Inferring Cell Differentiation Processes Based on Phylogenetic Analysis of Genome-Wide Epigenetic Information: Hematopoiesis as a Model Case

Kanako O. Koyanagi*

Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

*Corresponding author: E-mail: kkoyanag@ist.hokudai.ac.jp.

Accepted: January 28, 2015

Abstract

How cells divide and differentiate is a fundamental question in organismal development; however, the discovery of differentiation processes in various cell types is laborious and sometimes impossible. Phylogenetic analysis is typically used to reconstruct evolutionary processes based on inherent characters. It could also be used to reconstruct developmental processes based on the developmental changes that occur during cell proliferation and differentiation. In this study, DNA methylation information from differentiated hematopoietic cells was used to perform phylogenetic analyses. The results were assessed for their validity in inferring hierarchical differentiation processes of hematopoietic cells and DNA methylation processes of differentiating progenitor cells. Overall, phylogenetic analyses based on DNA methylation information facilitated inferences regarding hematopoiesis.

Key words: ancestral state estimation, epigenome, DNA methylation.

Introduction

During development, cells divide and differentiate to form a complete organism. In developmental and evolutionary biology, it is important to understand the cell differentiation process and the evolution of multicellular organisms. Recent advancement in omics technologies has enabled us to capture snapshots of the transcriptomes and epigenomes of various cell types at different developmental stages (Gifford et al. 2013). However, comprehensive monitoring of these processes for multiple differentiating cell types during development remains a laborious task. Tissue stem cells, which are present only in small numbers, can be particularly difficult.

Phylogenetic analysis is generally used to reconstruct evolutionary processes based on inherent genetic and morphological characters, which can change over many generations. Similarly, if developmental changes within an individual are inherited through cell proliferation and differentiation, the differentiation process could be reconstructed by phylogenetic analysis of these developmental changes.

Epigenetic changes are "mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence" (Riggs et al. 1996). Cellular differentiation occurs through a progression of gene expression patterns, changes of which are recorded as epigenetic changes in the genome as DNA methylation, histone modification, and chromatin structure (Rivera and Ren 2013). Differentiating cells possess specific epigenetic patterns that are inherited through cell division. Thus, phylogenetic analysis of epigenetic information could reconstruct cell differentiation processes.

DNA methylation is the most well-understood epigenetic change. It is essential for mammalian development (Smith and Meissner 2013) and influences cancer development (Estecio and Issa 2011). DNA methylation status is related to gene expression and stably maintained after cell division (Cedar and Bergman 2012; Jones 2012). Underscoring its similarity to DNA sequence and thus its potential for use in phylogenetic analysis, DNA methylation status is maintained in a semiconservative manner, although it is accomplished by the enzyme Dnmt1 (Cedar and Bergman 2012). Experimental and simulation studies of DNA methylation status at limited loci in colon crypt cells suggest that DNA methylation data can be used to infer stem cell population dynamics and histories of the human colon (Yatabe et al. 2001; Kim and Shibata 2004; Nicolas et al. 2007). Therefore, DNA methylation information could be compatible with phylogenetic analysis.

