Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1016/j.cell.2013.04.037">http://dx.doi.org/10.1016/j.cell.2013.04.037</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Elsevier B.V.</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Dec 21 14:14:48 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/87632">http://hdl.handle.net/1721.1/87632</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Resource

Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells

Casey A. Gifford,1,2,3 Michael J. Ziller,1,2,3 Hongcang Gu,1 Cole Trapnell,1,3 Julie Donaghey,1,2,3 Alexander Tsankov,1,2,3 Alex K. Shalek,4 David R. Kelley,1,3 Alexander A. Shishkin,1 Robbyn Issner,1 Xiaolan Zhang,1 Michael Coyne,1 Jennifer L. Fostel,1 Laurie Holmes,1 Jim Meldrim,1 Mitchell Guttmann,1 Charles Epstein,1 Hongkun Park,4 Oliver Kohlbacher,5 John Rinn,1,3,6 Andreas Gnirke,1 Eric S. Lander,1,7 Bradley E. Bernstein,1,8 and Alexander Meissner1,2,3,*

1Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
2Harvard Stem Cell Institute, Cambridge, MA 02138, USA
3Department of Stem Cell and Regenerative Biology
4Department of Chemistry
5Harvard University, Cambridge, MA 02138, USA
6Applied Bioinformatics, Center for Bioinformatics and Quantitative Biology Center, University of Tübingen, 72076 Tübingen, Germany
7Beth Israel Deaconess Hospital, Boston, MA 02115, USA
8Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
9Howard Hughes Medical Institute and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
*These authors contributed equally to this work
*Correspondence: alexander_meissner@harvard.edu
http://dx.doi.org/10.1016/j.cell.2013.04.037

SUMMARY

Differentiation of human embryonic stem cells (hESCs) provides a unique opportunity to study the regulatory mechanisms that facilitate cellular transitions in a human context. To that end, we performed comprehensive transcriptional and epigenetic profiling of populations derived through directed differentiation of hESCs representing each of the three embryonic germ layers. Integration of whole-genome bisulfite sequencing, chromatin immunoprecipitation sequencing, and RNA sequencing reveals unique events associated with specification toward each lineage. Lineage-specific dynamic alterations in DNA methylation and H3K4me1 are evident at putative distal regulatory elements that are frequently bound by pluripotency factors in the undifferentiated hESCs. In addition, we identified germ-layer-specific H3K27me3 enrichment at sites exhibiting high DNA methylation in the undifferentiated state. A better understanding of these initial specification events will facilitate identification of deficiencies in current approaches, leading to more faithful differentiation strategies as well as providing insights into the rewiring of human regulatory programs during cellular transitions.

INTRODUCTION

Coordinated changes to the epigenome are essential for lineage specification and maintenance of cellular identity. DNA methylation (DNAme) and certain histone modifications critically contribute to epigenetic maintenance of chromatin structures and gene expression programs (Zhou et al., 2011; Smith and Meissner, 2013). Genetic deletion of histone methyltransferases and the catalytically active DNA methyltransferases are embryonic or postnatal lethal (Li, 2002), providing evidence for their essential role in proper execution of developmental programs. Several groups have reported genome-wide maps of chromatin and DNA methylation in pluripotent and differentiated cell types. From these efforts, a global picture of the architecture and regulatory dynamics is beginning to emerge. For example, active promoters generally contain modifications such as H3K4me3 and H3K27ac, whereas active enhancers are generally enriched for H3K4me1 and H3K27ac (Heintzman et al., 2009; Creyghton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011). Repressed loci exhibit enrichment for H3K27me3, H3K9me2/3, DNAme, or a combination of the latter two modifications. The enrichment of repressive histone modifications, such as H3K27me3, which is initiated at CpG islands (CGI), is considered a facultative state of repression, whereas DNAme is generally considered a more stable form of epigenetic silencing (Smith and Meissner, 2013).

Recent studies have reported dynamics that suggest epigenetic priming such as the appearance of euchromatic histone modifications prior to gene activation during in vitro T cell differentiation (Zhang et al., 2012) and cardiac differentiation (Wamstad et al., 2012). These results are reminiscent of changes that occur during the early stages of reprogramming toward the induced pluripotent state (Koche et al., 2011) and highlight possible similarities between differentiation and dedifferentiation. In parallel to these advances, whole-genome bisulfite sequencing (WGBS) has been used to map DNAme genome
Figure 1. Generation and Characterization of hESCs and hESC-Derived Cell Types

(A) Left: Low (4x) and high (40x) magnification overlaid immunofluorescent images of the undifferentiated human embryonic stem cell (hESC) line HUES64 stained with OCT4 (POU5F1) and NANOG antibodies. Right: Established directed (two-dimensional) differentiation conditions were used to generate representative populations of the three embryonic germ layers: hESC-derived ectoderm, hESC-derived mesoderm, and hESC-derived endoderm. Cells were fixed and stained after 5 days of differentiation with the indicated antibodies. Representative overlaid images at low (10x) and high (40x) magnification are shown. DNA was stained with Hoechst 33342 in all images. Scale bars, 200 μm (4x), 100 μm (10x), and 30 μm (40x).

