIP-Seq Library Production

Cells collected by FACS were crosslinked in 1% formaldehyde for 15 min at room temperature, with constant agitation, followed by quenching with 125mM Glycine for 5 min at room temperature with constant agitation. Nuclei were isolated and chromatin was sheared using Branson sonifier until the majority of DNA was in the range of 200-700 base pairs. Chromatin was incubated with antibody overnight at 4°C, with constant agitation.

Co-immunoprecipitation of antibody-protein complexes was completed using Protein A or Protein G Dynabeads for 1 hr 4°C, with constant agitation. ChIPs were completed using previously reported methods (Mikkelsen et al., 2010). Sequencing library production details can be found in the Extended Experimental Procedures. Sequencing libraries were submitted for sequencing on the Illumina Hiseq 2000.

Immunoprecipitated DNA was end repaired using the End-It DNA End-Repair Kit (Epicenter), extended using a Klenow fragment (3′-5′ exo) (NEB), and ligated to sequencing adaptor oligos (Illumina). Each library was then PCR-amplified using PFU Ultra II Hotstart

Master Mix (Agilent), and a size range of 300-600 was selected for sequencing. We confirmed binding of OCT4, NANOG and SOX2 at the *NANOG* promoter using qPCR (Table S8).

ChIP-BS-Seq Library Construction

DNA was first subjected to end-repair in a 30-μl reaction containing 6 units T4 DNA polymerase, 2.5 units DNA Polymerase I (Large Klenow Fragment), 20 units T4 Polynucleotide Kinase (all New England Biolabs), dATP, dCTP, dGTP, and dTTP (0.125 mM each), and 1 × T4 Ligase buffer with ATP for 30 min at 20°C. DNA was then adenylated in a 20-μl reaction containing 10 units Klenow Fragment (3′→5′ exo-) (New England Biolabs), 0.5 mM dATP and 1 × NEB buffer 2 for 30 min at 37°C. DNA was then ligated to preannealed Illumina genomic DNA adapters containing 5-methylcytosine instead of cytosine (ATDBio) using T4 DNA ligase (New England Biolabs).

Adaptor-ligated DNA fragments were subsequently purified by phenol extraction and ethanol precipitation and size-selected on gel. 50 ng sheared and dephosphorylated *Escherichia coli* K12 genomic DNA was added to adaptor-ligated DNA as carrier during size-selection and bisulfite conversion. DNA was run on 2.5% Nusieve 3:1 Agarose (Lonza) gels. Lanes containing marker (50 bp ladder; New England Biolabs) were stained with SYBR Green (Invitrogen), and size regions to be excised were marked with tooth-picks and adaptor-ligated DNA fragments from 200–400 and 400–550 bp were excised. DNA was isolated from gel using the MinElute Gel Extraction kit (QIAGEN). The low and high libraries were kept separate in subsequent steps.

Adaptor-ligated and size-selected DNA was subjected to two subsequent 5-h bisulfite treatments using the EpiTect Bisulfite kit (QIAGEN) following the manufacturer's protocol for DNA isolated from FFPE tissue samples. PCR amplification was done with 1.25 units Pfu Turbo Cx Hotstart DNA Polymerase (Stratagene), primer LPX 1.1 and 2.1 (0.3 μM each), dNTPs (0.25 mM each), 1 × Turbo Cx buffer. Amplified libraries were purified with the MinElute PCR Purification kit (QIAGEN) and subsequently purified from gel essentially as described above; whole gels were stained with SYBR Green, and no carrier DNA was added. Final libraries were analyzed on analytical 4%–20% TBE Criterion precast gels (BioRad), and measured by Quant-iT dsDNA HS Assays (Invitrogen) (protocol adapted from Brinkman et al. (2012).

ChIP-Seq Data Processing and Analysis

ChIP-Seq data were aligned to the hg19/GRCh37 reference genome using bwa version 0.5.7 (Li and Durbin, 2009) with default parameter settings. Subsequently, reads were filtered for duplicates and extended by 200bp. Visualization of read count data was performed by converting raw bam files to .tdf files using IGV tools (Thorvaldsdóttir et al., 2013) and normalizing to 1 million reads.

In order to identify regions enriched for chromatin modifications we employed a two step approach, first identifying all regions enriched for any chromatin modification. Next, using this comparatively small region set, we determined the quantitative enrichment level as well as significance of enrichment using a Poisson background model based on the whole cell extract (WCE). Finally, we utilize conservative enrichment and significance cutoffs to binarize our enrichment signal in order to increase robustness and simplify subsequent analysis.