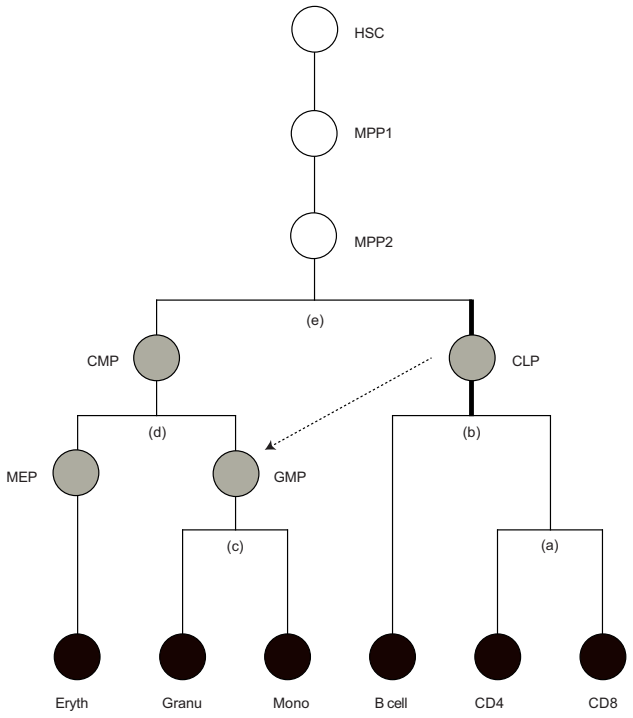

Fig. 1.—Differentiation model of hematopoiesis based on 13 cell types: HSC, MPP1, MPP2, CMP, MEP, GMP, CLP, Eryth, Granu, Mono, B cells, helper T cells (CD4), and cytotoxic T cells (CD8). For discussion of the thick branch and dotted line, see text.

In this study, I examined the feasibility of phylogenetic analysis using genome-wide DNA methylation information of differentiated cells to infer cellular differentiation processes in a murine hematopoietic system. Hematopoiesis is one of the best-studied cell differentiation processes and the hierarchical differentiation model is described in textbooks (fig. 1). Hematopoietic stem cells (HSC) can differentiate into myeloid or lymphoid lineages, producing six types of differentiated cells: Erythrocytes (Eryth), granulocytes (Granu), monocytes (Mono), B cells, helper T cells (CD4), and cytotoxic T cells (CD8). During hematopoiesis, multipotent progenitor 1 (MPP1), multipotent progenitor 2 (MPP2), common myeloid progenitor (CMP), megakaryocyte–erythroid progenitor (MEP), granulocyte–monocyte progenitor (GMP), and common lymphoid progenitor (CLP) cells proliferate and can be identified by molecular markers. Furthermore, DNA methylation has a critical role in hematopoiesis (Suarez-Alvarez et al. 2012; Jeong and Goodell 2014) and genome-wide methylation data have recently become available for all of these cell types (Bock et al. 2012). These data were used for the phylogenetic analyses described here. The aim of this study was to determine whether the DNA methylation data of adult differentiated cells could infer cell differentiation processes: The hierarchical differentiation process of hematopoietic cells and the DNA methylation processes of differentiating progenitor cells.

Results

To examine the validity of phylogenetic analysis based on DNA methylation data for inferring cell differentiation processes, I first examined how DNA methylation levels change during hematopoiesis. The genome-wide DNA methylation data for 13 murine hematopoietic cells contain information on methylation levels (0.0–1.0) for 83,505 sites. These sites were classified as "STABLE," "UP," "DOWN," and "OTHER" based on the changes during differentiation of each of six cell lineages (from HSC to Eryth, Granu, Mono, B cells, CD4, and CD8; see fig. 1). UP or DOWN classifications were assigned to 1,830–7,223 DNA methylation sites, depending on the cell lineage (fig. 2). These classifications indicated that these sites could be phylogenetically informative. Detailed clustering results for each lineage are available in the supplementary material S1, Supplementary Material online.

If DNA methylation patterns contain phylogenetic information, an inferred phylogenetic tree could reflect the cell differentiation process. On the basis of all 83,505 methylation sites, phylogenetic trees were inferred for six differentiated cell types—MEP, Granu, Mono, B cells, CD4, and CD8—with three progenitor cells—HSC, MPP1, and MPP2—as an outgroup. The erythrocyte lineage showed a high amount of demethylation (fig. 2), so MEP was used instead of Eryth to relax the long branch attraction effect (Felsenstein 1978). The methylation levels of most sites were near 0.0 or 1.0 (see supplementary material S2, Supplementary Material online). Therefore, the level (0.0–1.0) was first converted to binary (0 or 1), then phylogenetic trees were reconstructed with maximum parsimony (MP) and maximum likelihood (ML) methods (see Materials and Methods for detail). The reconstructed MP and ML trees had the same topology (fig. 3), which separated the leukocyte lineages (B cells, CD4, and CD8) with bootstrap probabilities of 100. This result was consistent with the known lineage (fig. 1), although the monophyly of myeloid lineages (MEP, Granu, and Mono) was violated by the outgroup.