(B) NanoString nCounter expression data (Z score log2 expression value of two biological replicates) for a time course of in vitro differentiation using the conditions shown in (A). 541 genes were profiled, and 268 changing by more than 0.5 are displayed. Selected lineage-specific genes are shown on the left for each category that was identified based on hierarchical clustering (see Table S1 for all). The average log2 expression value of two biological replicates is displayed. Error bars represent 1 SD.

(legend continued on next page)
The male hESC line HUES64, a National Institutes of Health (NIH)-et al., 2008) (see Extended Experimental Procedures). We chose of each embryonic germ layer, namely ectoderm (Lee et al., 2008). To better understand the molecular dynamics involved in hESC differentiation, we produced populations representative during human germ layer specification.

Directed Differentiation of hESCs
High-Resolution Transcriptional Measurements during

RESULTS

High-Resolution Transcriptional Measurements during Directed Differentiation of hESCs

To better understand the molecular dynamics involved in hESC differentiation, we produced populations representative of each embryonic germ layer, namely ectoderm (Lee et al., 2010), mesoderm (Evseenko et al., 2010), and endoderm (Hay et al., 2008) (see Extended Experimental Procedures). We chose the male hESC line HUES64, a National Institutes of Health (NIH)-approved line that readily differentiates into each of the three germ layers. These hESCs can be differentiated into a neuroectoderm-like progenitor population positive for SOX2 and PAX6 by inhibition of TGFβ, Wingless/integrase1 (WNT), and bone morphogenetic protein (BMP) signaling (Figure 1A, top). Alternatively, canonical mesoderm markers, such as GATA2 (Figure 1A, middle), can be induced using ACTIVIN A, BMP4, VEGF, and FGF2 treatment. Lastly, differentiation toward a definitive endoderm-like fate, positive for markers such as SOX17 and FOXA2 (Figure 1A, bottom), is induced using ACTIVIN A and WNT3A.

We began by measuring the expression of 541 selected genes, including many developmental transcription factors and lineage markers (Bock et al., 2011), at 24 hr intervals during differentiation toward each respective germ layer. We found that 268 of these genes exhibit expression changes (Z score log2 expression) during the first 5 days of differentiation (Figure 1B). Genes such as EOMES, T, FOXA2, and GSC are upregulated at 24 hr of mesoderm and endoderm induction, but not ectoderm differentiation (Figures 1B and 1C and Table S1 available online). GSC expression decreases within 48 hr of differentiation in the mesoderm-like population, whereas the expression level is maintained in the endoderm population (Figures 1B and 1C). EOMES and FOXA2 expression is also maintained in the endoderm population accompanied by upregulation of GATA6, SOX17, and HHEX (Figure 1B). After transient upregulation of mesendodermal markers, activation of mesodermal markers such as GATA2, HAND2, SOX9, and TAL1 is detected specifically in the mesoderm conditions (Figures 1B and 1C and Table S1). None of these markers are detected during early ectoderm differentiation, which instead upregulates neural markers such as PAX6, SOX10, and EN1 (Figures 1B and 1C and Table S1).

We found that POU5F1 (OCT4), NANOG, and, to some extent, SOX2 expression is maintained in our endoderm population (Figures 1B and 1D and Table S1). This is consistent with prior studies indicating that OCT4 and NANOG expression is detected during the course of early endoderm differentiation and supports NANOG’s suggested role in endoderm specification (Teo et al., 2011). SOX2 expression is downregulated in mesoderm and—to a lesser degree—in endoderm but is maintained at high levels in the ectoderm population (log2 expression 10.9) (Figure 1D and Table S1), whereas ZFP2 (REX1) is similarly downregulated in all three lineages (Figure 1B and Table S1). We confirmed that these populations indeed represent a precursor stage for each respective lineage by inducing them to differentiate further, which resulted in upregulation of genes such as OLG2 and
Figure 2. Epigenetic Remodeling Is Lineage Specific during Directed Differentiation

(A) WGBS (% methylation), ChIP-seq (read count normalized to 10 million reads), and RNA-seq (FPKM, read count normalized) for the undifferentiated hESC line HUES64 at three loci: *NANOG* (chr12:7,935,038-7,957,818), *GSC* (chr14:95,230,449-95,250,241), and *H19* (chr11:2,015,282-2,027,359). CGI are indicated in green.

(B) Size distribution of genomic regions enriched for at least one of our six histone modifications in at least one cell type (hESC, dEC, dME, and dEN) and/or classified as UMR or IMR in at least one cell type (n = 297,653).

(C) Definition of epigenetic states used in this study and the genomic space occupied by these in the four cell types under study.

(D) Median CpG content of the genomic regions in distinct epigenetic states defined in (C).

(E) Median expression level of epigenetic states used in this study (C) based on assignment of each region to the nearest RefSeq gene. Median was computed over the states in all four cell types and the corresponding expression profile.

(legend continued on next page)
SST in the ectoderm (Chambers et al., 2012), TRPV6 in the mesoderm (Evseeenko et al., 2010), and AFP and HGF in the later endoderm populations (Figure S1A) (DeLaForest et al., 2011). Lastly, multidimensional scaling confirmed that, at 24 hr, the mesoderm population is very similar to the endoderm, whereas the ectoderm population has already moved in an alternative direction (Figure S1B). These high-temporal-resolution gene expression signatures suggest that expression programs associated with the three unique cell populations, representing early stages of each germ layer, are established within a similar time frame of hESC differentiation.

