ORIGINAL ARTICLE

Regional DNA methylation differences between humans and chimpanzees are associated with genetic changes, transcriptional divergence and disease genes

Kei Fukuda[1], Kenji Ichiyanagi[1], Yoichi Yamada[2], Yasuhiro Go[3], Toshifumi Udono[4,5], Seitaro Wada[6], Toshiyuki Maeda[7], Hidenobu Soejima[7], Naruya Saitou[8], Takashi Ito[9] and Hiroyuki Sasaki[1]

Changes in gene expression have been proposed to have an important role in the evolutionary changes in phenotypes. Interspecific changes in gene expression can result not only from genetic changes in regulatory regions but also from epigenetic changes in such regions. Here we report the identification of genomic regions showing differences in DNA methylation between humans and chimpanzees (termed S-DMRs for species-specific differentially methylated regions) on chromosomes 21 and 22. These regional methylation differences are frequently associated with genes, including those relevant to a disease, such as Alzheimer's disease, diabetes mellitus or cancer. Methylation differences are often correlated with changes in promoter activity or alternative splicing. Comparative studies including other great ape species provide evidence for the contribution of genetic changes to some of these S-DMRs. Genetic changes responsible for the S-DMRs include gain or loss of CTCF-binding site and changes in CpG density in microsatellite repeats. Our results suggest that DNA methylation changes, often caused by small sequence changes, contribute to transcriptional and phenotypic diversification in hominid evolution.

INTRODUCTION

Humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) have approximately 1.2% difference in genomic sequence and yet show many phenotypic differences.[1] In addition to the changes in protein repertoire, structure and function, the differences in gene expression patterns are considered to have an important role in phenotypic diversification.[2–7] Indeed, comparative studies have identified transcriptional differences between humans and chimpanzees.[3,8–11] Such differences can arise not only from genetic changes, such as gene duplication, translocation and nucleotide changes in regulatory elements[12,13] but also from changes in epigenetic modifications in the regulatory regions. Although genetic differences between humans and chimpanzees have been extensively studied,[4] epigenetic differences are less well characterized.

Cytosine methylation at CpG sites is a major epigenetic modification of genomic DNA regulating gene expression.[14,15] In mammals, CpG sites are under-represented because of the high mutability of methylated cytosine. However, there are regions of 0.5–2 kilobases (kb) harboring many CpG sites, which are called CpG islands (CGIs). CGIs are often located in gene promoters and are generally unmethylated. Methylation of these regions represses the gene, and the methylated state is heritable through cell divisions, implying that methylation is important for maintaining proper expression patterns. Indeed, targeted disruptions of the genes coding for the DNA methyltransferases resulted in misregulation of genes and embryonic lethality in mice.[16,17] It is also found that the methylation level in regions immediately adjacent to CGIs, called CGI shores, is variable in development, carcinogenesis and reprogramming.[18,19] DNA methylation can contribute to phenotypic variation within the species.[20] Furthermore, although methylation is, for the most part, reprogrammed in each generation, a methylation pattern can be transmitted through the germline in certain sequences.[20] It is possible that some phenotypic differences induced by DNA methylation can be transmitted through multiple generations.

Several studies reported DNA methylation differences between the great ape species. Ernard *et al.*[21] detected methylation differences

[1]Division of Epigenomics and Development, Medical Institute of Bioregulation, and Epigenome Network Research Center, Kyushu University, Fukuoka, Japan; [2]Faculty of Electrical and Computer Engineering, Institute of Science and Engineering, Kanazawa University, Kanazawa, Japan; [3]Molecular Biology Section, Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan; [4]Sanwa Kagaku Kenkyusho Co., Ltd., Uto, Kumamoto, Japan; [5]Kumamoto Sanctuary, Wildlife Research Center, Kyoto University, Uto, Kumamoto, Japan; [6]Kyoto City Zoo, Okazaki Hoshojicho, Sakyo-ku, Kyoto, Japan; [7]Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan; [8]Division of Population Genetics, National Institute of Genetics, Mishima, Japan and [9]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo, Japan
Correspondence: Dr K Ichiyanagi or Professor H Sasaki, Division of Epigenomics and Development, Medical Institute of Bioregulation, and Epigenome Network Research Center, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.
E-mail: ichiyanagi@bioreg.kyushu-u.ac.jp or hsasaki@bioreg.kyushu-u.ac.jp
Received 13 March 2013; revised 26 April 2013; accepted 1 May 2013; published online 6 June 2013

between humans and chimpanzees in 22 selected promoters. Farcas *et al.*[22] identified inter- and intraspecific variations in methylation in the *CCRK* promoter but observed no correlation between methylation and expression. Pai *et al.*[23] performed a more comprehensive study on approximately 7700 genes in humans and chimpanzees, reporting that 12–18% of the genes showing expression differences had methylation differences. Martin *et al.*[24] performed methylation-sensitive restriction endonuclease digestion and deep sequencing and identified approximately 1500 CGIs differentially methylated in humans and chimpanzees. The study also showed that some of the methylation differences occurred in the germline. However, all these studies focused only on gene promoters and CGIs. More recently, Molaro *et al.*[25] compared the whole-genome methylation patterns of sperm from humans and chimpanzees at nucleotide resolution; however, the study did not allow investigation of the relationship between methylation and transcription.

In this study, we performed methylated DNA immunoprecipitation followed by tiling array hybridization (MeDIP-chip) to scan the entire lengths of chromosomes 21 and 22 for regional methylation differences between humans and chimpanzees. These two chromosomes were selected because of their small chromosome size and the availability of high-density tiling microarray. The two species have highly homologous chromosomes, and the only major difference is that the genes on human chromosome 2 are encoded by chimpanzee chromosomes 2a and 2b.[26,27] Our MeDIP-chip analysis on peripheral blood leukocytes (PBLs) and subsequent validation by bisulphite sequencing identified 16 genomic regions showing methylation differences between the two species. We termed these regions S-DMRs for species-specific differentially methylated regions. Many of these S-DMRs were associated with genes, notably with disease-related genes, and that the methylation differences were often correlated with expression differences. We discuss the possible involvement of epigenetic differences in diversification of phenotype.