The phylogenetic analysis identified 418 "homoplasy" sites, where the same DNA methylation changes (methylation or demethylation) occurred independently in multiple lineages. Such sites could hinder the correct phylogenetic inference. On the other hand, 2,392 sites were identified as "nonhomoplasy" sites, which could be suitable for phylogenetic inference. To characterize these DNA methylation sites, their target genes were examined and a gene enrichment analysis was performed on each group (homoplasy and nonhomoplasy sites). The results showed a clear difference between the two sets. Table 1 shows the enriched GO terms specific to nonhomoplasy sites. For example, transcription, immune response-regulating signal transduction, and hematopoiesis were specific to nonhomoplasy sites,

corresponding to the importance of their function in this differentiation process. On the other hand, homoplasy sites showed no specific function.

To determine whether the DNA methylation states in progenitor cells can be inferred from those of differentiated cells, the ancestral DNA methylation state for each internal node was estimated based on the adult differentiated cells. These results were compared with the progenitor cells (CLP, GMP, and CMP) for the tree representing known differentiation processes (fig. 1). The percent identities between internal nodes and progenitor cells were calculated among 3,182 sites (table 2), in which at least one differentiated cell had a different methylation state. Because the exact positions of progenitor cells in the tree (fig. 1) were unknown (e.g., CLP could be positioned somewhere in the branch from node e to b represented by a thick line in fig. 1), columns of the start and end positions (two internal nodes) for each progenitor cell are shaded in table 2. Underlined and italicized numbers represent the most similar progenitor cell for each internal node and the most similar internal node for each progenitor cell, respectively. The results demonstrated that 77.0–88.8%, 78.4–87.3% and 80.8–89.8% of the DNA methylation sites were correctly inferred for CLP, GMP and CMP, respectively.

Discussion

The results of this study supported the validity of a phylogenetic approach using DNA methylation information for inferring processes of cell differentiation in hematopoiesis. First, DNA methylation levels of most of the nonstable sites (91% on average) were gradually increasing or decreasing during cell differentiation. This result suggests that most DNA methylation changes are not highly variable (fig. 2). Variable changes (e.g., a zigzag pattern) were only found for limited sites in the monocyte and T-cell lineages (see supplementary material S1, Supplementary Material online). Thus, DNA methylation changes were compatible with phylogenetic analysis. Second, the phylogenetic trees inferred from DNA methylation information were similar to known cell differentiation processes of hematopoietic cells (fig. 3), although some inconsistencies were observed (discussed below). Thus, phylogenetic analysis using DNA methylation data was able to reconstruct the hierarchical differentiation process of hematopoietic cells. Third, target genes of "nonhomoplasious" sites were involved in immune response-regulating signal transduction and hematopoiesis (table 1). This result is consistent with their lineage-specific roles in the differentiation of hematopoietic cells (Novershtern et al. 2011). In addition, enrichment of target genes involved in transcriptional regulation reflects the regulated expression by DNA methylation during differentiation. Thus, the epigenetic changes in these genes could contain appropriate phylogenetic information and reflect cell differentiation processes. Finally, the ancestral state estimation was able to infer the DNA methylation states of differentiating progenitor cells from adult differentiated cells with 77.0–89.8% accuracy (table 2). Thus, a phylogenetic approach could infer DNA methylation processes. Cluster analysis performed with omics data often produces similar trees but cannot infer the states of progenitor cells. These results show phylogenetic analysis based on DNA methylation information of differentiated cells can be used to infer cellular differentiation processes, at least for hematopoiesis. While this manuscript was under submission, Capra and Kostka (2014) reported that a phylogenetic approach can be used to infer missing (masked) methylation states using both adult differentiated cells and progenitor cells also based on the data from Bock et al. (2012), which also shows the feasibility of phylogenetic analysis by DNA methylation.