First, we segmented the genome into non-overlapping windows and classified each window into either enriched or no enriched. This analysis was conducted separately for two groups, 1. H3K27ac, H3K4me3 using 200bp windows and H3K27me3, H3K9me3, K3K4me1 using 400bp windows. To compute the enrichment statistics on the window level, we determined the number of unique insert size extended sequence tags whose midpoint was located within the window of interest for the ChIP-Seq track of interest as well as the WCE. Next, we used the poisson model proposed in Mikkelsen et al. (2010) to determine nominal p-value of enrichment and computed the enrichment over the WCE. Only windows enriched at a significance level below $p < 10^{-5}$ (in case replicates were used both had to fulfill this criterion) and an enrichment above background greater than 3 (in case replicates were used, the average enrichment had to be greater than 2.5) were retained. However, for most enrichment analysis we employed only the replicate with the strongest signal.

Next, enriched windows within a distance of 850bp were merged into larger regions. Regions smaller than 400bp (600bp for broad marks) after merging were discarded as due to noise and regions greater than 10kb were split. This procedure was carried out for three groups of histone ChIP-Seq tracks separately: H3K4me3 &H3K27ac, H3K4me1 and H3K27me3 & H3K9me3 across all 4 cell types. The resulting three lists of enriched regions were then merged in a hierarchical fashion: first regions identified based on H3K4me3&H3K27ac and H3K4me1, retaining all H3K4me3&H3K27ac regions but merging or splitting enriched H3K4me1 regions.

After completion of this initial processing step, regions were again filtered for minimal size discarding regions smaller than 400bp. Next, the same procedure was repeated for the new H3K4me3, H3K27ac, H3K4me1 region set and the H3K27me3, H3K9me3 region list. Finally, the resulting list was merged with the regions classified as UMRs and IMRs, adding only regions not overlapping with any region identified so far. This procedure gave rise to the region catalog used in subsequent analysis.

In the second processing step, comparative analysis of ChIP-Seq experiments and assignment of chromatin states was carried out, excluding regions enriched for H3K9me3 only. First, for each region in the region catalog the significance and enrichment over WCE was determined using Poisson statistics (Mikkelsen et al., 2010) applied to the duplicate filtered and insert size extended sequencing tag counts overlapping each identified region. Regions with tag counts deviating at a significance level of $p \leq 0.001$ from the WCE and exhibiting enrichment over WCE ≥ 3 were classified as enriched. We chose these moderately stringent thresholds in order also pick up chromatin state changes that occur only in a subset of the investigated cell population and therefore have lower signal. However, this comes at the expense of a higher false positive rate. Next, we compared the enrichment levels for all four cell

types (hESC, dEC, dME, dEN) for each epitope separately. To that end we used the Poisson model based approach proposed in Mikkelsen et al. (2010) and defined regions deviating by $\geq$2.5 fold at a significance level of $p \leq 0.01$ as being different. Next, we reconciled these differential enrichment calls with our enrichment over background classification. Since in our setting we were mostly concerned with incorrectly called differences between cell states (false positives) due to heterogeneity in the distinct populations and varying ChIP-Seq library complexity, we redefined regions that were classified as enriched in hESC and not enriched in one of the differentiated cell types but exhibiting no significant difference according to our differential analysis as being enriched in the differentiated cell type under study. This approach yields a lower false positive rate in terms of dynamics at the expense of a higher false negative rate. However, at this point it still remains to be determined what magnitudes of differences in chromatin modifications are actually meaningful. In this sense, our binary classification approach is rather conservative and relies on previously established observations. Subsequently, we classified each genomic region identified in this way into one of 11 epigenetic states based on the binary classification of enrichment levels for the various modifications. DNA methylation levels were not taken into account when histone modification based states were assigned. Only states devoid of significant enrichment for one of the histone modifications were classified based on DNA methylation levels. Genomic regions were associated with their nearest RefSeq gene using the R package ChIPpeakAnno (Zhu et al., 2010) and classified into promoter, intragenic, distal (<50kb from TSS and not promoter) and intergenic.

TF ChIP-Seq Analysis

For OCT4, SOX2, NANOG and FOXA2 aligned read files were processed with macs version 1.4 (Zhang et al., 2008) using the following parameters: -g 2.7e9–tsize = 36–pvalue = 1e-5–keep-dup = 1 and the HUES64 WCE as input control. All other parameters were left at their default setting. For our 25bp libraries, tsize was set to 25. FDR was calculated using macs built in function essentially comparing the original read count distribution with a randomly shuffled distribution. Following this initial peak calling, only peaks significant at an FDR of 0.05 and present in both replicates were retained. As a second replicate for our OCT4 ChIP-Seq experiment we took advantage of publically available OCT4 data (Kunarso et al., 2010).