MATERIALS AND METHODS

Sample preparation
Blood samples (5–10 ml) were obtained from female volunteers (27–45 years old (yo)), female and male chimpanzees (10–44 yo, Primate Research Institute, Kyoto University, Kyoto, USA), a female gorilla (24 yo, Kyoto City Zoo) and a male orangutan (30 yo, Kyoto City Zoo). Research protocols and consent forms were reviewed and approved by the ethical committee of Kyushu University. We confirmed that differential white blood cell counts were similar. PBLs in the buffy coat were mixed with acid citrate dextrose and collected by centrifugation at $3000 \times g$ for 15 min at 25 °C. Frozen human tissues (cerebrum, 28-yo female; spleen, 37-yo female; pancreas, 40-yo female; testis, 24-yo male) were purchased from Wako Pure Chemical Industries (Osaka, Japan), and sperm was obtained from three volunteers (24–27 yo). Frozen tissues were obtained from a male chimpanzee (38 yo, Kumamoto Sanctuary) through Great Ape Information Network. Sperm was obtained from 11–22-yo male chimpanzees (Kumamoto Sanctuary). High-molecular weight genomic DNA was isolated by a standard procedure.[28] Total RNA was extracted from tissues and cells by an acid guanidium phenol–chloroform method using Isogen (Nippon Gene, Tokyo, Japan).

HpaII-McrBC PCR
The HpaII-McrBC PCR assay was performed as described previously.[29] In brief, genomic DNAs were digested either with HpaII, HhaI, MspI (TaKaRabio, Otsu, Japan) or McrBC (New England Biolabs, Ipswich, MA, USA) overnight at 37 °C. Following ethanol precipitation, DNAs were dissolved in TE (Tris-EDTA). Digested DNA of 50 ng was used for each PCR assay. Thermal cycling parameters and primer sequences were identical to those used in the previous study.[29]

MeDIP-chip
A mixture of genomic DNAs (a total of 5 μg) from four female humans or four female chimpanzees was fragmented by sonication into a 0.3- to 1-kb size range, denatured at 95 °C for 10 min and cooled on ice. The sample was incubated with 5 μg of anti-5-methylcytosine antibody (Aviva Systems Biology, San Diego, CA, USA) in 200 μl of IP buffer (10 mM sodium phosphate, 140 mM NaCl and 0.05% Triton-X100, pH 7.0) for 3.5 h at 4 °C. The antibody-bound DNA fragments were magnetically collected after incubation with Dynabeads M-280 (Life Technologies, Grand Island, NY, USA) coated with anti-mouse IgG antibody for 3 h at 4 °C. The beads were washed five times with 700 μl of IP buffer and resuspended in 250 μl of lysis buffer (50 mM TrisHCl, pH 8.0, 10 mM EDTA and 0.5% sodium dodecyl sulfate) containing proteinase K. After incubation at 55 °C for 30 min, DNA was recovered by phenol–chloroform extraction and ethanol precipitation. The DNA was amplified using WGA2 (Sigma Aldrich, St Louis, MO, USA) to give a total of approximately 10 μg DNA and then labelled with biotin-dUTP using GeneChip wild-type double-stranded DNA terminal labeling kit (Affymetix, Santa Clara, CA, USA). The labeled DNA was hybridized against a GeneChip Chromosome 21/22 2.0 R Array (Affymetrix) at 45 °C for 16 h. The array was washed and stained with Cy3-streptavidin in Fluidics Station 450 (Affymetrix) using GeneChip Hybridization, Wash and Stain kit (Affymetrix). The fluorescent signal intensity at each probe spot was measured on a GeneChip Scanner (Affymetrix). As a control, total genomic DNA was labeled and hybridized. Both the MeDIP and control data were obtained from three technical replicates. The data have been deposited in Gene Expression Omnibus (GSE35475).

A log2-transformed ratio of signals obtained from MeDIP and control samples were calculated for each probe using Tiling Analysis Software (TAS, Affymetrix). The data set was further processed using Bayesian Tool for Methylation Analysis (BATMAN),[30] which gave methylation scores in 100-bp windows. For this analysis, only data from probes with a unique sequence common to both humans (UCSC genome sequence version hg17) and chimpanzees (UCSC genome sequence version panTro2) were used. To process chimpanzee MeDIP-chip data using BATMAN, the genomic position of each probe in the human genome was converted to an orthologous position in the chimpanzee genome. The raw BATMAN scores were normalized (median = 0, s.d. = 1) for comparison. An S-DMR candidate was (1) a region of ≥300 bp (2) with a CpG density of ≥0.025 and (3) with an interspecific difference in BATMAN score >2.0. Furthermore, the average BATMAN score in one species was required to be >0.5 and that in the other less than −0.5.

Bisulphite sequencing
Genomic DNA was subjected to bisulphite-mediated C to U conversion using BisulFast (Toyobo, Osaka, Japan) and then used as a template for PCR with ExTaq HS (Takarabio). Amplification was performed by touchdown PCR as described previously.[31] The PCR products purified from an agarose gel were cloned into pGEM-Teasy vector (Promega, Madison, WI, USA) and at least 7 (typically 10–20) clones were sequenced for each locus. The primer sequences are listed in Supplementary Tables S1, S4, S5 and S6.

MassARRAY analysis
MassARRAY was performed on an EpiTYPER platform (Sequenom, San Diego, CA, USA). Primers for bisulphite-converted DNA were designed using EpiDesigner (Sequenom) (Supplementary Tables S2 and S3). Bisulphite conversion of DNA was performed as above. After PCR amplification, shrimp alkaline phosphatase treatment, reverse transcription and sequence-specific RNA cleavage (MassCLEAVE, Sequenom) were performed as described.[32] Cleavage products were loaded onto silicon chips (spectroCHIP, Sequenom) and analyzed by MALDI TOF mass spectrometry (MassARRAY Compact MALDI TOF, Sequenom). Methylation levels were obtained from mass spectra using the EpiTYPER software v1.0.5 (Sequenom).

Quantification of mRNA
cDNA was synthesized from total RNA of 1 μg with random hexamers as primers using a PrimeScript first-strand cDNA synthesis kit (Takarabio). Quantitative real-time PCR was performed with SYBR premix ExTaq II

reagents (Takarabio) and analyzed on a Thermal Cycler Dice Real Time System (Takarabio). The *ACTB* mRNA level was used as an internal control. The difference in relative mRNA level was calculated by the $\Delta\Delta$Ct method.[33] The primer sequences are listed in Supplementary Table S8.