The limitations of this study can be summarized in four points. First, the analyzed data were averaged DNA methylation levels for cell populations. As a result, heterogeneity in the source populations could affect the phylogenetic results. For example, lymphoid cells have the potential to differentiate into myeloid cells (dotted line in fig. 1 and Katsura and Kawamoto 2001), which might account for the inconsistent topology of the myeloid lineage (fig. 3). In addition, different granulocytic subtypes are derived from different progenitor cells: Neutrophils from CLP and basophils/eosinophils from CMP (Gorgens et al. 2013), which might also account for the paraphyly of the myeloid lineage (fig. 3). However, new single-cell technologies are emerging (Shapiro et al. 2013; Tsioris et al. 2014) and methods for obtaining genome-wide DNA methylation information from a single cell have been reported (Guo et al. 2013; Smallwood et al. 2014), which can overcome this problem. In addition, DNA methylation data from single cells are binary so that binarization of continuous values, which means loss of information, is not necessary, although continuous characters can be used for phylogenetic analysis (Webster and Purvis 2002). Second, phylogenetic inference could be incorrect for some cases. For example, global demethylation occurs in the erythrocyte lineage (fig. 2 and Shearstone et al. 2011). Such a lineage represents a long branch on a phylogenetic tree, which could lead to incorrect topology, known as the long branch attraction problem in MP (Felsenstein 1978). Third, the model of hematopoiesis itself has been under debate due to recent advancements in lineage tracing technologies—several models other than the classical model (fig. 1) exist as of this writing (Kawamoto et al. 2010; Ema et al. 2014). One of these models shows early branching of the erythrocyte lineage, which is consistent with the tree inferred in this study (fig. 3). Thus, further experimental studies are awaited. Finally, the results here are for hematopoiesis; the extent to which these conclusions can be generalized is unknown. These limitations could be the subject of future work.

The differentiation process estimated here could be different from the cell lineage tree, which is the history of cell division. The complete cell lineage was first revealed in

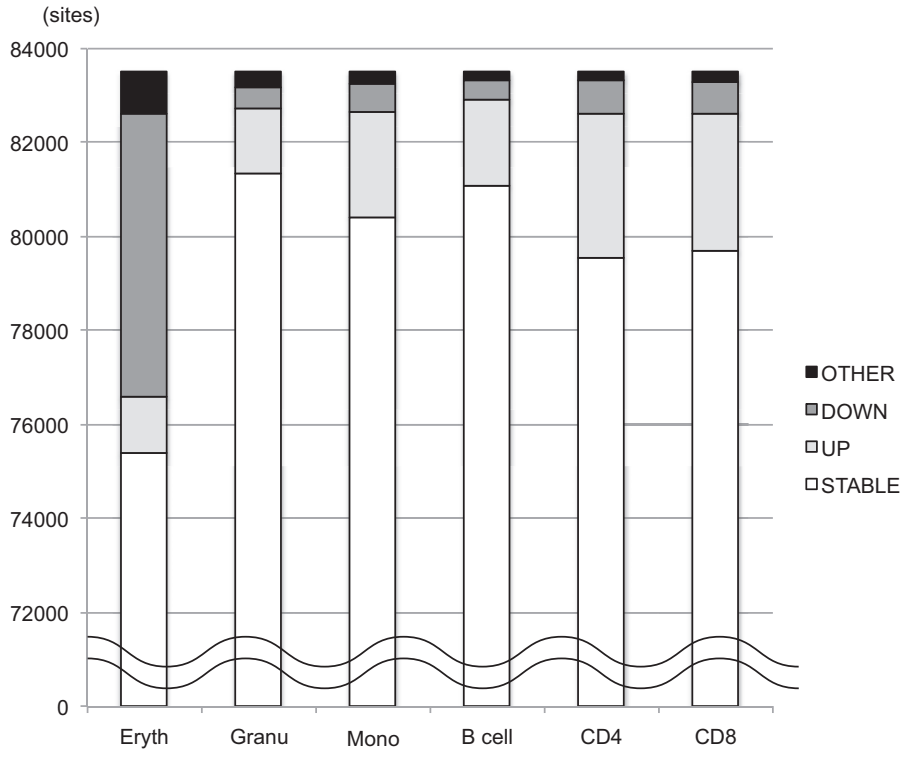

Fig. 2.—DNA methylation sites classified by their changes during cell differentiation. The categories are STABLE (the DNA methylation level was stable), UP (increasing), DOWN (decreasing), and OTHER (not classified into the aforementioned three categories).