ChIP BS-Seq Analysis

For the FOXA2 ChIP-bisulfite sequencing experiment, the bisulfite treated ChIP library was processed similarly to the WGBS processing described above and subsequently overlaid with the peak calling results from the FOXA2-ChIP-Seq library that was not bisulfite treated.

RNA-Seq Data Processing and Differential Expression Analysis

Strand specific libraries were constructed as described in the main text using a strand specific method (Levin et al., 2010). Reads were mapped to the human genome (hg19) using TopHat v2.0.6 (Trapnell et al., 2009) (http://tophat.cbcb.umd.edu) with the following options: ''—library-type fr-firststrand'' and ''—transcriptome-index'' with a TopHat transcript index built from RefSeq. Transcript expression was estimated with an improved version of Cuffdiff 2 (Trapnell et al., 2013) (http://cufflinks.cbcb.umd.edu). Cuffdiff was run with the following options: ''—min-reps-for-js-test 2 –dispersion-method per-condition'' against the UCSC iGenomes GTF file from Illumina (available at http://cufflinks.cbcb.umd.edu/igenomes.html). The workflow used to analyze the data is described in detail in Trapnell et al. (2012) (alternate protocol B).

To identify a gene or transcript as DE, Cuffdiff 2 tests the observed log-fold-change in its expression against the null hypothesis of no change (i.e., the true log-fold-change is zero). Because of measurement error, technical variability, and cross-replicate biological variability might result in an observed log-fold-change that is nonzero, Cuffdiff assesses significance using a model of variability in the log-fold-change under the null hypothesis. This model is described in detail in Trapnell et al. (2013). Briefly, Cuffdiff two constructs, for each condition, a table that predicts how much variance there is in the number of reads originating from a gene or transcript. The table is keyed by the average reads across replicates, so to look up the variance for a transcript using the table, Cuffdiff estimates the reads originating from that transcript, and then queries the table to retrieve the variance for that number of reads. Cuffdiff 2 then accounts for read mapping and assignment uncertainty by simulating probabilistic assignment of the reads mapping to a locus to the splice isoforms for that locus. At the end of the estimation procedure, Cuffdiff 2 obtains an estimate of the number of reads that originated from each gene and transcript, along with variances in those estimates. The read counts are reported along with FPKM values and their variances. Change in expression is reported as the log fold change in FPKM, and the FPKM variances allow the program to estimate the variance in the log-fold-change itself. Naturally, a gene that has highly variable expression will have a highly variable log-fold-change between two conditions.

The modifications made to Cuffdiff 2 improve sensitivity in calling differentially expressed (DE) genes and transcripts while maintaining a low false positive rate. They stem from the method used to calculate the variability in the log fold change in expression. In Trapnell et al., Cuffdiff 2 used the ''delta method'' to estimate the variance of the log fold change estimate for a gene or transcript. This method yields a simple equation that takes as input the mean and variance of the transcript's expression in two conditions and produces a variance for the log fold change. However, the equation contains no explicit accounting for the number of replicates used to produce those estimates – they are assumed to be perfectly accurate.

The improved version of Cuffdiff 2 more accurately estimates the variance in the log-fold-change using simulated draws from the model of variance in expression for each of the two conditions. Imagine an experiment that has *n* replicates in condition A and *m*

replicates in condition B. To estimate the distribution of the log-fold-change in expression for a gene G under the null hypothesis, Cuffdiff first draws *n* times from the distribution of expression of G according to the algorithm's model of expression. Cuffdiff then takes the average of the *n* draws to obtain an expression "measurement." Then, Cuffdiff draws *m* from the *same* distribution and again takes their average. Cuffdiff then takes the log ratio of these averages, places this value in a list, and then repeats the procedure until there are thousands of such log-fold-change samples in the list. The software then makes a similar list, this time using the expression model for condition B – the null hypothesis assumes both sets of replicates originate from the same condition, but we do not know whether A or B is the better representative of that condition, so we must draw samples from both and combine them. To calculate a *p*-value of observing the real log-fold-change under this null model, we simply sort all the samples and count how many of them are more extreme than the log fold change we actually saw in the real data. This number divided by the total number of draws is our estimate for the *p*-value.

Cuffdiff 2 reports not only genes and transcripts that are significantly differentially expressed between conditions, but also *groups* of transcripts (i.e., the isoforms of a gene) that show significant changes in expression relative to one another. The test for this is similar to what is described in Trapnell et al., but comparably modified along the lines described above for single genes or transcripts. Draws of expression are made for each transcript in a group according to the number of replicates in the experiment. These are averaged, and the shift in relative transcript abundance for the draw is made using the Jensen-Shannon metric. These draws are added to a list and used to calculate *p*-values for significance of observed shifts in relative abundance under the null hypothesis.