Statistical analysis and cluster analysis

Interspecific differences in DNA methylation and mRNA levels were statistically evaluated using the nonparametric Mann–Whitney U-test at the following web site:

http://www.gen-info.osaka-u.ac.jp/testdocs/tomocom/wilc1-e.html.

For hierarchical cluster analysis to construct a phylogenetic tree, we used the R (http://www.r-project.org/) function hclust.

RESULTS

Identification of S-DMRs on chromosomes 21 and 22

We first used the HpaII–McrBC PCR method[29] to compare methylation levels of 146 orthologous CGIs (92 in promoter regions) on chromosome 21 in PBLs from humans and chimpanzees. This analysis revealed that 43 CGIs were methylated and 102 were unmethylated in both the species (data not shown). The methylated CGIs were mostly non-promoter CGIs. However, CGI in the *PDXK* promoter, designated as S-DMR1, showed an interspecific methylation difference, with human sequences being more highly methylated (Table 1 and Supplementary Figure S1).

To identify S-DMRs more comprehensively, we used an Affymetrix genome-tiling array for human chromosomes 21 and 22. Pooled DNAs from PBLs from age-matched females (four from each species) were immunoprecipitated with an anti-5-methylcytosine antibody and hybridized to the array (MeDIP-chip). To avoid hybridization artefacts, we only used data from probes carrying a single-copy sequence common to both the species. Furthermore, differences in CpG density were taken into consideration; we processed the signal-intensity data using BATMAN[30] that gives methylation scores in 100-bp windows. We confirmed that the normalized BATMAN scores (median = 0, s.d. = 1) were well correlated with the methylation levels of 78 loci measured by bisulphite sequencing (Supplementary Figure S2).

When the average BATMAN scores for all 100-kb regions were plotted along the chromosome arms, very similar overall methylation patterns were observed in the two species (Figure 1a). Even in much smaller windows (1 kb, with $\geqslant 5$ BATMAN scores), average scores were highly similar between the species (Pearson's $r = 0.81$). It has been reported that changes in DNA methylation occur predominantly in CGI shores rather than within CGIs during cell differentiation, reprogramming and carcinogenesis.[18,19] This trend was observed in our interspecific comparison: the differences in BATMAN score were larger in shores than in CGIs for 1069 orthologous CGI regions (Figure 1b). To identify S-DMRs, we selected regions of $\geqslant 300$ bp that showed a high degree of interspecific difference in BATMAN score (for criteria, see Materials and Methods section). Using bisulphite sequencing and MassARRAY analyses, significant interspecific methylation differences ($> 30\%$) were confirmed at 15 candidate regions in four individuals from each species ($P < 0.05$, Mann–Whitney U-test) (Figure 1c). Thus, 16 S-DMRs (including S-DMR 1) were identified (Figure 1c and Table 1). Twelve of the 16 S-DMRs were associated with genes and four were intergenic. Of the twelve gene-associated S-DMRs, four were located in promoters, four in exons and the remaining four in introns. Two of the gene-associated S-DMRs were located in CGIs and three in CGI shores. The methylation statuses of these S-DMRs were investigated in other tissues (cerebrum, spleen, pancreas, testis and sperm) from several individuals using MassARRAY. With one exception (spleen, of which the major cell component is lymphocyte), the methylation levels of these S-DMRs were similar between the two species (Supplementary Figure S3), suggesting that the differences observed in PBLs and spleen arise late in development.

Correlation of methylation status and sequence at S-DMRs in the four great ape species

We next attempted to gain insight into the evolution of the S-DMRs. We thus extended the analysis to western gorilla (*Gorilla gorilla*) and Bornean orangutan (*Pongo pygmaeus*) and examined whose (human or chimpanzee) methylation status is unique (or common) among the species. We found six human-specific S-DMRs, with four being more methylated and two being less methylated only in humans, and one

Table 1 S-DMRs identified in this study

S-DMR ID	Methylated species	Location in hg19[a]	Genomic feature	
S-DMR 1[b]	Human	chr21:45138212-45138399	Promoter (*PDXK*)	CGI
S-DMR 2	Human	chr21:17443463-17443762	Promoter (*C21orf34*)	
S-DMR 3	Human	chr22:27179656-27180155	Intergenic	
S-DMR 4	Human	chr21:45846606-45847105	Exon (*TRPM2*)	CGI shore
S-DMR 5	Human	chr21:39629164-39629563	Promoter (*KCNJ15*[c])	
S-DMR 6	Human	chr21:36060464-36060863	Intron (*CLIC6*)	
S-DMR 7	Human	chr22:28198756-28199155	Promoter (*MN1*[c])	CGI shore
S-DMR 8	Human	chr22:27040656-27041255	Intergenic	
S-DMR 9	Human	chr21:27372163-27372662	Exon (*APP*[c])	
S-DMR 10	Chimpanzee	chr21:46898606-46899005	Intron (*COL18A1*[c])	CGI
S-DMR 11	Chimpanzee	chr22:46872891-46873190	Intron (*CELSR1*)	
S-DMR 12	Chimpanzee	chr22:40405710-40406409	Exon (*FAM83F*)	
S-DMR 13	Chimpanzee	chr22:39626110-39626509	Exon (*PDGFB*[c])	CGI shore
S-DMR 14	Chimpanzee	chr22:23296856-23297155	Intergenic	
S-DMR 15	Chimpanzee	chr21:46274606-46274905	Intron (*PTTG1P*)	
S-DMR 16	Chimpanzee	chr21:44201365-44201664	Intergenic	