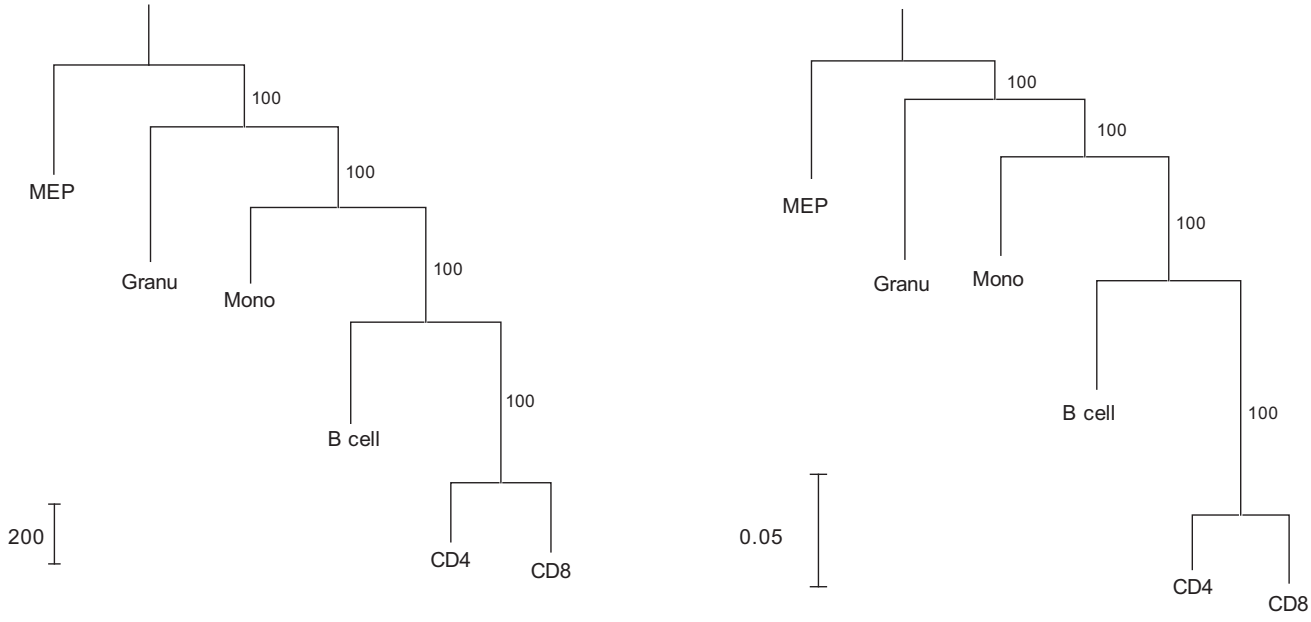

Fig. 3.—MP (left) and ML (right) trees inferred from DNA methylation information. A total of 83,505 sites were used. Numbers represent bootstrap probabilities. HSC, MPP1, and MPP2 were used as an outgroup.