Clustering of gene expression profiles was achieved with the csDendro() function from CummeRbund (http://compbio.mit.edu/cummeRbund/). This function first transforms the FPKMs of all genes in each sample by adding 1 and then takes the logarithm. Next, it converts each genes transformed expression into a fraction of the total transformed expression. The distances between these transformed expression profiles are then measured by the Jensen-Shannon metric. The distances are then used to build a dendrogram via complete linkage hierarchical clustering using the R function hclust().

Motif Analysis

Predefined sets of genomic regions were scanned for occurrences of motifs contained in the Transfac professional database (2009) using the FIMO program from the MEME suite (Grant et al., 2011). Only motifs with at least one known associated human transcription factor and detected at a significance level of $p \leq 10^{-5}$ were used for further analysis. Next, the total number of occurrences was calculated for each motif. To correct for sequence composition, we trained a Hidden Markov Model on each set of input sequence sets and generated 10 sets of number and size matched region sets using the inferred probabilities as controls. Subsequently, these sequence sets were also subjected to the same motif identification procedure and motif enrichment results were averaged over the 10 control runs. We defined the final motif enrichment score as the fraction of total motif occurrences in the region set of interest and the total number of motif occurrences in the averaged control region set. To determine differentially enriched motifs between region sets from different hESC-derived cell types, we calculated the fraction of motif scores between the two conditions, retaining only motifs with a differential enrichment ≥ 1.2.

For the H3K27ac motif analysis, we computed overall motif enrichment scores for each region class separately as described above. Next, we correlated the motif enrichment scores only focusing on those motifs with scores ≥ 1.2. To that end we multiplied the motif enrichment score for the cell type of interest with the $\log_2$ fold change of the associated transcription factor in that cell type, giving rise to a new combined motif score. If multiple TFs mapped to one motif, we took the average motif score. For each cell type we rank ordered the motifs according to their enrichment scores and report the top 20 motifs with their raw motif score in Figure 5D.

For the H3K4me1 analysis, we wanted to focus on all potential TFBS gaining H3K4me1 and not only those that also become expressed as in the H3K27ac analysis. First, we again determined the motif enrichment scores over background. To focus on motifs differentially enriched between the different cell types, we subtracted the mean motif enrichment across hESd cell types for each motif separately from the enrichment level and rank ordered the motifs. For each cell type, we then report the top 20 enriched motifs in Figure 6G.

SUPPLEMENTAL REFERENCES

Chen, A.E., Egli, D., Niakan, K., Deng, J., Akutsu, H., Yamaki, M., Cowan, C., Fitz-Gerald, C., Zhang, K., Melton, D.A., and Eggan, K. (2009). Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. Cell Stem Cell *4*, 103–106.

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

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods *7*, 709–715.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S., and Rosen, E.D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. Cell *143*, 156–169.

Press, W.H. (2007). Numerical Recipes: the Art of Scientific Computing, Third Edition (Cambridge, UK: Cambridge University Press).