Global Transcriptional Dynamics between hESCs and hESC-Derived Cell Types

Based on these results, we selected day 5 as the optimal time point to capture early regulatory events in well-differentiated populations representing all three germ layers. To reduce heterogeneity, we used FACS to enrich populations based on previously reported surface markers (Figure S1C); populations isolated by FACS are referred to as dEC for the ectoderm, dME for the mesoderm, and dEN for the endoderm. Expression analysis of the sorted populations confirms further enrichment for the desired populations (Figure 1E and Table S1).

We next expanded on our selected gene signature profiles by performing strand-specific RNA-seq on poly-A fractions from each day-5-differentiated FACS-isolated population and undifferentiated HUES64 (Table S2). Hierarchical clustering based on the global expression profiles of each cell type reveals that the dME population is the most distantly related cell type and that dEN and dEC are more similar to each other than to dME or hESCs (Figure 1F). This was unexpected, given that the dME and dEN populations are putatively derived through a common mesendoderm precursor stage (Figures 1B and 1C and Table S1), whereas the dEC does not upregulate markers associated with this stage (EOMES, T, and GSC; Table S1). Overall, 14,196 RefSeq-defined coding and noncoding transcripts (~38% of defined transcripts) are expressed (FPKM > 1) in at least one of the populations, with 11,579 (81.6% of the total number of transcripts detected within our cell types) being expressed in all three populations. Examining the overlap of genes expressed (FPKM > 1) in each population reveals that the dME population exhibits expression of the largest number of unique genes (n = 448, Figure 1G), such as RUNX1 (FPKM: 3.4) and HAND2 (FPKM: 17.8). Examining genes unique to pairs of the differentiated cell types also reveals that dEC and dME have the least in common (n = 37; Figure 1G), whereas dEC and dEN have the most number of transcripts in common (n = 171; Figure 1G), which is consistent with our clustering analysis. Genes such as PAX6 (dEC FPKM: 25.9; dEN FPKM: 5.6) and NKX6.1 (dEC FPKM: 2.3; dEN FPKM: 3.3), which are each required for both brain (Ericson et al., 1997) and pancreas development (Sander et al., 1997), are expressed in both the dEC and dEN. Canonical markers of embryonic development such as FOXA2 (FPKM: 12.7) in the dEN and EN1 (FPKM: 5.8) in the dEC are restricted to their expected germ layers at our early stages (Table S2).

Notably, we also identified 1,296 splicing events (FDR = 5%) as well as alternative promoter usage within our populations (Table S3) (Trapnell et al., 2013). For example, we detected expression of multiple isoforms of DNMT3B (p = 5 x 10^-6). Expression of DNMT3B isoform 1 (NM_006892) was restricted to the undifferentiated hESCs (FPKM: 214.3), whereas the differentiated cell types predominantly express an alternative isoform, DNMT3B isoform 3 (NM_175849) (dEC FPKM: 33.9; dME FPKM: 14.2; dEN FPKM: 20.0) (Figure 1H). The presence of this isoform, as well as others, has previously been reported in more advanced stages of embryonic development as well as in normal adult (Robertson et al., 1999) and cancerous tissues (Ostler et al., 2007). Our results suggest that this switch coincides with the exit from the pluripotent state, regardless of the specified lineage. We also identified expression of three PITX2 isoforms, with differential splicing leading to different isoform expression between the dEN and dME (Table S3). In the chick, PITX2 is essential for heart looping, and each isoform is responsible for executing distinct functions (Yu et al., 2001). Taken together, this suggests that both transcript levels and isoform expression contribute to cellular identity.

Generation of Comprehensive Reference Epigenome Maps

To gain a more complete picture of the underlying molecular dynamics and to investigate the regulatory events during the specification of the three germ layers, we collected ~12 million cells of the respective dE, dME, and dEN populations, as well as HUES64. All samples were subjected to ChIP-seq (H3K4me1, H3K4me3, H3K27me3, H3K27ac, H3K36me3, and H3K9me3) and WGBS (Figure 2A), producing a total of 28 data sets with more than 12 billion aligned reads (data are publicly available through the NIH Roadmap Epigenomics Project data repositories: http://www.roadmapepigenomics.org/; Table S4).