Table 1
Enriched GO Annotation for Nonhomoplasious Sites

Enriched GO Annotation[a]		*P* Value	Benjamini[b]
Transcription	GO:0006350	2.68E-07	5.52E-05
Immune response-regulating signal transduction	GO:0002757, GO:0002764, GO:0002429, GO:0050851, GO:0002768	2.98E-07	5.76E-05
Negative regulation of transcription	GO:0016481, GO:0000122, GO:0045934, GO:0045892, GO:0051172, GO:0051253, GO:0010629, GO:0009890, GO:0010558, GO:0031327, GO:0010605	9.33E-07	1.70E-04
Positive regulation of transcription	GO:0045944, GO:0031328	1.37E-06	2.12E-04
Hematopoiesis	GO:0002520, GO:0048534	7.67E-06	8.79E-04
Intracellular signaling cascade	GO:0007242	1.16E-05	1.24E-03
Embryonic development	GO:0048568, GO:0043009, GO:0009792	4.16E-05	3.57E-03
Blood vessel morphogenesis	GO:0048514	4.94E-05	4.12E-03
Tube development	GO:0035295, GO:0035239	6.25E-05	5.08E-03
Cell migration	GO:0016477	8.60E-05	6.63E-03

[a]The representative biological terms associated with the clusters were manually summarized. *P* value and Benjamini of the top GO terms are shown.
[b]Multiple-testing corrected *P* value.

Table 2
Percent Identities of DNA Methylation Sites between Internal Nodes and Progenitor Cells (%)

Internal Nodes	Progenitor Cells								
	MP-ACCTRAN			MP-DELTRAN			ML		
	CLP	GMP	CMP	CLP	GMP	CMP	CLP	GMP	CMP
(a) CD4, CD8	68.4	58.4	63.9	69.1	59.9	65.4	64.2	55.6	60.3
(b) B cell, CD4, CD8	83.5	73.6	79.6	84.2	74.6	80.6	77.0	65.7	71.6
(c) Granu, Mono	86.0	*87.3*	88.7	*87.3*	*87.1*	89.1	70.9	81.8	73.2
(d) MEP, Granu, Mono	86.7	85.6	*89.3*	86.4	84.8	*89.8*	75.9	78.4	80.8
(e) MEP, Granu, Mono, B cell, CD4, CD8	*88.8*	84.0	89.2	86.4	83.7	88.8	*84.6*	*85.4*	*86.8*

Note.—Shaded columns represent the start and end points of differentiation for each progenitor cell (see fig. 1). Columns with highest identity in each line and row are shown in underline and italic, respectively. Alphabets in parentheses correspond to those in figure 1.

Caenorhabditis elegans by direct observation (Sulston et al. 1983) and has been analyzed for other organisms by various cell lineage tracing methods (Stern and Fraser 2001; Blanpain and Simons 2013). Cell lineage trees can also be analyzed by phylogenetic analysis of somatic mutations such as microsatellites (Frumkin et al. 2005), polyguanine repeats (Salipante and Horwitz 2006), and substitutions (Behjati et al. 2014); however, the number of mutations per genome is rather small compared with the number of epigenomic changes. Cell lineage trees represent the history of cell divisions, whereas a differentiation process estimated by epigenomes would not reflect cell divisions. The same epigenetic status can be maintained after cell division, whereas it can change during development without cell division. Thus, the differentiation process estimated in this study could be considered as an average landscape of epigenetic changes through hematopoiesis rather than a history of cell divisions. Combining the phylogeny of epigenomes and the cell lineage tree, together with transcriptome and proteome data from single cells will deepen our understanding of organismal development.

Materials and Methods

Genome-wide DNA methylation data for murine hematopoietic cells were obtained from supplementary table S2 of Bock et al. (2012). These data include high-confidence DNA methylation measurements determined by reduced representation bisulfite sequencing (RRBS), which is an enrichment strategy for capturing the majority of CpG islands and promoters in the genome (Gu et al. 2011). DNA methylation levels (0.0–1.0) are described for each 1-kb genomic region (called DNA methylation sites in this study) with sufficient RRBS coverage. Uncertain DNA methylation sites lacking concordance between two biological replicates were excluded from the analysis. In total, 83,505 DNA methylation sites were available for HSC, six differentiating progenitor cells (MPP1, MPP2, CMP,

MEP, GMP, and CLP), three differentiated myeloid cells (Eryth, Granu, and Mono), and three differentiated lymphoid cells (CD4, CD8, and B cells).

