DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells

Christoph Bock,1,2,3,7 Isabel Beerman,2,3,4,6 Wen-Hui Lien,5,6 Zachary D. Smith,1,2,3,6 Hongcang Gu,1,6 Patrick Boyle,1 Andreas Gnirke,1 Elaine Fuchs,5,* Derrick J. Rossi,2,3,4,* and Alexander Meissner1,2,3,*

1Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
2Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
3Harvard Stem Cell Institute, Cambridge, MA 02318, USA
4Immune Disease Institute, Program in Cellular and Molecular Medicine, Boston Children’s Hospital, Boston, MA 02116, USA
5Howard Hughes Medical Institute, Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, NY 10065, USA
6These authors contributed equally to this work
7Present address: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria

SUMMARY

DNA methylation is a mechanism of epigenetic regulation that is common to all vertebrates. Functional studies underscore its relevance for tissue homeostasis, but the global dynamics of DNA methylation during in vivo differentiation remain underexplored. Here we report high-resolution DNA methylation maps of adult stem cell differentiation in mouse, focusing on 19 purified cell populations of the blood and skin lineages. DNA methylation changes were locus specific and relatively modest in magnitude. They frequently overlapped with lineage-associated transcription factors and their binding sites, suggesting that DNA methylation may protect cells from aberrant transcription factor activation. DNA methylation and gene expression provided complementary information, and combining the two enabled us to infer the cellular differentiation hierarchy of the blood lineage directly from genome-scale data. In summary, these results demonstrate that in vivo differentiation of adult stem cells is associated with small but informative changes in the genomic distribution of DNA methylation.

INTRODUCTION

DNA methylation provides a mechanism for robust and epigenetically heritable gene silencing. Mouse knockout studies have shown that the mammalian DNA methyltransferases Dnmt1, Dnmt3a, and Dnmt3b are essential for embryonic development, and that loss of DNA methylation interferes with tissue homeostasis (Bird, 2002). These studies clearly establish the functional relevance of DNA methylation as an epigenetic mark, but they provide only partial insight into the mechanisms by which DNA methylation contributes to the regulation of gene expression and cellular identity. Large-scale DNA methylation mapping has recently emerged as a complementary approach to functional knockout studies, enabling the genome-wide identification of genomic regions that change their DNA methylation states during cellular differentiation. In vitro studies have mapped DNA methylation patterns in pluripotent stem cells (Lister et al., 2009) and in their differentiating progeny (Meissner et al., 2008; Mohn et al., 2008; Stadler et al., 2011), and in vivo studies have started to uncover the DNA methylation dynamics of adult stem cell differentiation and cellular lineage commitment (Hodges et al., 2011; Ji et al., 2010). However, conclusions from in vitro studies tend to be confounded by the presence of cell culture artifacts, while early in vivo studies relied on relatively low-resolution mapping technologies or were based on heterogeneous cell populations.

To dissect the in vivo patterns of epigenetic regulation associated with the differentiation of mammalian cells, we established genomic maps of DNA methylation at single-base resolution for two types of adult stem cells (hematopoietic stem cells and hair follicle bulge stem cells) and for a broad selection of blood and skin cell types that are derived from these stem cells. Comparative analysis of 19 highly purified cell types identified DNA methylation changes associated with the differentiation of adult stem cells into progenitor cells, with lymphoid versus myeloid lineage choice among blood progenitor cells and with the specification of terminally differentiated cells in both lineages. Condensing our observations into a bioinformatic model of differentiation-associated changes, we were able to infer—with good accuracy—the hierarchy of cellular differentiation in the blood lineage based on the combination of DNA methylation and gene expression data.

RESULTS

Single-Base Resolution DNA Methylation Maps of Adult Stem Cell Differentiation

We established genomic DNA methylation maps for 19 highly purified cell populations from the blood and skin lineages of adult mice (Figure 1A). Stem cells, progenitor cells and terminally differentiated cells were purified by fluorescence-activated cell
sorting (FACS) under stringent conditions (Table S1 available online). This approach overcomes the heterogeneity of surgically obtained tissues and avoids artifacts that emerge during continued passaging of nonpluripotent cells in vitro (Meissner et al., 2008). For each cell type, two biological replicates were collected from different mice, purified in independent sorting
DNA Methylation in Adult Stem Cell Differentiation

experiments and subjected to genomic DNA methylation mapping (Table S2). In order to provide additional reference points for the identification of blood- and skin-specific DNA methylation patterns, we also mapped DNA methylation in embryonic stem cells (ESCs) and in two primary tissues (brain and liver). Furthermore, we obtained gene expression data for the same cell types, in part from public sources (Table S3 and the Experimental Procedures), allowing us to compare DNA methylation and gene expression differences between cell types (Figure 1B).

DNA methylation mapping was performed by reduced representation bisulfite sequencing (RRBS), which provides single-base pair resolution and highly quantitative data for a defined subset of cytosines in the genome (Figure S1). We selected RRBS as the most suitable method for this study because it can be applied to very rare cell types (Smith et al., 2012) and because its focus on a defined set of consistently sampled genomic regions confers sensitivity for detecting small differences and minimizes the number of measurements lost due to poor coverage (Bock et al., 2010). Our RRBS analyses covered on average 1.64 million individual CpGs throughout the mouse genome (Table S1), with excellent reproducibility between biological replicates (Pearson’s $r > 0.99$ for most cell types, Figure S1A). Because RRBS preferentially assays genomic regions with medium to high CpG density, these maps are particularly suitable for studying DNA methylation at putative gene-regulatory elements (Figures S1B and S1C).