Integrated Analysis of Epigenetic State Transitions

After completing our basic quality control (see Extended Experimental Procedures and Table S4), we focused our analysis on previously identified informative chromatin states associated with various types of regulatory elements (Ernst et al., 2011; (F) Epigenetic state map of regions enriched for one of four histone modifications in at least one cell type or classified as UMR/IMR in at least one cell type and changing its epigenetic state upon differentiation in at least one cell type (n = 157,433). State definitions are listed in (C). (G) Regions bound by OCT4, SOX2, and NANOG, as determined by ChIP-seq and organized using the chromatin states in (F). (H) Enrichment of OCT4, SOX2, and NANOG within various classes of dynamic genomic regions changing upon differentiation of hESC, computed relative to all regions exhibiting the particular epigenetic state change in other cell types. Epigenetic dynamics are categorized into three major classes: repression (loss of H3K4me3 or H3K4me1 and acquisition of H3K27me3 or DNAme), maintenance of open chromatin marks (H3K4me3, H3K4me1, and H3K27ac), and activation of previously repressed states. See also Figure S2.
WGBS data into three DNAme states: highly methylated regions (HMRs: > 60%), intermediately methylated regions (IMRs: 11%–60%), and unmethylated regions (UMRs: 0%–10%). The latter two differ from the highly methylated background of the genome and likely indicate functional importance as previously suggested (Stadler et al., 2011). We next assigned each genomic region to one of the resulting states (Figure S2A, see Extended Experimental Procedures for details).

Across the four cell populations, we identified 297,653 genomic regions spanning 400 base pairs (bp) to 10 kilobases (kb) (Figure 2B) and covering a total of 512,784,035 bp (21.4% of the human genome) (Figure 2C) that were enriched for at least one of the eight chromatin states and/or classified as an IMR or UMR in at least one of the cell types. Of the identified epigenetic states, H3K27me3-enriched regions and HMRs covered the most base pairs (Figure 2C), and the combination of H3K4me3 and H3K27me3 exhibits the highest CpG content (Figure 2D), which was expected given the close association between H3K27me3 and high CpG density (Ku et al., 2008). The majority of epigenetically dynamic regions are not located near promoters (6.8% +2 kb to −500 bp of the TSS promoters; 48.8% > 50 kb upstream of TSS intergenic; 15.1% > 500 bp downstream of TSS intragenic/gene body) (Figure S2B). As expected, regions of open chromatin exhibit the highest median expression value (Figure 2E). However, overall, we find that many (62%–67%) epigenetic remodeling events are not directly linked to transcriptional changes based on the expression of the nearest gene.

The loss of H3K4 methylation (me1 and me3) is commonly associated with a transition to high DNAme (Figure 2F), which is most prominent in the dEN population and genes involved in neural development (i.e., neural tube development, $q = 9.6 \times 10^{-12}$). We identified 4,639 proximal bivalent domains in hESCs and observed that 3,951 (85.1%) of these domains resolve their bivalent state in at least one hESC-derived cell type (Figures 2F and S2C). When we specifically investigated the promoters of TF-encoding genes, we found that 463 of these promoters are in a bivalent state in hESCs, and 400 of them change in at least one differentiated cell type (Figure S2D). The majority transitions to H3K4me3 only or H3K27me3 only in a lineage-specific manner, as shown for ISL1 (Figure S2E). In dME, H3K4me3 is gained at the ISL1 locus while H3K27me3 is lost, leading to expression (FPKM: 14.3). The lineage-specific dynamics in this region are interesting given that this gene has known roles in all three germ layers, although at later time points (Piffet et al., 1996; Ahlgren et al., 1997; Cai et al., 2003). Notably, in contrast to the limited overall association between many epigenomic dynamics and changes in expression, we found that a large proportion of these bivalent TFs (27%) change their expression level during the differentiation (Figure S2D).

**Pluripotent TF Binding Is Linked to Chromatin Dynamics during Differentiation**

To further explore potential regulators of chromatin dynamics during the exit from pluripotency, we performed ChiP-seq for OCT4, NANOG, and SOX2 in HUES64 (Figures S2F and S2G and Table S8). We found that regions bound by all three factors (n = 1,556), by SOX2 only (n = 955), or by NANOG only (n = 14,523) are frequently associated with inter- and intragenic regions (Figures S2H–S2J, top). In contrast, regions bound by OCT4 only (n = 8,599) are more frequently associated with promoter regions (Figure S2H). Examination of regions bound by OCT4, NANOG, and SOX2 in hESCs showed that H3K4me1 regions enriched for OCT4 binding sites frequently become HMRs in all three differentiated cell types, whereas NANOG and SOX2 sites are more prone to change to an HMR state in dME (Figure 2F). In general, many regions associated with open chromatin that are bound by NANOG are more likely to retain this state in dEN compared to dME and dEC (Figure 2F). We also found that regions enriched for H3K27ac in hESCs that maintain this state in dEN or dEC are likely to be bound by SOX2 and NANOG. This is in agreement with the reported role of SOX2 during ectoderm development and differentiation (Wang et al., 2012) but also supports our observation that SOX2 expression is maintained in the dEN. Motif enrichment analysis detected the GATA3 motif in regions bound by OCT4 and SOX2 that transition to an active state in dEC. Furthermore, we found that regions bound by OCT4, NANOG, and SOX2 that gain an active mark in dEC are enriched for the motifs PAX9, p63, and STATs (Table S5). Examining epigenetic dynamics at sites of OCT4, NANOG, and SOX2 binding further supports the observation that some pluripotency-associated TFs are also involved in downstream specification.

**Gain of DNAme Occurs at Open Chromatin Enriched for TF Motifs**

We next utilized the WGBS data that cover ~26 million CpGs (at ≥5 coverage) across all four cell types. Hierarchical clustering analysis of the WGBS data, which included human adult liver and hippocampus for comparison, revealed that the pluripotent hESCs and the hESC-derived cell types form a separate cluster arm with respect to the somatic tissues (Figure 3A).