Abbreviations: CGI, CpG island; S-DMR, species-specific differentially methylated region.
[a]The extent (size) of S-DMR 1 (188 bp) was determined by bisulphite sequencing while that of the others was determined only roughly based on BATMAN (Bayesian Tool for Methylation Analysis) scores in 100-bp windows. Thus the extents of S-DMRs 2–16 are ($100 \times n$) bp, where n is $\geqslant 3$ (300 bp).
[b]This S-DMR was identified by the HapII-McrBC assay and the others by the MeDIP-chip assay.
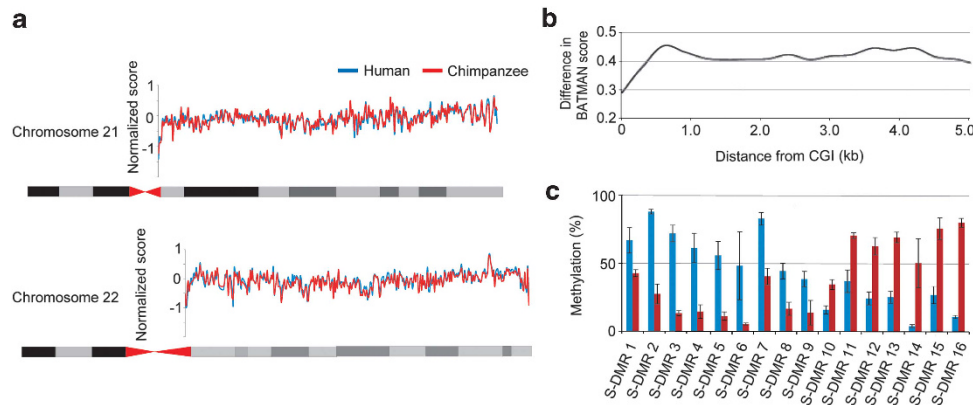[c]Genes implicated in human diseases.

Figure 1 Overall similarities and regional differences in DNA methylation between humans and chimpanzees revealed by MeDIP-chip (methylated DNA immunoprecipitation). (**a**) The average Bayesian Tool for Methylation Analysis (BATMAN) score for each 100-kb window is plotted along chromosomes 21 and 22 in humans (blue) and chimpanzees (red). The genomic sequences of the short arms are not available. The chromosome banding pattern and the centromere (red) are shown for each chromosome. (**b**) The interspecific difference in BATMAN score is plotted against the distance from the boundaries of 1069 orthologous CpG islands (CGIs). The value at distance 0 represents the average difference in BATMAN score in the CGIs. (**c**) The DNA methylation levels of the 16 S-DMRs (species-specific differentially methylated regions) are shown for each species. The methylation levels were measured by MassARRAY assays (S-DMRs 1–8 and 10–16) or by bisulphite sequencing (S-DMR 9) in four humans (blue) and four chimpanzees (red). The interspecific difference in average methylation level is significant for all S-DMRs ($P < 0.05$, Mann–Whitney U-test). Error bars represent s.d.
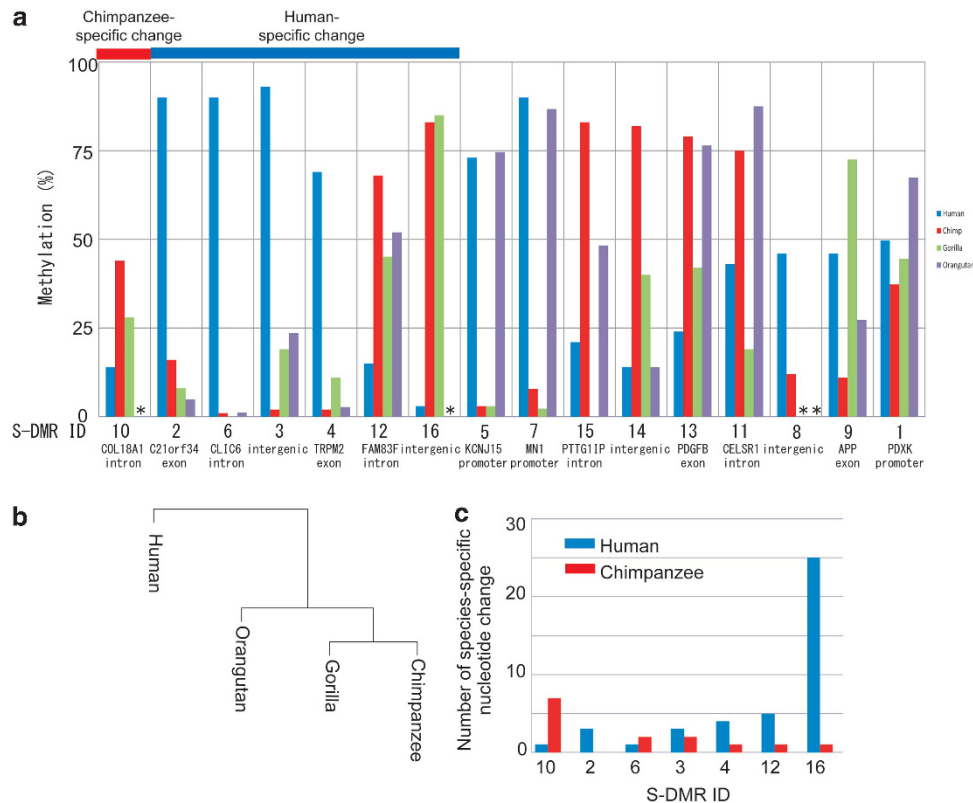

Figure 2 DNA methylation statuses of the S-DMRs (species-specific differentially methylated regions) in four great ape species. (**a**) The DNA methylation levels of the 16 S-DMRs in peripheral blood leukocytes from humans (blue), chimpanzees (red), gorillas (green) and orangutans (purple) are shown. The S-DMRs with human-specific methylation changes and those with chimpanzee-specific methylation changes are indicated at the top. Asterisks mark the S-DMRs that could not be analyzed because of the species-specific deletion or insertion. (**b**) A phylogenetic tree based on the methylation data is shown. A hierarchical cluster analysis was performed with the methylation data from the 16 S-DMRs based on the neighbor-joining method using the R function hclust. The length of each vertical branch is proportional to the degree of difference in DNA methylation. (**c**) The numbers of human-specific (blue) and chimpanzee-specific (red) nucleotide changes identified by sequence alignment are shown for chimpanzee-specific (S-DMR 10) and human-specific S-DMRs (S-DMRs 2, 6, 3, 4, 12 and 16).

chimpanzee-specific S-DMR, which is more methylated only in chimpanzees (Figure 2a). As shown by the phylogenetic tree constructed by the hierarchical clustering method using the methylation data from all S-DMRs (Figure 2b), the overall methylation status of chimpanzees was closer to that of gorillas and orangutans than that of humans. These results suggest that the methylation status changed

more rapidly in the human lineage than in the chimpanzee lineage, consistent with the results of Martin *et al*.[24]

Interspecific differences in DNA methylation may be purely epigenetic or may arise secondarily to differences in DNA sequence, as reported for the intraspecific methylation differences.[34–36] We compared the sequences of the one chimpanzee-specific S-DMR and six human-specific S-DMRs between the four great ape species. Among the six human-specific S-DMRs, five carried more human-specific sequence changes than chimpanzee-specific sequence changes and the chimpanzee-specific S-DMR carried more chimpanzee-specific sequence changes (Figure 2c). Thus there is a good correlation between the methylation changes and sequence changes, suggesting that the S-DMRs may arise from genetic changes.