To characterize how DNA methylation changes throughout cell differentiation, I first performed *k*-means clustering ($k = 100$) for 83,505 DNA methylation sites in each cell lineage (fig. 1). For example, the erythrocyte lineage differentiates from HSC > MPP1 > MPP2 > CMP > MEP to erythrocyte. The DNA methylation levels (0.0–1.0) for these six cell types represent the putative time-course methylation changes through differentiation. These six values were treated as a vector for each DNA methylation site. On the basis of these vectors, 83,505 sites were clustered into 100 clusters using the kmeans() function in R (3.0.2) with Lloyd's algorithm.

Each cluster was then classified as STABLE, UP, DOWN, or OTHER based on the pattern of methylation changes during cell differentiation. A third-order polynomial was fitted to the pattern for each cluster using lm() in R. If the estimated polynomial function was flat, where the difference between the maximum and the minimum values of the function was within 0.2 and all gradients for each time point (cell) had values between −0.1 and 0.1, the cluster was classified as STABLE. If the estimated polynomial function was increasing, where all gradients had positive values (greater than −0.1 after accounting for fluctuation), the cluster was classified as UP. If the polynomial function was decreasing, where all gradients had negative values (less than 0.1 after accounting for fluctuation), the cluster was classified as DOWN. The remaining clusters were classified as OTHER. According to this procedure, all the DNA methylation sites belonging to any clusters were classified into STABLE, UP, DOWN, and OTHER.

For phylogenetic analyses, the DNA methylation level (0.0–1.0) was transformed into binary data as 0 for 0.0–0.4 (unmethylated) and 1 for 0.4–1.0 (methylated). The rationale for the cut-off value of 0.4 was based on Bock et al. (2012) who reported "genomic regions with intermediate DNA methylation levels in the range of 40% to 60% turned out to be even more powerful predictors." Adult differentiated cells (Granu, Mono, B cells, CD4, and CD8) and MEP (see Results section) were used for the phylogenetic analyses with progenitor cells (HSC, MPP1, and MPP2) as an outgroup.

MP Method: On the basis of the binary DNA methylation data, the MP tree was inferred using PAUP 4.0 (Swofford 2003). The character type was treated as undirected (cost of methylation was equal to that of demethylation) and an exhaustive search was performed. Branch support was estimated by 1,000 bootstrap replicates. To examine whether the DNA methylation states of progenitor cells can be inferred from adult differentiated cells, the ancestral state for each node was inferred with accelerated transformation (ACCTRAN) and delayed transformation (DELTRAN) algorithms based on the fixed tree topology shown in figure 1. A methylation site whose CI was estimated as 1.0 was defined as a site of nonhomoplasy, and a methylation site whose CI was estimated as less than 1.0 and showed same DNA methylation changes (methylation or demethylation) in multiple lineages was defined as a site of homoplasy in this study.

ML Method: On the basis of the binary DNA methylation data, the ML tree was inferred using RAxML v8.1.2 (Stamatakis 2014) with the GTR + GAMMA + I model. Branch support was estimated by rapid bootstrap analysis with 1,000 replicates (-f a option). Ancestral states for internal nodes were estimated (-f A option) based on the fixed tree topology shown in fig. 1.

To characterize each DNA methylation site, target genes for each site were obtained from supplementary table S2 of the article by Bock et al. (2012) and then gene enrichment analysis was performed. To compare the characteristics of DNA methylation sites identified as nonhomoplasy sites and those identified as homoplasy sites, enriched Gene Ontology (GO) annotation terms of biological processes were explored by using the Functional Annotation Chart tool of the Database for Annotation Visualization and Integrated Discovery (DAVID) v6.7 (Huang da et al. 2009). GO terms with a Benjamini-corrected *P* value, which is a multiple-testing correction of the EASE score (a modified Fisher's exact test), less than 0.01 were considered significantly enriched. Enriched GO terms were compared between the two sets (nonhomoplasy vs. homoplasy sites), so that enriched GO terms specific to each set were identified. To summarize the results, GO terms were clustered by using the Functional Annotation Clustering tool (default settings).