We defined DMRs as exhibiting a significant (p ≤ 0.05) minimal difference of CpG methylation level of 0.1 among our four cell types. The majority of all DMRs occur at CpG-poor intergenic regions in line with previous reports (Figure 3B, bottom) (Stadler et al., 2011). The dEN exhibits more than twice the number of regions that gain DNAme compared to dEC and dME (Figure 3B, top). Interestingly, only 65 of the total number of DMRs identified are shared between all three populations. However, reaffirming that our populations are depleted of pluripotent cells, this group of DMRs includes the regulatory region of OCT4 (Figure S3A). In line with the small number of shared regions, more than 60% of regions that gain DNAme are lineage specific (Figure 3C) and include loci such as SMAD3 (dEC), CTNNB1 (dME), and FOXA2 (dEN). FOXA2 has an upstream CGI that exhibits gain of DNAme (Figure 3D), and transcription in dEN is initiated downstream of this DMR at an alternative TSS, suggesting that TSS usage may be regulated, stabilized, or reflected by DNAme (Maunakea et al., 2010).
We find significant enrichment of various TF motifs as DNAme targets upon differentiation (Table S6), which has some analogy to the gain of methylation observed at myeloid targets in the lymphoid lineage in vivo (Ji et al., 2010; Deaton et al., 2011; Bock et al., 2012). To extend this observation, we examined the DNAme state at regions bound by SOX2, OCT4, and NANOG in hESCs. For example, two regions 20 kb downstream of DBX1, a gene associated with early neural specification, are bound by all

Figure 3. Global DNA Methylation Dynamics—Gain of DNAme
(A) Hierarchical clustering of hESCs, hESC-derived populations (dEC, dME, and dEN), human adult hippocampus, and human adult liver based on mean DNAme levels of 1 kb tiles across the human genome using Pearson Correlation Coefficient (PCC). y axis indicates sample distance in terms of 1 minus PCC. Red box indicates cell types interrogated in this study.
(B) Regions that significantly (p < 0.05) increase their DNAme levels by at least 0.1 between hESCs and the differentiated cell types. The color code indicates the DNAme state found in hESCs. Bottom: Genomic features associated with DMRs gaining DNAme in each of the differentiated cell types based on RefSeq gene annotation and de novo discovered promoters by RNA-seq.
(C) The overlap of these differentially methylated regions (DMRs) that increase their DNAme level in the three hESC-derived populations.
(D) DNAme levels and RNA-seq expression values of FOXA2 (chr20:22,559,343-22,571,189) in hESCs and differentiated cell types. The heat map below shows the DNAme values of individual CpGs within the highlighted region. The average DNAme value for the entire highlighted region is shown on the right in red. CGI are shown as green bars. Expression values (FPKM) are displayed on the right. The arrows indicate two known TSSs.
(E) DNAme levels and OCT4, SOX2, and NANOG ChIP-seq at the DBX1 locus (chr11:20,169,548-20,277,940).
(F) Promoters (left) and distal elements (right) that gain DNAme separated by the changes in FPKM at associated genes.
(G) Chromatin state in hESCs at regions that gain DNAme during differentiation. Regions devoid of any detected chromatin marks are categorized according to their DNA methylation state in hESCs.

See also Figure S3.
three TFs and gain DNAme in dME and dEN. In contrast, this region maintains low levels of DNAme in dEC, which has activated transcription of DBX1 (Figure 3E). We generally find that cobound sites gain DNAme in the dME and dEN, but not dEC (Figure S3B). Further supporting the functional relevance of these dynamics, we find that regions that gain DNAme frequently coincide with DNaseI hypersensitive sites (Figure S3C) (Thurman et al., 2012). Transcriptional silencing was less frequently correlated with gain of DNAme at distal elements than at promoters (Figure 3F).

In examining the chromatin state of regions that gain DNAme during differentiation, we find that most regions exhibited enrichment of one or more histone modifications in hESCs (Figure 3G). These results confirm that, in particular, distal regulatory elements show highly dynamic regulation of DNA methylation during specification.

**Loss of DNA Methylation Is Biased toward dEC**

Loss of DNAme is asymmetric between the three populations (Figure 4A, top) and occurs in a more lineage-specific fashion than gain (Figure 4B). However, loss also occurs mainly at intergenic regions (Figure 4A, bottom). Notably, the dEC has the most DMRs, and many were associated with neuronal gene categories (for instance: neural tube development, q = 3.13 x 10^{-13}). This includes the ectodermal TF POU3F1, which has a bivalent promoter in hESCs, resolves to a H3K4me3-only state (Figure S4A), and exhibits transcriptional activation in dEC (Table S2). Chromatin dynamics and activation at this locus coincide with specific loss of DNAme at a putative regulatory element downstream of the 3′ UTR in dEC (Figure 4C). On a global scale, an immediate correspondence between loss of DNAme and expression, such as that observed at POU3F1, occurs at about half of the regions (Figure 4D). More than 70% of DMRs that lose DNAme during differentiation are enriched for one of our profiled histone modifications, H3K4me1 or H3K27ac in particular (Figure 4E).