When comparing DNA methylation maps between cell types, we observed high correlations within the blood and skin lineages, with Person’s $r$ ranging from 0.96 to 0.99 for most comparisons (Figure S1A). Nucleated erythrocytes were the only cell type that deviated substantially from all other cell types of the blood and skin lineages (Figures S1A and S1D), because they exhibited global loss of DNA methylation as reported previously (Shearstone et al., 2011). In contrast, all other differences observed within the two lineages were locus specific and relatively modest in magnitude. To be able to study these differences on a genomic scale, we developed a bioinformatic method that identifies differentially methylated regions and differentially expressed genes within the data set in a sensitive and robust manner (Supplemental Experimental Procedures). Detailed results of the differential DNA methylation and gene expression analyses are listed in Table S4, and the supplementary website (http://invivomethylation.computational-epigenetics.org/) provides genome browser tracks for visualizing all data at single-base resolution.

DNA Methylation Reflects Cellular Lineage Choice and Differentiation Stage

Gene expression profiles have been shown to accurately reflect cell type and differentiation stage (Lien et al., 2011; Novershtern et al., 2011), which supports their relevance for understanding cellular differentiation. To assess whether DNA methylation maps reflect cellular identity in similar ways, we performed hierarchical clustering based on DNA methylation data of all biological replicates (Figure 1C). This analysis could accurately distinguish among blood and skin cell types and among the additional reference samples (brain and liver tissue, ESCs). Blood stem and progenitor cells clustered together and separately from lymphocytes, while the skin cell types clustered according to the physiological compartment from which they were derived (hair follicle and interfollicular epidermis).

For comparison, we also performed hierarchical clustering based on gene expression profiles for the same cell types (Figure 1D). Overall, we obtained similar results, while also uncovering one global difference between the two clustering trees. For DNA methylation data (Figure 1C), the branching points within blood and skin lineages were located substantially lower in the tree than for the clustering based on gene expression data (Figure 1D). This observation indicated that between-lineage differences in DNA methylation dominate over within-lineage differences, while the magnitude of change for gene expression is more evenly distributed throughout the tree. To quantify this disparity in the way that epigenetic and transcriptional data reflect cellular identity, we compared the number of differentially methylated regions and the number of differentially expressed genes within and between lineages (Table S4). Indeed, DNA methylation differences were five times as frequent between lineages as they were within the blood and skin lineages, while gene expression differences were only twice as frequent between lineages as they were within lineages.

In order to assess the accuracy with which DNA methylation and gene expression data reflect cellular identity, we repeated the hierarchical clustering several thousand times based on randomly selected region sets of varying sizes (Figure 1E). Ten randomly selected 1 kilobase genomic regions were generally sufficient for achieving 50% correct clustering, while 1,000 or more randomly selected regions were required to achieve near-perfect clustering accuracies. These results were similar to the clustering accuracies observed for gene expression data, although near-perfect accuracy was already achieved based on the expression levels of 100 or more randomly selected genes (Figure S2). To further evaluate the predictive power of DNA methylation, the random selection was restricted to subsets of genomic regions that share a certain characteristic. We found that CpG islands and CpG island shores gave rise to accurate clustering results based on fewer data points than were required for promoter regions (Figure S2), which is consistent with prior reports proposing CpG island shores as hotspots of informative DNA methylation differences between samples (Irizarry et al., 2009). However, genomic regions with intermediate DNA methylation levels in the range of 40% to 60% turned out to be even more powerful predictors than CpG island shores (Figure S2). These observations suggest that moderately methylated regions (a large percentage of which overlap with CpG island shores) exhibit DNA methylation levels that accurately reflect cellular identity.

DNA Methylation and Gene Expression Changes Intersect at Cell Type-Specific Genes

While DNA methylation maps and gene expression profiles reflected cellular lineage choice and differentiation stage with comparable accuracy (Figures 1 and S2), we did not observe strong overlap between cell type specific DNA methylation and expression differences (Figure S3). This observation can be reconciled with the hypothesis that DNA methylation plays an
Figure 2. DNA Methylation and Gene Expression Changes Intersect at Cell Type-Specific Genes

(A) Correlation between differences in DNA methylation and in the expression of associated genes, focusing on differentially methylated regions in all cell types (left bars), among blood cells (center bars) and among skin cells (right bars).

(B) DNA methylation difference between skin and non-skin samples (core promoter).

(C) Blood-specific unmethylated & expressed genes.

(D) Skin-specific unmethylated & expressed genes.

Please cite this article in press as: Bock et al., DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells, Molecular Cell (2012), http://dx.doi.org/10.1016/j.molcel.2012.06.019
important role for gene regulation during development if we assume that it affects a moderate number of specific developmental genes, rather than a large and unspecific set of genes (Hemberger et al., 2009). In support of this hypothesis, the correlation of DNA methylation and gene expression across cell types was indeed modestly negative (Figure 2A), consistent with DNA methylation’s role as a repressive epigenetic mark. The negative correlation between DNA methylation and gene expression was more pronounced for gene promoters than for the wider gene locus (Figure 2A), suggesting a direct link between DNA methylation at gene-regulatory elements and the expression levels of the associated genes.