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

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

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

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

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

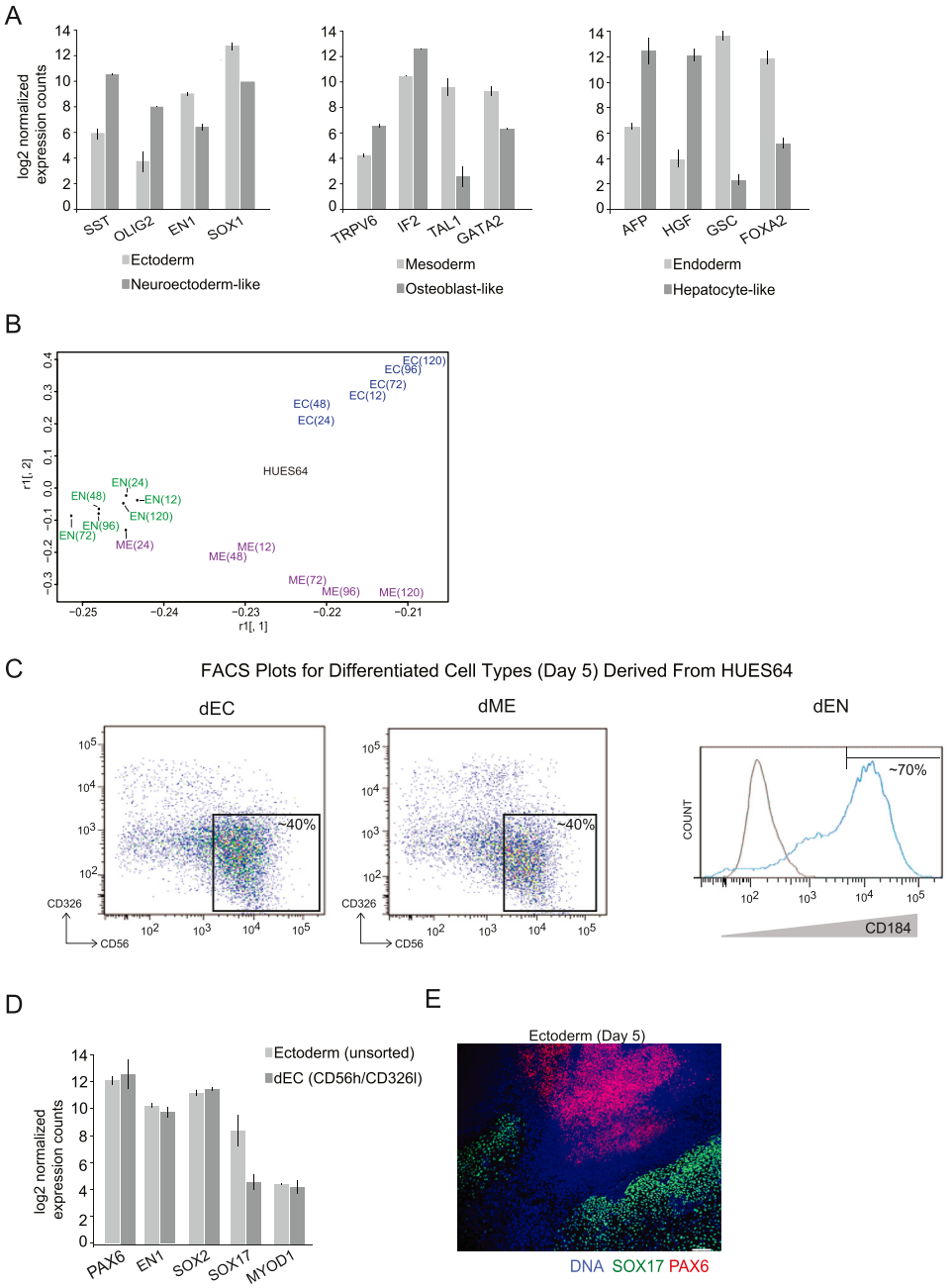

Figure S1. Additional Characterization of the Differentiated Populations, Related to Figure 1

(A) Median Nanostring expression values (log_2) of populations derived from dEC, dME and dEN.

(B) Multidimensional scaling of populations included in differentiation time course.

(C) Representative FACS plots used to isolate differentiated populations. Square boxes (left and middle panels) and line (in right panel) indicate population collected for further analysis. Approximate percent of population collected is given.

(D) Average Nanostring expression values (log_2) of unsorted ectoderm versus CD56high/CD326low sorted dEC cells.

(E) Immunofluorescent staining of SOX17 and PAX6 in day 5 ectoderm population (40x, scale bar equals 200um).

A

State	hESCs	dEC	dME	dEN
H3K4me3 & H3K27me3	11,828	9,174	2,737	4,776
H3K4me3 & H3K27ac	11,583	6,815	9,182	9,959
H3K4me3	15,443	11,262	14,408	36,180
H3K27me3 & H3K4me1	3,064	5,481	674	1,640
H3K27ac	20,345	12,330	29,818	23,186
H3K4me1	40,035	54,836	48,159	29,402
H3K27me3	15,374	16,921	20,405	20,574
H3K9me3	61,426	56,272	60,980	58,431
IMR	10,272	12,667	16,930	13,652
HMR	63,499	63,724	51,111	59,339
None	15,030	18,312	13,186	10,543