Taken together, our hESC differentiation system reveals several interesting DNAme dynamics, including the lineage-specific silencing of regulatory regions in default or alternative lineages. The asymmetric loss may also explain why our chromatin state analysis revealed fewer regions that gained H3K4me3 in the dEC population.

**Gain of H3K27ac Reveals Putative Regulatory Elements**

In addition to methylation on H3K4, open chromatin is also demarcated by enrichment of H3K27ac. It has also been suggested that the combination of H3K4me1 and H3K27ac at
distal regions identifies active enhancer elements, whereas H3K4me1 and H3K27me3 correspond to poised enhancer elements (Rada-Iglesias et al., 2011). To extend these observations, we focused specifically on regions that gain H3K27ac during differentiation and found that more than half of the identified regions are HMRs in hESCs (Figure 5A), whereas another large fraction is enriched for H3K4me1 in hESCs (Figure 5A). The majority of regions that gain H3K27ac are intergenic (Figure S5A), as shown for the RUNX1 locus (Figures 5A and 5B).

We next placed each region into one of three distinct categories (repressed, poised, and open) based on their state in hESCs and subsequently performed gene set enrichment analysis using the GREAT toolbox (Figure 5C) (McLean et al., 2010). This analysis reveals enhancer dynamics in line with the lineage-specific differentiation trajectory for dEC and dME (Figure 5C). In contrast, the dEN population shows an unexpected enrichment for early neuronal genes (e.g., neural tube development, Figure 5C). This observation is consistent with the correlation that we reported between our dEC and dEN RNA-seq data, suggesting that similar networks are induced in the early stages of both our ectoderm and endoderm specification (van Arensberg et al., 2010).

Moreover, we find strong enrichment of downstream effector genes of the TGFβ, VEGF, and BMP pathways in dME, directly reflecting the signaling cascades that were stimulated to induce the respective differentiation. In dEN, we find enrichment of genes involved in WNT/β-CATENIN and retinoic acid (RA) signaling (Figure 5C). Although we did not use RA, this signaling cascade has previously been implicated in endodermal tissue development, including pharyngeal and pancreatic cell types (Wendling et al., 2000; Oström et al., 2008). Concordantly, we also find high levels of SMAD3 motif enrichment in the repressed dME and dEN, particularly in the poised putative enhancer populations (Figure 5D). Similarly, we observe enrichment of key lineage-specific TF motifs such as the ZIC family proteins in dEC, TBX5 in dME, and SRF in dEN. Interestingly, we also find the FOXA2 motif highly overrepresented in dEN—in which the factor is active, and also dEC, in which the factor is inactive but becomes expressed at a later stage of neuronal differentiation (Kriks et al., 2011), but not in dME (Figure 5D).

**Acquisition of H3K4me1 without Transcriptional Activation Suggests Epigenetic Priming**

Many regions that exhibit high DNAme in hESCs and transition to H3K4me1 in one lineage remain HMRs in the two alternative cell types (Figures 2F and 6A). These regions are typically >10 kb from the nearest TSS (Figure 6B). GREAT analysis shows a strong enrichment for categories associated with brain development such as cerebellum morphogenesis in dEC (q < 10−30), TGFβ pathway targets (q < 10−10) in dME, and suppression of EMT in dEN (q < 0.0001). To understand whether regions that gain H3K4me1 in our system are associated with somatic identity, we took advantage of published microarray data for 24 human tissues and determined genes upregulated in these tissues with respect to hESCs (termed Tissue Atlas, see Extended Experimental Procedures). Reaffirming the relevance of our dynamics, we found that regions that gain H3K4me1 in dEC are associated with fetal brain and specific cell types found within the adult brain (Figure 6C). The dME H3K4me1 pattern was associated with a range of interrogated tissues, such as heart, spinal cord, and stomach, which may be due to heterogeneity of the tissues collected (Figure 6C). The dEN associations were interesting given that, as with the RNA-seq and H3K27ac trends, H3K4me1 was again associated with brain-related categories (Figure 6C).

Overall, less than half of the genes that gain H3K4me1 exhibit immediate transcriptional changes (Figure 6D). CYP2A6 and CYP2A7 (Figure 6E) are representative examples that do not show a corresponding change in expression, whereas LMO2 does (Figure 6F). To investigate these regions in more detail, we carried out motif enrichment analysis and found lineage-specific enrichment of TF motifs near regions that gain H3K4me1. Whereas the FOXA2 motif is enriched in all three cell types, the DBX1 motif is associated with the gain of H3K4me1 in dEC (Figure 6G), which coincides with its transcriptional activation in this cell type (FPKM: 5.36). Conversely, the GLI3, HIC1, and CTF1 motifs are strongly enriched at regions that gain H3K4me1 in dEN (Figure 6G).

To further assess whether this DNAme to H3K4me1 switch acts as a priming event, we differentiated the HUES64 endoderm population for 5 additional days in the presence of BMP4 and FGF2, leading to HNF4x-positive hepatoblast-like (dHep) cells (Table S2 and Figure S6A). Interestingly, of the motifs enriched in dEN that gain H3K4me1, HIC1, KLF4, and CTF1 (Figure 6G), several of these genes become expressed at the next stage of differentiation (Figure 6H). Lastly, 1,346 of these putatively primed regions are enriched for the active enhancer mark H3K27ac in human liver (Figure 6I).