Given the relatively small number of genes with overlapping DNA methylation and gene expression changes, we reasoned that genes exhibiting consistently negative association between these two properties may constitute strong candidates for a cell type specific functional role. Indeed, we observed well-established marker genes exhibiting lineage-specific decreases in DNA methylation and concomitant increases in gene expression (box II in Figures S4A and 2B). For the blood lineage, this included transcription factors (TFs) involved in hematopoietic regulation (Sfpi, Lmo2), cellular surface antigens (Cd27, Cd93), and cytokines (Il16), and for the skin lineage, we detected multiple keratin genes (Krt15, Krt23, Krt27, Krt35) and several transcription regulators with an established role in skin differentiation (Cebpb, Gata3, Hoxa5). To confirm these observations in a more quantitative manner, we performed an extended gene set enrichment analysis (see the Supplemental Experimental Procedures) on all genes exhibiting lineage-specific decreased DNA methylation and concomitantly increased expression levels. This analysis detected significant enrichment of relevant gene sets, epigenetic signatures and gene-regulatory binding events that are characteristic of the blood and skin lineages, respectively (Figures 2C and 2D), supporting that this combined epigenetic and transcriptional filtering strategy is useful for identifying lineage-specific genes.

The utility of intersecting DNA methylation and gene expression differences was not restricted to the comparison between lineages, but also identified relevant genes associated with cellular differentiation within a lineage. For example, TFs associated with gene regulation in progenitor cells (Fli1, Sfpi) were characterized by lower DNA methylation and higher gene expression levels in common lymphoid progenitors (CLPs), while genes encoding T cell associated surface markers (Cd6, Cd8a, Cd8b1) were specifically unmethylated and highly expressed in CD8-positive T cells (Figure 2E). Furthermore, we found that cell type specific correlations between DNA methylation and gene expression frequently extended beyond gene promoters, allowing us to link distal regulatory elements to their target genes. A case in point was the reduced DNA methylation levels found exclusively in T cells at a putative enhancer element located 60 kilobases upstream of the Tcf7 gene, which encodes a TF that is specifically expressed in T cells (Figures 2F and 2G). In a similar way, we identified putative gene-regulatory elements for the T cell-specific surface marker gene Cd8b1 and the lymphoid TF gene encoding Lef1, both of which are specifically unmethylated and expressed in T cells (Figure S4B). These findings illustrate how locus-specific negative correlations between DNA methylation and gene expression can help identify cell type specific genes and their associated gene-regulatory regions.

**Loci and Binding Sites of Myeloid TFs Become Robustly Methyllated in Lymphoid Cells**

Within the hematopoietic system, the cell types of the lymphoid lineage control adaptive immunity, while myeloid cells mediate innate immune response and a number of other physiological functions of the blood. On an evolutionary timescale, the myeloid lineage is older and detectable already in primitive vertebrates, whereas the lymphoid lineage is exclusive to vertebrates. Functional evidence suggests that DNA methylation is crucial for lymphoid differentiation but largely dispensable for myeloid differentiation (Bröskie et al., 2009; Trowbridge et al., 2009); and a substantial number of DNA methylation changes have been identified that associate with lymphoid lineage differentiation (Ji et al., 2010).

Notably, our analysis identified roughly twice as many genomic regions with higher methylation in CLPs than regions with higher methylation in common myeloid progenitors (CMPs) (Figure 3A and Table S3). Regions that were differentially methylated between CLPs and CMPs tended to retain their increased or decreased DNA methylation levels in terminally differentiated lymphocytes (CD4, CD8, B cells), while the picture was more diverse for myeloid cells (Figure 3A). In terms of gene expression, we observed that lineage-specific downregulation was consistently retained downstream of both lymphoid and myeloid progenitor cells, while lineage-specific upregulation was reversed in terminally differentiated cells of both lineages.

---

(B) Scatterplot of differences in DNA methylation and in the expression of associated genes between skin cells and nonskin samples (i.e., ESCs, brain tissue, liver tissue, blood cells). Genomic regions that exhibit a significant positive correlation (I and III) or negative correlation (II and IV) are shown in boxes, and selected genes are highlighted.

(C) Enrichment of genomic properties among regions exhibiting blood-specific reduced DNA methylation and an associated increase in gene expression (box II in Figure S4A). Full results are available from http://invivomethylation.computational-epigenetics.org/.

(D) Enrichment of genomic properties among regions exhibiting skin-specific reduced DNA methylation and an associated increase in gene expression (box II in B).

(E) Scatterplot of differences in DNA methylation and in the expression of associated genes between two lymphoid cell types (CLP and CD8).

(F) Interplay of DNA methylation and gene expression at the Tcf7-Vdac1 gene locus. Expression of the Tcf7 gene is specific to T cells (left orange column), while Vdac1 is a widely expressed housekeeping gene (right orange column). The promoters of Tcf7 and Vdac1 are unmethylated in all cell types (left and right columns in gray). In contrast, a putative enhancer located upstream of Tcf7 and Vdac1 (Figure S1B) is specifically unmethylated in T cells (black column).

(G) Correlation between DNA methylation levels of the putative Tcf7 enhancer highlighted in (F) and the expression levels of the Tcf7 gene within the blood lineage. Genomic region: chr1:3,027,000–3,028,000.