Sequence changes at CTCF-binding motifs in S-DMRs

S-DMR 6, which is located 20 kb downstream of the transcription start site of *CLIC6* (Figure 3a), is a human-specific S-DMR, which is more methylated in humans (Figures 1c and 3a). It carried one human-specific sequence change (G to C) (Figures 2c and 3a) and two chimpanzee-specific sequence changes (Figure 2c). Interestingly, the human-specific sequence change site was polymorphic even in the human population (rs2236610), with some individuals carrying G, a nucleotide found in chimpanzees, gorillas and orangutans. Among the 17 human PBL samples that we examined, four were heterozygous for this polymorphism and only the allele carrying C at this site showed heavy methylation (Figure 3b). This allelic bias in methylation was also observed in the testis and cerebrum from additional heterozygous humans (Figures 3c and d). These results strongly suggest that the methylation level at S-DMR 6 is determined by the sequence of the region (*cis*-element) rather than by a *trans*-acting factor(s). The interspecific methylation difference in S-DMR 6 was also observed in sperm and some somatic tissues (Figure 3e), suggesting that the difference is germline-established and ubiquitous. Based on the published results of chromatin immunoprecipitation
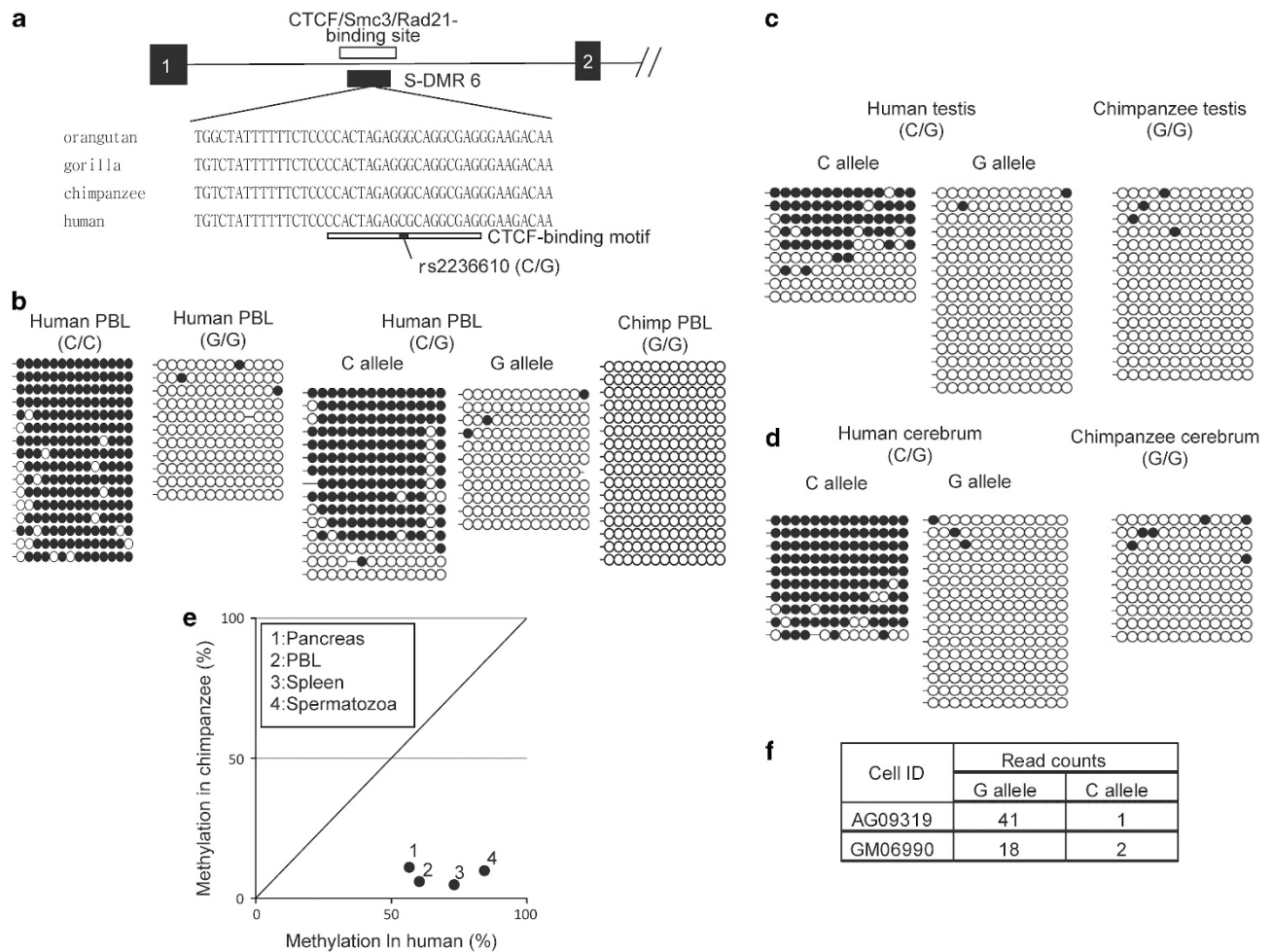

Figure 3 CTCF binding and S-DMR (species-specific differentially methylated region) 6. (**a**) The locations of S-DMR 6 (filled rectangle) and the CTCF-binding site (open rectangle) determined by ChIP-seq (chromatin immunoprecipitation followed by sequencing; Birney *et al*.[37]) are shown with the *CLIC6* structure. Exons are shown as filled rectangles and numbered. The human-specific change (G to C) is located in a CTCF-binding motif (underlined). This site is polymorphic (C/G) in the human population (rs2236610). (**b**) The results of bisulphite sequencing are shown for human peripheral blood leukocytes (PBLs) of different genotypes. The sample from a homozygous chimpanzee (G/G) is an unmethylated control. Open and closed circles represent unmethylated and methylated CpG sites, respectively. (**c, d**) The results of bisulphite sequencing are shown for testis (**c**) and cerebrum (**d**) of heterozygous humans (C/G). Data from the same tissues of homozygous chimpanzees (G/G) are also shown. (**e**) The methylation levels measured in pancreas, PBLs, spleen and sperm from homozygous humans (C/C) are plotted against those measured in the same tissues from chimpanzees (G/G) are plotted. (**f**) The CTCF ChIP-seq data for a human fibroblast cell line (AG09319) and a human lymphoblastoid cell line (GM06990) are used to count the reads mapped to the respective alleles.