**Loss of DNA Methylation and Acquisition of H3K27me3 at Putative Regulatory Elements**

More surprisingly, we observe distal/intergenic regions that switch from high DNAme to H3K27me3 (n = 3,985 in dEN) (Figure 7A). This transition frequently occurred within CpG-poor regions, which is distinct from the common CpG-island-centric targets of Polycomb Repressive Complex 2 and H3K27me3 (Figure 7B). This switch is highly lineage specific, and DNAme is generally retained in the alternative two cell types (Figures 7C and 2F).

Motif enrichment analysis, combined with the evaluation of publicly available TF binding site (TFBS) data from the ENCODE project, indicated that many regions exhibiting this transition in dEN were near binding sites of the pioneering factor FOXA2. This TF has putative roles in chromatin decompaction, but its distinct functions and limitations remain somewhat unclear (Li et al., 2012). To investigate this association, we performed ChIP-seq for FOXA2 in the endoderm population. This analysis reveals that FOXA2 binding sites frequently overlap with regions that transition from HMR to H3K27me3 (Figure 7D). We also confirmed that gain of H3K27me3 at dEN FOXA2 binding sites occurs predominantly in dEN and not dEC or dME (Figure 7E). A notable example of this transition can be seen at the ALB locus, where H3K27me3 is gained at AFP and AFM, proximal to FOXA2 binding sites (Figure 7F). Many regions that exhibit this transition are required for later stages of development as with AFP and AFM in the liver (Figure 7F) or HBB1 in the dME. As expected, the majority of these regions do not yet exhibit significant changes in expression (Figure S7A).
Figure 5. H3K27ac Dynamics at Putative Gene Regulatory Elements

(A) Number of regions and associated epigenetic state distribution in hESCs of regions that are transitioning to H3K27ac in the three populations.

(B) Normalized ChIP-seq tracks (H3K4me1, H3K27me3, and H3K27ac) for the RUNX1 region (chr21:36,091,108-36,746,447) with corresponding RNA-seq data in dME.

(C) GO categories enriched in regions transitioning to H3K27ac in the cell type indicated on the right compared to hESCs as determined by GREAT analysis. Regions gaining H3K27ac were split up by state of origin in hESC into repressed (none, IMR, HMR, and HK27me3), poised (H3K4me1/H3K27me3), and open (H3K4me3/H3K27me3, H3K4me3, and HK4me1). Color code indicates multiple testing adjusted q value of category enrichment.

(D) TF motifs enriched in regions changing to H3K27ac in the cell type indicated on the right compared to hESCs. Color code indicates motif enrichment score incorporating total enrichment over background as well as differential expression of the corresponding transcription factor in the respective cell type. Regions were split up by state of origin in hESCs similar to (C). For each region class, the eight highest-ranking motifs are shown.

See also Figure S5.
Figure 6. Characterization of H3K4me1 Dynamics at Putative Distal Regulatory Elements

(A) Overlap of regions gaining H3K4me1 in the three differentiated populations relative to hESCs.

(B) Genomic distribution of all regions gaining H3K4me1 compared to hESCs in at least one of the three differentiated populations.

(C) Tissue signature enrichment levels of genes assigned to regions specifically gaining H3K4me1 in the differentiated populations indicated on the bottom. For tissue signature definitions, see Extended Experimental Procedures.

(D) Number and distribution of gene expression changes of genes assigned to regions gaining H3K4me1 in the differentiated populations. Associated genes were classified as either being up/downregulated or unchanged relative to hESCs.

(E) Normalized ChIP-seq tracks (H3K4me1 and H3K36me3) for the LMO2 locus (Chr.11:33,865,134-33,977,858). Read counts on y axis are normalized to 10 million reads for each cell type. CGIs are indicated in green.

(F) Normalized ChIP-seq tracks (H3K4me1 and H3K36me3) for the CYP2A6/CYP2A7 region (Chr19: 41,347,260-41,395,599). Read counts on y axis are normalized to 10 million reads for each cell type. CGIs are indicated in green.

(G) Normalized motif enrichment scores for the top 15 motifs enriched in regions specifically transitioning to H3K4me1 in the differentiated cell type indicated on the bottom. Motif highlighted in red corresponds to a TF that is upregulated at the next stage (hepatoblast) of endoderm differentiation, whereas motifs highlighted in red are specifically upregulated in dEN but are downregulated at the dHep stage.

(legend continued on next page)
A previous report found that FOXA1/FOXA2 could bind to regions exhibiting DNAme (Sérandour et al., 2011), which is not a characteristic shared by all TFs (Gifford and Meissner, 2012). Regions bound by these factors subsequently lost DNAme and gained euchromatic histone modifications in our populations. We therefore compared DNAme at FOXA2 binding sites in hESCs to dEN and found a slight reduction specifically in the dEN (Figure 7G). To more directly assess this relationship, we interrogated the DNAme state of regions isolated by FOXA2-ChIP-BS sequencing in dEN (Brinkman et al., 2012). Interestingly, we saw a major depletion of DNAme at sites isolated by FOXA2-ChIP (Figure 7G). To determine whether these regions exhibit transcriptional activation after further differentiation, we examined again our dHep RNA-seq data and found that 50 genes, which were bound by FOXA2 and gained H3K27me3 in dEN, increased their expression at the next stage of differentiation (Figure 7H and Table S2). We also find H3K27ac enrichment at 197 loci in the human liver that had experienced the gain of H3K27me3 in dEN (Figure 7I).