See also Figure S4.
DNA Methylation in Adult Stem Cell Differentiation

Figure 3. Myeloid TFs and Their Binding Sites Become Methylated in Lymphoid Cells

(A) DNA methylation of genomic regions that are differentially methylated between CLPs and CMPs. Focusing only on regions that are significantly more highly methylated in CLPs than in CMPs (left) or vice versa (right), the diagram shows these regions’ mean DNA methylation levels for all cell types of the blood lineage. To be able to plot multiple regions with different DNA methylation levels on the same scale, DNA methylation levels (y axis) are shown as absolute differences relative to the mean methylation levels of CLPs and CMPs in each genomic region.

(B) Correlation between DNA methylation and expression levels at myeloid transcription-regulatory genes within the blood lineage. Shown are the mean DNA methylation levels (x axis) of the 1 kilobase tiling regions that exhibit the highest association with the genes’ expression levels (y axis). Genomic regions: Gata2, chr6:88,150,000–88,151,000; Tal1, chr4:114,732,000–114,733,000; Lmo2, chr2:103,798,108–103,798,109.

(C) Enrichment of genomic properties among regions that are significantly more highly methylated in CLPs than in CMPs (left) or vice versa (right).

(D) Interplay of DNA methylation and gene expression at the Zfp1m1 gene locus. Lymphoid-specific reduced expression of the Zfp1m1 gene is associated with increased DNA methylation levels at an intronic Gata1 binding site (center) and also at a putative enhancer element further toward the 3’ end of the gene (right). This putative enhancer overlaps with ChIP-seq binding peaks of multiple myeloid TFs, namely Gata1, Gata2, Lmo2,Lyf1, Runx1, Tal1, Erg, Fli1, Ght1b, and Meis1.

(E) Correlation between DNA methylation levels of the putative Zfp1m1 enhancer highlighted in (D) and the expression levels of the Zfp1m1 gene within the blood lineage. Genomic region: chr8:124,833,000–124,834,000. See also Figure S5 and Table S5.

Two additional lines of evidence support the hypothesis that DNA methylation plays an important role for silencing myeloid regulatory programs in the lymphoid lineage. First, we observed that lymphoid cells exhibited increased promoter DNA methylation levels at several key regulators of myeloid lineage commitment, which was correlated with robust transcriptional repression in these cell types (Figure 3B). Examples included the TFs Gata2 and Tal1, both of which interfere with normal differentiation when overexpressed in lymphoid cells (Goardon et al., 2002; Tipping et al., 2009). Another interesting case was Lmo2, which blocks T cell differentiation (Pike-Overzet et al., 2007) and causes T cell leukemia (McCormack et al., 2003) when aberrantly expressed in lymphoid cells. Although the promoter region of Lmo2 is largely depleted of CpG sites, we observed

(Figure S5A). These observations indicate that lymphoid versus myeloid lineage choice is associated with stable silencing of genes from the alternative lineage at the stage of multipotent progenitor (MPP2) differentiation into CLPs or CMPs. Furthermore, our data indicate that DNA methylation is more robustly used for myeloid gene repression within lymphoid cells than vice versa.

Please cite this article in press as: Bock et al., DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells, Molecular Cell (2012), http://dx.doi.org/10.1016/j.molcel.2012.06.019
a significant correlation between Lmo2 expression levels and the DNA methylation levels of a single CpG. This CpG is located 41 bp upstream of the transcription start site and overlaps with a Gata1 binding site, suggesting that it may interfere with Gata1 TF binding when methylated (Figure 3B). While this analysis cannot discriminate between causal and consequential changes, DNA methylation levels of a single CpG have been shown to act as a regulatory mechanism for TF binding and gene expression in other systems (Xu et al., 2009). Second, we found that binding sites of important myeloid TFs were strongly enriched among genomic regions that exhibited higher DNA methylation levels in CLPs compared to CMPs (Figure 3D and 3E). Previous ChIP-seq studies had identified binding sites for the known myeloid TFs Gata1, Gata2, Lmo2, and Runx1 (Hannah et al., 2011), all of which we found to be enriched among the CLP-specific hypermethylated regions. This trend was statistically significant on the genomic scale and observed for several regulatory elements located near well-characterized myeloid genes. For example, we observed differential DNA methylation at Gata1 and/or Gata2 binding sites in the Zfp811 gene, which encodes an important cofactor of Gata TFs (Figures 3D and 3E), for the erythrocyte-specific Rhd and Hpn genes as well as for the myeloid master regulator Sfpi1 (Figure S5B). In summary, these observations suggest that cell type-specific DNA methylation of myeloid TFs and their binding sites may provide a checkpoint against accidental activation of myeloid regulatory programs in lymphoid cells.

**Blood Stem Cell Differentiation Is Associated with Hypermethylation and Downregulation of Homeobox Genes**

Hematopoietic stem cells (HSCs) were the earliest identified adult stem cells and the first that could be purified and characterized with high stringency (Spangrude et al., 1988). The study of HSCs has shaped the functional definition of stem cells, which comprises the capacity for long-term self-renewal and the potency to differentiate into several types of specialized cells. Bona fide HSCs are rare, require complex purification procedures, and cannot be cultivated in vitro without losing their stemness, which likely explains why this important cell type has so far eluded comprehensive epigenetic characterization.