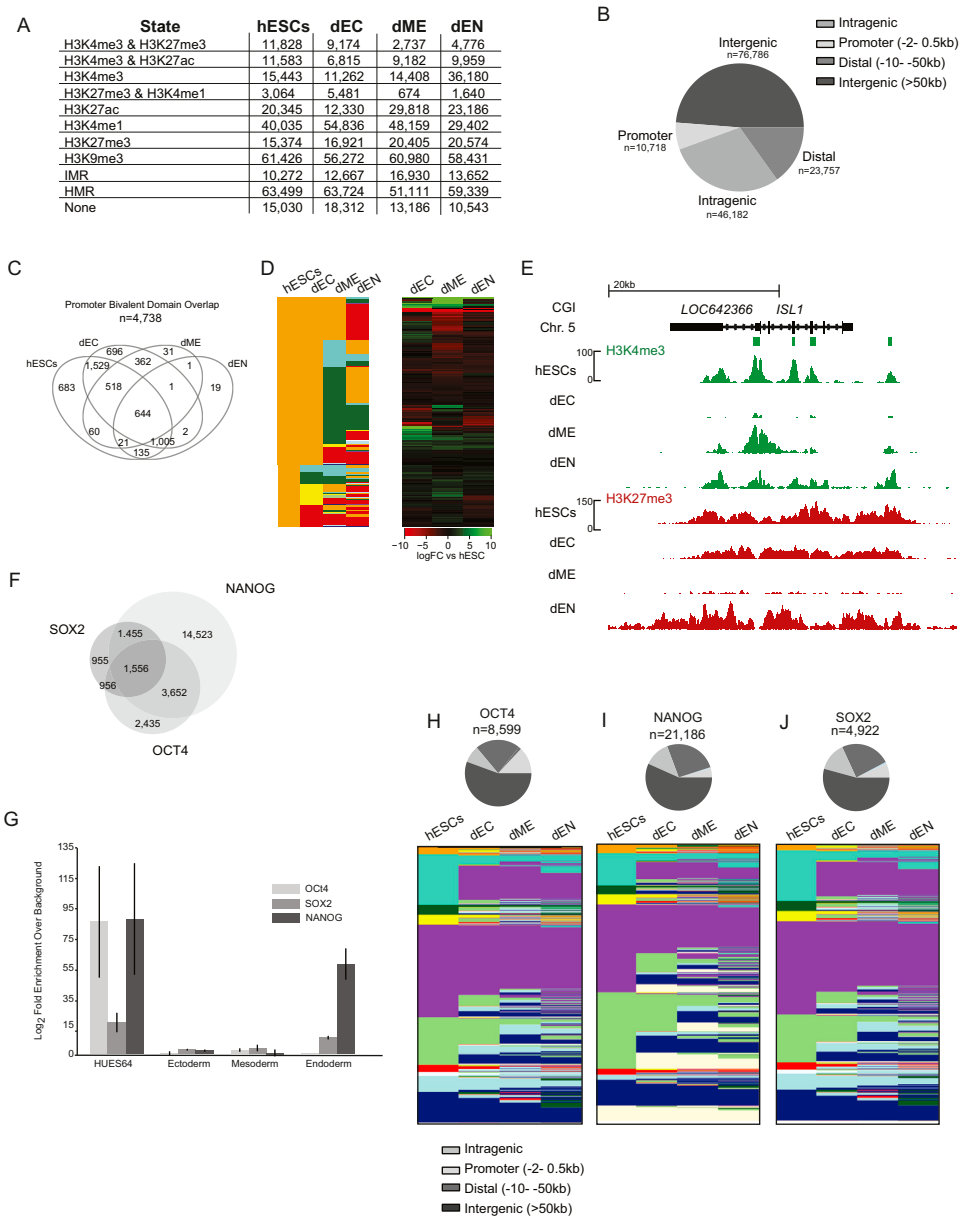

Figure S2. Overall Chromatin Dynamics and TFS Binding, Including OCT4, SOX2, and NANOG, Related to Figure 2

(A) Total number of regions found in each state, in each population.

(B) Genomic features associated with all regions that change their epigenetic state in at least one cell type.

(C) Venn diagram showing the overlap of identified proximal bivalent domains.

(D) Left: Chromatin state map for all TFs that are bivalent in hESCs and change their epigenetic state in at least one cell type (n = 400). Right: Hierarchical clustering ordered heatmap of TF expression (logFC relative to hESCs).

(E) Normalized ChIP-seq tracks of H3K4me3 and H3K27me3 at the *ISL1* locus (chr5:50,661,163-50,703,879) indicating H3K27me3 is selectively maintained at high levels in dEC but not dME, in contrast to dEN, where H3K4me3 increases while H3K27me3 is lost, promoting active transcription. Read counts on *y* axis are normalized to 10 million reads and CpG islands (CGI) are indicated in green.

(F) Venn diagram of the overlap between OCT4, NANOG and SOX2 binding sites identified in hESCs (total overlap = 1,556).

(G) Fold enrichment of OCT4, NANOG and SOX2 binding at the NANOG locus in hESCs, and each differentiated population on day 5 of differentiation.

(H) Genomic features of OCT4 binding sites (top) and the associated epigenetic states (bottom) (n = 8,599).

(I) Genomic features of NANOG binding sites (top) and the associated epigenetic states (bottom) (n = 21,186).