Supplementary Material

Supplementary materials S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The author thanks H. Watanabe for valuable comments. This work was supported by JSPS KAKENHI Grant Number 25440186.

Literature Cited

Behjati S, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature 513:422–425.
Blanpain C, Simons BD. 2013. Unravelling stem cell dynamics by lineage tracing. Nat Rev Mol Cell Biol. 14:489–502.
Bock C, et al. 2012. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell 47:633–647.
Capra JA, Kostka D. 2014. Modeling DNA methylation dynamics with approaches from phylogenetics. Bioinformatics 30:i408–i414.
Cedar H, Bergman Y. 2012. Programming of DNA methylation patterns. Annu Rev Biochem. 81:97–117.
Ema H, Morita Y, Suda T. 2014. Heterogeneity and hierarchy of hematopoietic stem cells. Exp Hematol 42:74–82.
Estecio MR, Issa JP. 2011. Dissecting DNA hypermethylation in cancer. FEBS Lett. 585:2078–2086.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410.
Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E. 2005. Genomic variability within an organism exposes its cell lineage tree. PLoS Comput Biol. 1:e50.
Gifford CA, et al. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell 153:1149–1163.
Gorgens A, et al. 2013. Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. Cell Rep 3:1539–1552.
Gu H, et al. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6: 468–481.
Guo H, et al. 2013. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. 23:2126–2135.
Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.
Jeong M, Goodell MA. 2014. New answers to old questions from genome-wide maps of DNA methylation in hematopoietic cells. Exp Hematol 42:609–617.
Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 13:484–492.
Katsura Y, Kawamoto H. 2001. Stepwise lineage restriction of progenitors in lympho-myelopoiesis. Int Rev Immunol 20:1–20.
Kawamoto H, Ikawa T, Masuda K, Wada H, Katsura Y. 2010. A map for lineage restriction of progenitors during hematopoiesis: the essence of the myeloid-based model. Immunol Rev 238:23–36.
Kim KM, Shibata D. 2004. Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. BMC Gastroenterol 4:8.
Nicolas P, Kim KM, Shibata D, Tavare S. 2007. The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput Biol. 3:e28.
Novershtern N, et al. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 144:296–309.
Riggs AD, Martienssen RA, Russo VEA. 1996. Epigenetic mechanisms of gene regulation. Plainview (NY): Cold Spring Harbor Laboratory Press.
Rivera CM, Ren B. 2013. Mapping human epigenomes. Cell 155:39–55.
Salipante SJ, Horwitz MS. 2006. Phylogenetic fate mapping. Proc Natl Acad Sci U S A. 103:5448–5453.
Shapiro E, Biezuner T, Linnarsson S. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet. 14:618–630.
Shearstone JR, et al. 2011. Global DNA demethylation during mouse erythropoiesis in vivo. Science 334:799–802.
Smallwood SA, et al. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11:817–820.
Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. Nat Rev Genet. 14:204–220.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Stern CD, Fraser SE. 2001. Tracing the lineage of tracing cell lineages. Nat Cell Biol. 3:E216–E218.
Suarez-Alvarez B, Rodriguez RM, Fraga MF, Lopez-Larrea C. 2012. DNA methylation: a promising landscape for immune system-related diseases. Trends Genet. 28:506–514.
Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol. 100: 64–119.
Swofford DL. 2003. PAUP*: phylogenetic analysis using parsimony (and other methods) version 4. Sunderland (MA): Sinauer Associates.
Tsioris K, Torres AJ, Douce TB, Love JC. 2014. A new toolbox for assessing single cells. Annu Rev Chem Biomol Eng 5:455–477.
Webster AJ, Purvis A. 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. Proc Biol Sci. 269: 143–149.
Yatabe Y, Tavare S, Shibata D. 2001. Investigating stem cells in human colon by using methylation patterns. Proc Natl Acad Sci U S A. 98: 10839–10844.

Associate editor: Takashi Gojobori