In our data set, HSC differentiation into progenitor cells was associated with a moderate number of DNA methylation changes (Figure 4A and Table S5), gain of DNA methylation being more common than its loss. These changes occurred gradually during adult stem cell differentiation into multipotent and lineage-committed progenitor cells and were robustly maintained in terminally differentiated cells. We observed a somewhat more complex picture when analyzing the gene expression changes associated with HSC differentiation (Figure S6A). On the one hand, stem-cell specific genes were gradually downregulated during differentiation and retained their reduced expression in terminally differentiated cells, which resembles the DNA methylation dynamics of HSC differentiation. In contrast, those genes that were specifically upregulated in progenitor cells became downregulated again in terminally differentiated cells (Figure S6A). This progenitor-specific gene cluster was highly enriched for cell cycle-associated genes, reflecting the highly proliferative nature of progenitor cells compared to relatively quiescent stem cells and terminally differentiated cells.

In order to identify putative HSC-specific regulator genes, we intersected DNA methylation and gene expression data, and we determined those genes that became downregulated and hypermethylated during HSC differentiation into progenitor cells. Examples of such genes included the hormone receptor Lhcg and the TF Smad6 (Figure S6B). Perhaps most remarkably, we identified four HSC-specific homeobox genes that appear to be repressed by DNA methylation in terminally differentiated cells (Figures 4B and S6B). These were the well-characterized oncogenes Hoxa9 and Pbx1, the candidate oncogene Hoxb5 (Bullinger et al., 2004) and the putative tumor suppressor gene Hoxa5, which is aberrantly methylated in some leukemia patients (Strathdee et al., 2007). Each of these genes followed its own trajectory in terms of downregulation and gain of DNA methylation, arguing against a single epigenetic switch that deactivates multiple homeobox genes during HSC differentiation. For example, the Hoxa5 locus was already partially methylated in HSCs and accumulated further DNA methylation while its expression decreased gradually; in contrast, the Hoxb5 locus gained modest levels of DNA methylation only in lymphoid cells, although it was already fully transcriptionally repressed within MPP2 cells (Figure 4C). In summary, these results suggest that certain homeobox genes accumulate DNA methylation during HSC differentiation, which may protect them from aberrant activation in progenitors and terminally differentiated cell types.

**Skin Stem Cells Remodel their Epigenome in Response to Cell-Cycle Activation and Cellular Differentiation**

Two distinct populations of skin stem cells can be purified from the hair follicle: quiescent telogen bulge stem cells (TBSCs) and activated anagen bulge stem cells (ABSCs). Both cell types are capable of long-term self-renewal and have the potency to differentiate into several types of specialized cells, thus fulfilling the defining criteria of stem cells (Blanpain et al., 2004). During the normal hair cycle, TBSCs and ABSCs give rise to short-lived matrix transit-amplifying cells (MTAcs), from which the seven differentiated lineages of the hair follicle are derived (including companion layer differentiated cells, CLDCs). Hair follicle bulge stem cells can also regenerate the epidermis, but only under the special conditions of wound repair. A recent study mapped three histone modifications (histone H3K4me3, H3K27me3, H3K79me2) during stem cell maintenance and differentiation (Lien et al., 2011). This histone modification data set provided us with the opportunity to compare the DNA methylation changes observed in our data set with a catalog of chromatin states in the same cell types.

In contrast to HSCs, hair follicle stem cells can be purified during their quiescent resting stage (TBSCs) and after their activation (ABSCs), when follicles are growing hair and stem cells undergo self-renewal. This characteristic allowed us to assess the prevalence of DNA methylation changes associated with stemness versus those associated with cell proliferation. In pairwise comparisons of DNA methylation maps (Figures 5A and S7A), we observed marginally higher correlation between the
two stem cell populations (ABSC versus TBSC, Pearson’s $r = 0.993$) than between the proliferating stem and progenitor cell populations of the hair follicle (ABSC versus MTAC, Pearson’s $r = 0.992$). Direct comparison between quiescent stem cells and proliferating progenitor cells resulted in a lower correlation (TBSC versus MTAC, Pearson’s $r = 0.987$). An even lower
correlation was observed when comparing hair follicle stem cells with epidermal progenitor cells (TBSC versus EPro, Pearson’s $r = 0.974$) and with terminally differentiated cells of the epidermis (TBSC versus EDif, Pearson’s $r = 0.982$), consistent with the fact that the interfollicular epidermis is maintained by a separate stem cell population under conditions of normal tissue homeostasis and may therefore constitute a distinct cellular lineage.