followed by sequencing (ChIP-seq) in cultured human cells,[37] this region binds CTCF, a so-called insulator protein that regulates long-range genomic interactions as well as transcriptional elongation, alternative splicing and DNA methylation.[38,39] Interestingly, the polymorphic site corresponds to a highly conserved nucleotide of the CTCF-binding motif[40] (Figure 3a). Based on the calculation by Regulatory Sequence Analysis Tools,[41] the sequence with C at the polymorphic site was less likely to be a CTCF-binding site ($P = 2.7 \times 10^{-5}$) than the sequence with G ($P = 4.6 - 10^{-7}$). Indeed, the above mentioned CTCF ChIP-seq data from the heterozygous human cell lines shows that, of the 62 reads covering this polymorphic site, 59 were from the G allele, whereas only three were from the C allele (Figure 3f). Because loss of CTCF binding can cause a gain in DNA methylation,[42] this observation could explain both inter- and intraspecific differences in methylation. The same mechanism may account for the methylation changes in S-DMR 14 (chimpanzee-specific loss of a CTCF-binding motif by a 2-bp deletion) and S-DMR 16 (human-specific gain of a CTCF-binding motif by two nucleotide changes) (Supplementary Figure S4). Taken together, these results shed light on the role of CTCF-binding sites in the generation of inter- and intraspecific variations in regional DNA methylation beyond the binding sites.
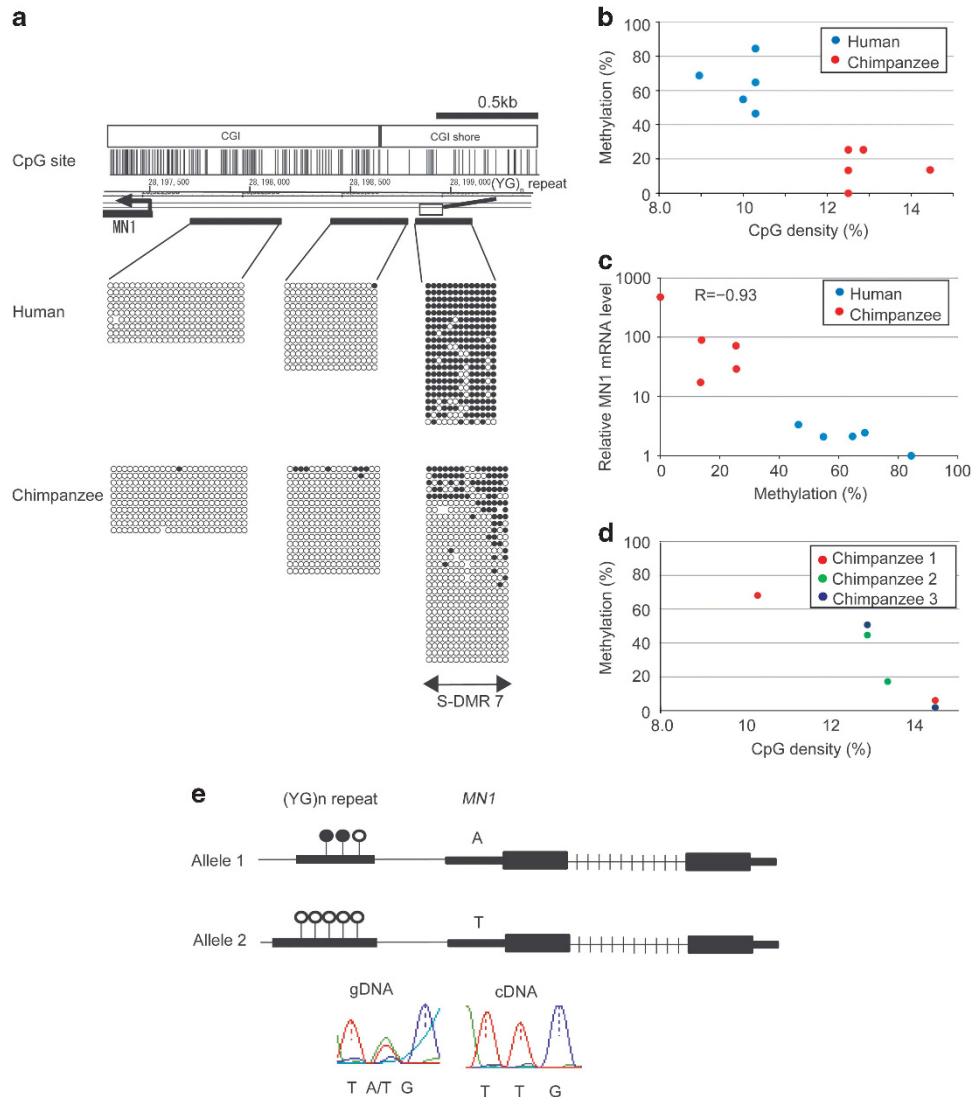

Figure 4 Genetic, epigenetic and transcriptional differences of *MN1*. (a) The structure of the *MN1* promoter region is shown with the results of bisulphite sequencing in representative human and chimpanzee peripheral blood leukocyte samples. The CpG island (CGI) and CGI shore are indicated at the top. The locations of the (YG)ₙ repeat and CpG sites are also shown. The transcription start site and direction of *MN1* transcription are indicated by an arrow. (b) The DNA methylation level of the (YG)ₙ repeat measured by bisulphite sequencing is plotted against the CpG density of the repeat in humans (blue) and chimpanzees (red). (c) The *MN1* mRNA level (normalized by *ACTB*) is plotted against the DNA methylation level of the (YG)ₙ repeat in humans (blue) and chimpanzees (red). The Pearson's correlation coefficient (*r*) between the mRNA (log-scaled) and methylation levels are also shown. (d) The DNA methylation level of the (YG)ₙ repeat is plotted against the CpG density of the repeat in each allele of three heterozygous chimpanzees. The two alleles from the same individual are shown in the same colour. (e) The result obtained in a doubly heterozygous chimpanzee is shown. In this individual, allele 1 carries seven CpGs in the (YG)ₙ repeat and an A at the polymorphic site in the 5′ untranslated region (UTR). Allele 2 has 13 CpGs in the repeat and a T in the 5′ UTR. Open and filled circles represent unmethylated and methylated CpG sites, respectively. The results of direct sequencing of the PCR products from genomic DNA (gDNA, left) and cDNA (right) are shown at the bottom. The sequences of (YG)ₙ repeat and 5′ UTR were determined by sequencing of PCR products of the respective regions. The primer sequence is listed in Supplementary Table S7.