(J) Genomic features of SOX2 binding sites (top) and the associated epigenetic states (bottom) (n = 4,922).

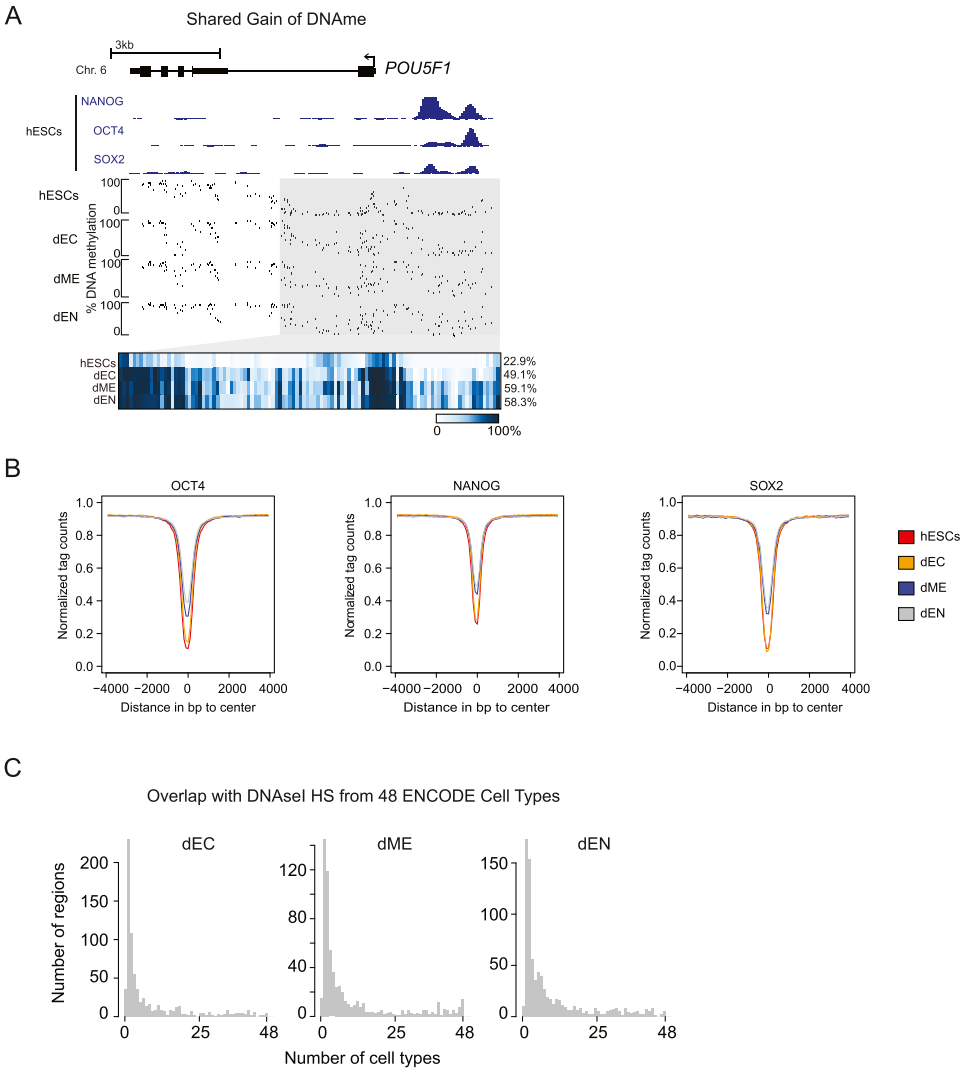

Figure S3. DNA Methylation Dynamics at *POU5F1* and Regions Associated with TF Binding, Related to Figure 3

(A) Gain of DNAme at the *POU5F1* locus (chr6:31,135,410-31,141,237). NANOG, OCT4, SOX2 ChIP-seq tracks (hESCs only) and DNAme levels in hESCs and differentiated cell types. Individual CpG methylation values across the locus are displayed using the IGV. The heatmap below shows the DNAme values of individual CpGs within the gray region. The average DNAme value for the entire highlighted region is shown on the right in red. The TSS is indicated by the arrow. Gain of DNAme is seen at the distal enhancer, as well as over the TSS, in all three differentiated cell types.

(B) Composite plots of DNAme levels in hESC and differentiated populations across SOX2/OCT4/NANOG binding sites in hESCs. Average CpG methylation levels were computed for 100bp tiles across an 8kb region centered at the middle of each transcription factor binding site.

(C) Frequency distribution of overlapping DMRs gaining DNAme in the differentiated populations with DNase I hypersensitive sites across 48 ENCODE cell types (Thurman et al., 2012).