Figure 5. Skin Stem Cells Undergo Epigenetic Remodeling upon Cell-Cycle Activation and upon Differentiation
(A) Pairwise scatterplots of DNA methylation observed in skin stem cells (TBSC, ABSC) and skin progenitor cells (MTAC). Genomic regions that were significantly more highly methylated in TBSCs than in MTACs are highlighted in red across all three diagrams, and genomic regions that were significantly less methylated in TBSCs than in MTACs are highlighted in green. Numbered red squares highlight two specific genomic regions that are displayed in detail in (B).
(B) DNA methylation and histone modifications at the Epha2 gene locus. Red boxes refer to 1 kilobase genomic regions, with the mean DNA methylation level of each region shown as a percentage value. The frequency plots display ChIP-seq read counts of the promoter-associated histone modification H3K4me3 (green) and the transcription-associated histone modification H3K79me2 (blue) across the locus, based on published ChIP-seq data (Lien et al., 2011). Genomic region: chr4:140,853,000–140,883,000.
(C) Correlation between DNA methylation and gene expression levels at the Hoxc6 and Sox9 genes within the skin lineage. Genomic regions: Hoxc6, chr15:102,842,000–102,843,000; Sox9, chr11:112,646,000–112,647,000.
See also Figure S7.
To test whether skin stem cell differentiation is associated with similarly consistent DNA methylation changes as we observed for HSC differentiation, we identified all genomic regions that were differentially methylated between TBSCs and MTACs (Table S5) and highlighted them in each of the scatterplots (Figures 5A and S7A). Genomic regions that were more highly methylated in MTACs than in TBSCs (green) were on average also more highly methylated in TBSCs than in stem cells of the blood and skin lineages (Figures 6 and S8). A total of 248 genomic regions were more highly methylated in stem cells of the blood and skin lineages (Figures 6 and S8). Similarly, 258 genomic regions were more highly methylated in MTACs than in TBSCs (green) were on average also more highly methylated in MTACs than in stem cells of the blood and skin lineages (Figures 6 and S8). These shared signatures indicate that the DNA methylation changes in adult stem cell differentiation are to a large degree lineage specific.

**Stem Cells Exhibit Incomplete Programming of DNA Methylation toward the Blood and Skin Lineages**

In order to identify common themes associated with adult stem cell differentiation, we performed a systematic comparison between stem cells and committed progenitor cell types in the blood and skin lineages (Figures 6 and S8). A total of 248 genomic regions were more highly methylated in stem cells of both lineages when compared to the corresponding progenitor cells, which constitutes a modest enrichment over random chance (odds ratio = 1.3, p < 10^{-3}). Similarly, 258 genomic regions were less methylated in stem cells than in progenitor cells (odds ratio = 1.4, p < 10^{-3}). These shared signatures were statistically significant but small in absolute terms, indicating that the DNA methylation changes in adult stem cell differentiation are to a large degree lineage specific.

When we performed an extended gene set enrichment analysis separately on the blood and skin data and compared
DNA Methylation in Adult Stem Cell Differentiation

enriched properties, we observed more pronounced similarities (Figure 6A). For both lineages, genomic regions characterized by lower DNA methylation in stem cell than in progenitor cells were significantly enriched for open chromatin and DNase hypersensitive sites associated with other somatic lineages (brain, fibroblasts, heart, skeletal muscle, kidney) and with ESCs. We also observed enrichment of blood-specific TF binding, open chromatin and putative gene-regulatory elements among those genomic regions that were more highly methylated in blood stem cells than in progenitor cells. These results suggest that blood and skin stem cells retain open chromatin at gene-regulatory elements associated with other lineages, which is lost upon differentiation. At the same time, lineage-specific regulatory regions become increasingly unmethylated when adult stem cells differentiate into progenitor cells.

We performed a similar comparison for gene expression data and obtained quite different results (Figure S8), highlighting that DNA methylation and gene expression reflect complementary aspects of adult stem cell differentiation. Stem cell specific genes of the blood and skin lineage exhibited considerable overlap (odds ratio = 3.3, p < 10^{-99}), and an even stronger overlap was observed among genes that were specifically expressed in committed progenitor cell types as opposed to stem cells (odds ratio = 4.7, p < 10^{-99}). Stem cell-specific genes frequently coded for TFs carrying zinc-finger domains, overlapped with published gene signatures of stem cells, and exhibited an active chromatin state in the two adult stem cell populations for which epigenome data was available (TBSC, ABSC). In contrast, progenitor-specific genes exhibited strong enrichment for cell cycle, proliferation, and cancer gene signatures, consistent with the highly proliferative nature of most progenitor cell populations.

Combining DNA Methylation and Gene Expression Data Enables Inference of Cellular Lineage Hierarchies

Our efforts to dissect the role of DNA methylation in adult stem cell differentiation uncovered certain properties of cellular differentiation that were consistently observed among the blood and skin lineages (Figure 7A). First, gene-regulatory regions associated with other lineages and with ESCs became increasingly methylated during adult stem cell differentiation. Second, gene-regulatory elements of the chosen lineage were only partially demethylated in stem cells and exhibited significantly reduced DNA methylation levels in committed progenitors and in terminally differentiated cells. Third, committed progenitor cells were characterized by a strong proliferative gene expression signature, which was essentially absent from both stem cells and terminally differentiated cells. Finally, all cell types clustered by their known biological similarity in an unsupervised multidimensional scaling analysis of the complex data sets onto a two-dimensional map to depict the similarities and differences of cell types as spatial distances (Figure 7C). In this map, we tagged the most undifferentiated cell type as stem cell (green), highly proliferating cell types as progenitors (orange) and all remaining cell types as terminally differentiated cells (blue). Finally, we inferred a cellular differentiation tree by starting from the stem cells and iteratively connecting each cell type to the spatially closest stem cell or progenitor cell type that was already part of the tree.