DNA methylation differences correlated with gene expression divergence

We wished to know whether the genes associated with the S-DMRs show interspecific transcriptional differences. Among the five genes that we examined (four with a promoter S-DMR and one with an exonic S-DMR), *MN1* (encoding a transcription co-activator) and *APP* (encoding the amyloid βA4 precursor protein) showed differences in gene expression in PBLs (Figures 4c and 5d).

S-DMR 7, located in the CGI shore of the *MN1* promoter (1.3 kb upstream of the transcription start site), was more methylated in humans than in chimpanzees (Figures 1c, 2a, 4a and 4b), and this difference was associated with an approximately 30-fold decrease in mRNA level in human PBLs (Figure 4c). The S-DMR contained a $(YG)_n$ repeat $(Y = C$ or $T)$ and showed a lower CpG density in humans than in chimpanzees. Interestingly, the CpG density varied even within each species and was roughly inversely correlated with the methylation level (Figure 4b). The methylation level of the S-DMR was then inversely correlated with the *MN1* expression level (Figure 4c, Pearson's $r = -0.93$). To assess the role of the $(YG)_n$ repeat more directly, we examined the allelic methylation differences in three chimpanzees heterozygous for the polymorphism. In all the cases, the allele with a lower CpG density was more methylated than the other allele (Figure 4d). In one of these chimpanzees, we identified an additional single-nucleotide change in the transcribed region, which revealed that *MN1* mRNA is derived almost exclusively from the less methylated allele (Figure 4e). These results indicate that the sequence and epigenetic status of the microsatellite repeat are the major determinants of the variations in *MN1* expression.

S-DMR 9, located in exon 7 of *APP* (Figure 5a), was more highly methylated in humans than in chimpanzees (Figure 5b). Because exon 7 is one of the alternatively spliced exons, we measured the levels of *APP* mRNA isoforms with and without exon 7 sequence. Although the overall *APP* mRNA level did not differ in PBLs between the species (Figure 5c), the relative level of the exon 7-containing mRNA isoform was significantly higher in humans than in chimpanzees ($P < 0.05$ by Mann–Whitney *U*-test; Figure 5d). Thus, a higher methylation of exon 7 was associated with a more frequent usage of this exon in these species. The results are consistent with the recent report that alternative splicing patterns are regulated by DNA methylation.[43]

S-DMRs and disease genes

It is noteworthy that the above mentioned genes, *MN1* and *APP*, are involved in human diseases: *MN1* in meningioma and acute myeloid leukemia[44] and *APP* in Alzheimer's disease and cerebral amyloid angiopathy.[45] Using Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim), we found that the following genes have a relevance to diseases: *KCNJ15* is implicated in the susceptibility to type II diabetes mellitus;[46] *COL18A1* is responsible for Knobloch syndrome;[47] and *PDGFB* is responsible for dermatofibrosarcoma protuberans.[48] Thus, among the 12 S-DMR-associated genes identified in this study, 5 were associated with human diseases in some way.

DISCUSSION

Profiling of DNA methylation in PBLs by MeDIP-chip and subsequent confirmation by bisulfite sequencing and MassARRAY identified 16 regional methylation differences between humans and chimpanzees on chromosomes 21 and 22. These S-DMRs spanned ⩾300 bp and were frequently associated with genes (12 of the 16 DMRs). Interestingly, many (5 of the 12) of the genes are involved in or have some relevance to human diseases. Because chromosomes 21 and 22 comprise approximately 1.3% of the entire genome, we estimate the total number of S-DMRs to be >1200.

In two of the genes associated with an S-DMR, we observed transcriptional differences correlated with the methylation differences.
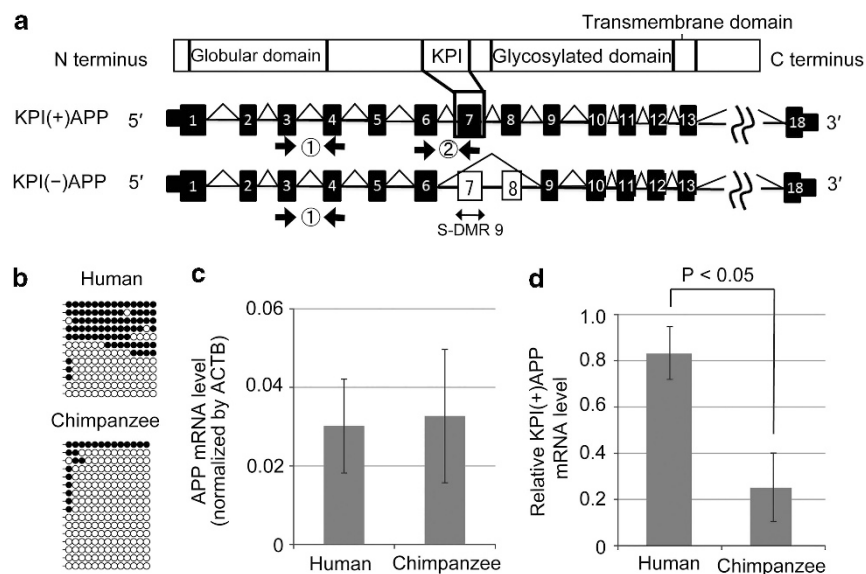

Figure 5 Alternative splicing of *APP* (amyloid βA4 precursor protein) mRNA and S-DMR (species-specific differentially methylated region) 9. (**a**) The overall *APP* protein structure and exon–intron organization are shown. KPI represents the Kunitz-type protease inhibitor domain encoded by exon 7. KPI(+)APP and KPI(−)APP represent the mRNAs with and without exon 7, respectively. S-DMR 9 corresponds to exon 7. The arrows represent PCR primers for total *APP* mRNA and KPI(+) mRNA. (**b**) The results of bisulphite sequencing of S-DMR 9 in representative human and chimpanzee peripheral blood leukocyte (PBL) samples. (**c**) The levels of total *APP* mRNA (normalized by *ACTB*) measured in four human and four chimpanzee PBL samples are shown. Error bars indicate s.d. (**d**) The levels of KPI(+) mRNA relative to those of total *APP* mRNA in four human and four chimpanzee PBL samples. The interspecific difference was evaluated by Mann–Whitney *U*-test.