DISCUSSION

Using directed differentiation of hESCs to three distinct, FACS-enriched populations representing early stages of embryonic development, we provide an extensive data set and highlight several insights on the transcriptional and epigenetic dynamics that occur during human in vitro lineage specification.

Among other things, we describe two very interesting but distinct lineage-specific dynamics from high DNAme to H3K4me1 or H3K27me3. These transitions occur at many sites that do not significantly change gene expression during our developmental biology. We provide an extensive data set and highlight several insights on the transcriptional and epigenetic dynamics that occur during human in vitro lineage specification.

In conclusion, examination of our data yields powerful insights on transcriptional and epigenetic dynamics during hESC specification and represents valuable reference maps for many applications, including regenerative biology and the study of human developmental biology.

EXPERIMENTAL PROCEDURES

For full details, see Extended Experimental Procedures.

Cell Culture

Human ESCs were cultured as previously described (Bock et al., 2011). All details regarding the differentiation conditions are in the Extended Experimental Procedures.

Gene Expression Signatures

RNA was hybridized to a custom probe set, processed using the Nanostring prep station, imaged using the Nanostring nCounter, and analyzed as previously described (Bock et al., 2011).

Sequencing Library Construction

For WGBS sequencing libraries, genomic DNA was fragmented using a Covaris S2 sonicator. DNA fragments were cleaned up, end repaired, A-tailed, and ligated with methylated paired-end adapters (purchased from ATDBio). For ChIP-seq libraries, ChIP-isolated DNA was end repaired, A-tailed, and ligated to barcoded Illumina adaptors, and the library was amplified using PFU Ultra II Hotstart Master Mix (Agilent). For RNA-seq libraries, polyadenylated RNA was isolated using Oligo dT beads (Invitrogen) and fragmented to 200–600 base pairs and then ligated to RNA adaptors using T4 RNA Ligase (NEB), preserving strand of origin information. ChIP-BS libraries were constructed as previously described (Brinkman et al., 2012).

Sequencing Data Processing and Analyses

Details regarding the WGBS, ChIP-seq, ChIP-bisulfite seq, and RNA-seq bioinformatics analysis can be found in the Extended Experimental Procedures. All data were aligned to hg19, and the accession numbers for each data set are included in Table S4.

ACCESSION NUMBERS

The accession numbers for the all of the sequencing data reported in this paper are listed in Table S4.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.04.037.
ACKNOWLEDGMENTS

We would like to thank Kendell Clement for support of the WGBS data visualization and Loyal Goff for RNA-seq visualization, as well as Zachary Smith and Jing Liao for critical reading of the manuscript, and the SCRB FACS Core for advice regarding FACS analysis. We also thank other members of the Meissner Lab and Epigenomics Platform at the Broad Institute for helpful discussion. A.M. is supported by the Pew Charitable Trusts and is a New York

Figure 7. DNAme to H3K27me3 Switch and FOXA2 Binding
(A) Distribution of genomic features associated with region gaining H3K27me3 (n = 22,643) upon differentiation to any of the three hESC-derived cell types compared to hESC.
(B) CpG content distribution of regions gaining H3K27me3 upon differentiation. For reference, the CpG content distribution of CpG islands is shown.
(C) Epigenetic state distribution in hESC, dEC, and dME of regions that gain H3K27me3 in the dEN population compared to hESC.
(D) Binding profile of FOXA2 in dEN (n = 357), OCT4 (n = 32), SOX2 (n = 12), and NANOG (n = 124) in hESC across regions that gain H3K27me3 in dEN upon differentiation.
(E) Composite plot of median normalized tag counts (RPKM) of regions bound by FOXA2 in dEN and gaining H3K27me3 in dEN compared to hESC (n = 357).
(F) Normalized H3K27me3 and H3K4me3 ChIP-seq tracks for hESCs, dEN, and human adult liver tissue at the ALB locus (chr4:74,257,882-74,377,753). Black bars (bottom) indicate TF binding of OCT4, SOX2, or NANOG in hESCs. Read counts on y axis are normalized to 10 million reads.
(G) Distribution of methylation levels of regions bound by FOXA2 and gaining H3K27me3 in dEN. DNAme information is depicted for hESC and dEN WGBS data sets and two biological replicates of FOXA2 ChIP-bisulfite experiments in dEN (n = 357).
(H) Gene expression profile of genes upregulated at the hepatoblast stage relative to dEN that are associated with regions bound by FOXA2 and gaining H3K27me3 in dEN (n = 50).
(I) Fraction of regions gaining H3K27me3 in dEN and being enriched for H3K27ac in human liver (n = 197).
See also Figure S7.
REFERENCES