This analysis was conducted without entering any prior knowledge on cell type specific marker genes, functional similarities between cell types or observations from lineage tracing studies. Instead it was solely based on a combination of two genomic data sets (DNA methylation and gene expression), well-established statistical methods, and a small set of rules that we derived from the observations made in the current study. The resulting predictions closely recapitulate our current understanding of hematopoiesis and its hierarchical differentiation tree rooted on the HSC (Figure 7C). For example, the inferred map accurately reflects the establishment of lymphoid and myeloid lineages downstream of MPP2 cells, and it correctly positions lymphocytes downstream of CLP, while placing monocytes and granulocytes downstream of GMPs. On the other hand, the predictions for the erythrocyte lineage are more debatable. While it is reasonable that the model classified nucleated erythrocytes as progenitor cells (they are a highly proliferating cell type that retains the potency for differentiation into enucleated erythrocytes), it would be consistent with widely accepted models of hematopoietic differentiation to place erythrocytes downstream of MEPs, and MEPs downstream of CMPs, although some evidence exists that is supporting the same relationship of MPPs and MEPs that we inferred from the data (Adolfsson et al., 2005).

When we applied the same approach to the skin lineage the proliferation ranks accurately reflected increased proliferation of progenitor cells relative to stem cells and terminally differentiated cells (Figure S9A). Furthermore, the two-dimensional map identified biologically plausible properties of the skin lineage hierarchy, such as strong separation between cell types of the hair follicle and epidermis and stepwise changes among the hair follicle cells (Figure S9B). However, all cell types scored equally in terms of their differentiation rank (Figure S9A), which precluded us from performing a confident classification of skin cell types into stem cells, progenitors, and terminally differentiated cell types. For this reason, we could not add directional arrows denoting differentiation trajectories into an otherwise highly accurate map of the skin lineage (Figure S9B). It is...
possible that higher plasticity within the skin lineage constitutes the biological reason for our inability to infer skin differentiation hierarchies, but technical reasons related to the smaller number of cell types, less developed cell sorting strategies or insufficient robustness of the bioinformatic methods could also explain this result. In summary, cellular lineage inference based on the combination of DNA methylation and gene expression data can be applied to the cell types of the blood lineage and may provide a useful complement to more classical approaches. However, further validation on additional cell types both within and beyond the blood lineage will be required to refine the proposed method.

**DISCUSSION**

In order to foster our understanding of DNA methylation in cellular lineage commitment, we established high-resolution genomic maps of DNA methylation for 19 cell types of the blood and skin lineages that were purified directly from the mouse, without exposing them to cell culture. These data constitute a single-base resolution resource of the DNA methylation dynamics of adult stem cell differentiation in vivo. Through integrated bioinformatic analyses, we discovered epigenetic regulatory patterns that were in part shared between the blood and skin lineages, raising the possibility that these patterns may constitute general principles of adult stem cell differentiation and epigenetic regulation of cellular lineage commitment. For example, we found that adult stem cells were characterized by reduced levels of DNA methylation at gene-regulatory elements associated with other lineages and with ESCs, while lineage-specific gene-regulatory elements (such as the binding sites of hematopoietic TFs) were more highly methylated in stem cells than in progenitor cells. This observation suggests that it is more feasible and biologically meaningful to establish an
epigenetic signature of stemness based on enrichment of certain chromatin features than by trying to identify a single set of stem cell-associated genomic regions that would be unmethylated in all stem cell populations. We also observed that gain of DNA methylation and loss of gene expression were robustly maintained downstream of adult stem cells, indicating that cellular identity is in part defined by epigenetic switches that change their state only once in the course of the differentiation. Such switches might act as gatekeepers and prevent differentiated cells from aberrantly expressing stem cell-associated genes. In a similar way, DNA methylation appears to provide a two-tier epigenetic barrier against spuriously expression of myeloid TFs in lymphoid cells. We observed that both the gene loci and the binding sites of key myeloid specification factors (Gata2, Tal1, and Lmo2) became robustly methylated during lymphoid differentiation, suggesting that DNA methylation may interfere not only with their transcription but also with the DNA binding of these TFs in case they are spuriously transcribed in a lymphoid cell. Given that all of these TFs are well-characterized oncogenes, it seems plausible to speculate that simultaneous epigenetic repression of TF gene loci and their binding sites across the genome may help protect lineage-committed cells from TF-induced oncogenesis.

Across all cell types analyzed, we found that DNA methylation and gene expression data reflected cellular lineage choice and differentiation stage with similar accuracy, although DNA methylation differences were comparatively rare among the blood and skin lineages, modest in magnitude and not strongly correlated with differences in gene expression. From a practical perspective, these results support the value of DNA methylation as a fingerprint of cellular identity, which may prove valuable for inferring the tissue of origin of samples that lack RNA of sufficient quality and quantity (e.g., archival tumors, forensic samples, museum collections). Furthermore, we could show that a relatively straightforward bioinformatic method was able to accurately infer the lineage hierarchy among blood cell types based on the combination of DNA methylation and gene expression profiles. While this method requires further validation and refinement based on larger data sets and additional cellular lineages, it may ultimately provide a fast and unbiased method for mapping the topology of cellular differentiation hierarchies.