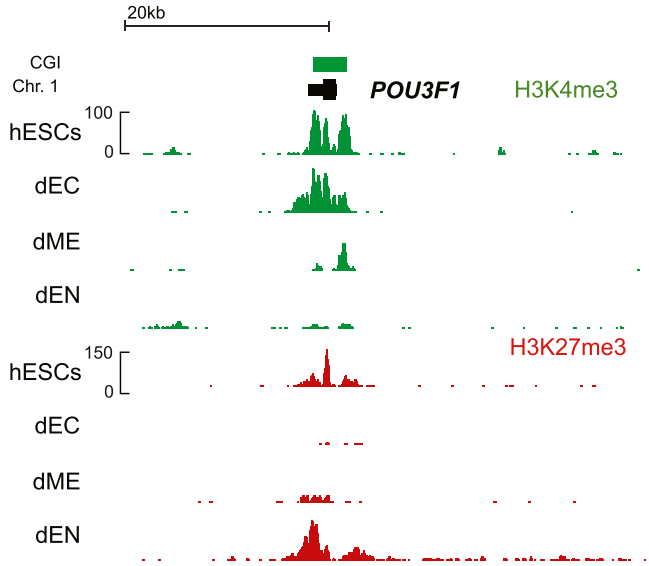

Figure S4. H3K4me3 and H3K27me3 Enrichment at *POU3F1*, Related to Figure 4

H3K4me3 and H3K27me3 tracks at the *POU3F1* locus (chr1:38,493,152-38,532,618) show lineage-specific resolution of the bivalent domain.

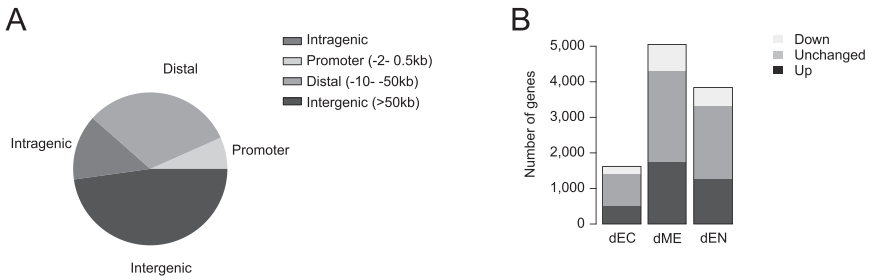

Figure S5. Genomic Features and Expression Changes at Regions that Gain H3K27ac, Related to Figure 5

(A) Genomic features associated with gain of H3K27ac during differentiation. See key on the right.

(B) Classification of gene expression associated with regions gaining H3K27ac in each germ layer into either upregulated (FDR < 0.05), downregulated (FDR < 0.05), or unchanged.

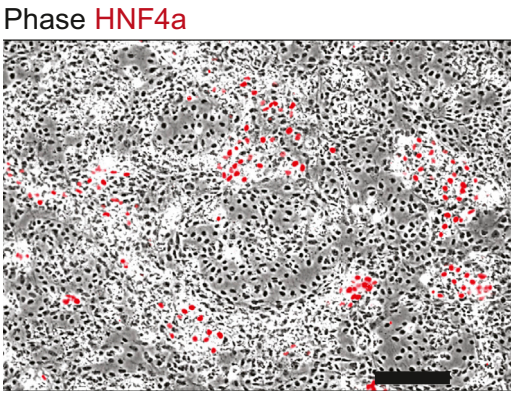

Figure S6. HUES64-Derived Hepatoblast-like Population, Related to Figure 6
Immunofluorescent HNF4α stain of hepatoblast-like population after 10 days of differentiation (10x, scale bar = 100 μm).

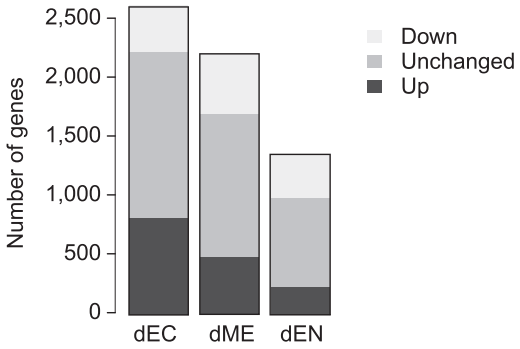

Figure S7. Gene Expression Changes at Genes that Gain H3K27me3, Related to Figure 7

Classification of gene expression associated with regions gaining H3K27me3 in each germ layer into either upregulated (FDR < 0.05), downregulated (FDR < 0.05), or unchanged.