**EXPERIMENTAL PROCEDURES**

**DNA Methylation**

Blood and skin cells were obtained from adult mice and purified by FACS using stringent sorting criteria (see the Supplemental Experimental Procedures and Table S1). RRBS libraries for DNA methylation analysis were prepared from 30 ng input DNA per biological replicate following a published protocol (Gu et al., 2011) and sequenced by the Broad Institute’s Genome Sequencing Platform on Illumina Genome Analyzer II or HiSeq 2000 machines. Bioinformatic data processing and quality control were performed as described previously (Bock et al., 2011).

**Gene Expression**

All gene expression data were based on the Affymetrix GeneChip Mouse Genome 430 2.0 Array. Microarray profiles for a subset of the cell types used in this study have been published previously (Chambers et al., 2007; Ji et al., 2010; Lien et al., 2011), and the corresponding data are publicly available from the Gene Expression Omnibus (GEO) repository (accession numbers: GSE20244, GSE6506, and GSE31028). Microarray data that were not publicly available have been submitted to GEO (GSE38557). All microarray data were obtained as CEL files, and they have been quality controlled and normalized in the same way in order to minimize batch effects (see the Supplemental Experimental Procedures).

**Data Analysis**

DNA methylation data were mapped to a 1 kilobase tiling region set of the mouse genome, giving rise to a total of 95,086 genomic regions for which quantitative DNA methylation levels were available throughout the data set. Similarly, gene expression data were mapped to Ensembl gene identifiers, resulting in 20,666 genes with robust data across the entire data set. Data from biological replicates were integrated and cell type-specific differences detected with a bioinformatic method that makes it possible to treat DNA methylation and gene expression data in an equivalent way (see the Supplemental Experimental Procedures). The extended gene set enrichment analysis combined two gene-based approaches—functional enrichment analysis with DAVID (http://david.abcc.ncifcrf.gov/) and a parametric gene set enrichment analysis based on MSigDB (http://www.broadinstitute.org/pgsql/ —with two region-based approaches, namely an enrichment analysis from chromatin annotations collected from Cistrome (http://cistrome.org/) and other sources, and a genomic feature analysis using Epigraph (http://epigraph.mpi-inf.mpg.de/). With the exception of the DAVID and EpiGRAPH analyses, which were performed with the publicly available web servers, all data analyses were performed with the R statistics package (http://www.r-project.org/).

**Lineage Inference**

The differentiation rank was calculated as the mean rank of DNA methylation for gene-regulatory elements specific to ESCs (higher levels resulted in higher ranks) and for gene-regulatory elements specific to the blood or skin lineage (lower levels resulted in higher ranks). In both cases, DNase hypersensitivity hotspot data and ChIP-seq peaks for transcription factor binding as well as chromatin modifications were obtained from public sources, and the DNA methylation levels at 1 kilobase tiling regions overlapping with these peaks were compared to the background of 1 kilobase tiling regions that did not overlap. Similarly, the proliferation rank was calculated based on the expression levels of genes included in proliferation-associated gene signatures obtained from MSigDB. Based on these scores, cell types were classified as stem cells (lowest differentiation rank), progenitor cells (proliferation score above the mean), and terminally differentiated cells (all remaining cell types). In the next step, multidimensional scaling was performed on the combination of the DNA methylation and gene expression profiles, using Pearson’s r as distance function (higher correlations corresponding to smaller distances) and averaging the normalized distance matrix between DNA methylation and gene expression (in order to give both data types equal weight). Finally, cell types were iteratively connected by arrows, starting from the stem cells and in the order of increasing differentiation scores. With each step, the next cell type in the list was connected to the spatially closest cell type that had already been selected and had been classified as either stem cell or progenitor cell.

**Supplementary Website**

Additional material is available from http://invivomethylation.computational-epigenetics.org/, including raw DNA methylation data (BAM and BED files), preprocessed DNA methylation and gene expression data tables, and genome browser tracks for interactive visualization of the DNA methylation data. All genomic coordinates in this paper and on the supplementary website refer to the mm9 (NCBI37) assembly of the mouse genome.

**ACCESSION NUMBERS**

The GEO accession number for all unpublished gene expression data reported in this paper is GSE38557.
SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, nine figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.06.019.

ACKNOWLEDGMENTS

We would like to thank other members of the Meissner, Rossi, and Fuchs labs for their support. From Rockefeller University, we also thank S. Matzel at the flow cytometry facility, as well as Nicole Stokes and the Comparative Biology Center animal facility. Furthermore, we thank Tarjei S. MikkelSEN for his contributions to the early parts of the project and Fontina Kelley for supporting the sequencing efforts. C.B. was supported by a Feodor Lynen Fellowship (Alexander von Humboldt Foundation) and a Charles A. King Trust Postdoctoral Fellowship (Charles A. King Trust, N.A., Bank of America, Co-Trustee). W.-H.L. was supported by a Harvey L. Karp Postdoctoral Fellowship and a Jane Coffin Child Postdoctoral Fellowship. E.F. is a Howard Hughes Medical Institute Investigator. A.M. was supported by the Massachusetts Life Science Center Investigator. A.M. was supported by a grant (to E.F.) from the NIH/NIAMS (R01AR31737) and partially supported by a grant (to D.R. and A.M.) and partially supported by a grant (to E.F.) from the NIH/NIAMS (R01AR31737).

Received: November 29, 2011
Revised: April 9, 2012
Accepted: June 21, 2012
Published online: July 26, 2012

REFERENCES


Molecular Cell

DNA Methylation in Adult Stem Cell Differentiation