Expression of *MN1* was inversely correlated with the methylation level of the S-DMR located in the promoter region. In addition to its role in diseases, *MN1* is required for appropriate osteoblast proliferation, motility, differentiation and function.[49] Furthermore, mouse *Mn1* has been shown to be important for the development of membranous bones of the cranial skeleton.[50] It is conceivable that *MN1* has a role in the structural difference of the cranium between humans and chimpanzees.[51] The other gene *APP* showed a difference in splicing pattern that was correlated with the methylation status of the exonic S-DMR. The higher methylation in humans was associated with more frequent usage of this exon in the mRNA, consistent with the recent findings.[43,52] Interestingly, the exon encodes the KPI domain implicated in senile plaque formation in Alzheimer's disease.[53] Given that amyloid β peptide in the blood stream is implicated in Alzheimer's disease,[54] it is tempting to speculate that the interspecific splicing difference in PBLs provides an explanation for the absence of this disease in chimpanzees.[55]

The DNA methylation pattern changes dynamically during development, when each cell lineage acquires its own methylation pattern.[14,56] We found that, with only one exception, almost all interspecific differences identified in PBLs are tissue specific; other tissues, including sperm, showed no significant difference between humans and chimpanzees. The results strongly suggest that most methylation differences arise late in development in a tissue-specific manner. However, using CG decay as a hallmark of methylation in the germline, Martin *et al.*[24] proposed that the methylation differences that they found in human and chimpanzee neutrophils occurred in the germline. As available data are still limited, further studies are needed to know which type of difference is more prevalent.

Regarding the evolutionary aspect of the S-DMRs, our study involving gorillas and orangutans provided important insights. First, as more than half of the S-DMRs showed methylation differences in multiple great ape species, these regions may be highly prone to methylation change. Second, as we identified six human-specific S-DMRs but only one chimpanzee-specific S-DMR, humans appear more prone to methylation change than other great apes, consistent with the previous report (Martin *et al.*[24]). It is unknown whether this applies to other tissues, given that humans show a larger change in gene expression in some tissues but not in others.[57] Third, our findings argue that genetic changes are the major cause of methylation changes. The allelic biases in methylation observed at S-DMRs 6 and 7 in heterozygous individuals were the strong evidence for this notion. Consistently, the methylation differences identified between the laboratory mouse strains are caused mainly by sequence changes.[34,36]

How can relatively small changes in DNA sequence cause regional changes in methylation? We showed that the sequence changes in the CTCF-binding motifs are the likely cause of the interspecific methylation changes found in some S-DMRs. CTCF is an evolutionarily conserved zinc-finger protein that has a central role in chromatin boundary formation and global organization of chromatin architecture.[39] Furthermore, it is known that binding of CTCF to chromatin protects DNA from methylation.[58] Therefore, although CTCF-binding sites are largely conserved in mammals,[59] if there is a change, it can cause a regional methylation change between the species. In addition to CTCF-binding sites, microsatellite repeats containing variable number of CpG sites can be another genetic factor for regional methylation differences. Our findings on S-DMR 7 are consistent with the general principle that regions with high CpG density are less highly methylated than low CpG regions.[60] The wide distribution and high mutability of microsatellites[61] suggest that they could be a frequent cause of interspecific methylation differences. It is known that microsatellite variations in promoters and gene bodies can change gene expression through transcriptional initiation, elongation or RNA splicing.[62–64] In addition to the genetic changes that alter transcription factor-binding sites,[5–7] those that alter epigenetic modification may be important for the divergence of gene expression pattern and phenotype during evolution.

In conclusion, our study revealed regional DNA methylation differences between humans and chimpanzees and showed that many of them arise from small genetic changes. We also showed that some of these differences are correlated with expression differences. Further studies will provide insights into the genetic and epigenetic mechanism of phenotypic divergence during hominid evolution.

1   Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69–87 (2005).
2   Karaman, M. W., Houck, M. L., Chemnick, L. G., Nagpal, S., Chawannakul, D., Sudano, D. *et al.* Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome. Res.* **13,** 1619–1630 (2003).
3   Caceres, M., Lachuer, J., Zapala, M. A., Redmond, J. C., Kudo, L., Geschwind, D. H. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100,** 13030–13035 (2003).
4   Kehrer-Sawatzki, H. & Cooper, D. N. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum. Mutat.* **28,** 99–130 (2007).
5   Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165,** 349–357 (1969).
6   Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3,** e245 (2005).
7   King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116 (1975).
8   Uddin, M., Wildman, D. E., Liu, G., Xu, W., Johnson, R. M., Hof, P. R. *et al.* Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl Acad. Sci. USA* **101,** 2957–2962 (2004).
9   Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4,** e1000271 (2008).
10 Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309,** 1850–1854 (2005).
11 Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome. Res.* **14,** 1462–1473 (2004).
12 Blekhman, R., Oshlack, A. & Gilad, Y. Segmental duplications contribute to gene expression differences between humans and chimpanzees. *Genetics* **182,** 627–630 (2009).
13 Chabot, A., Shrit, R. A., Blekhman, R. & Gilad, Y. Using reporter gene assays to identify cis regulatory differences between humans and chimpanzees. *Genetics* **176,** 2069–2076 (2007).
14 Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
15 Chong, S. & Whitelaw, E. Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* **14,** 692–696 (2004).
16 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
17 Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69,** 915–926 (1992).

18 Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R. *et al*. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41,** 1350–1353 (2009).

19 Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P. *et al*. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41,** 178–186 (2009).

20 Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat. Rev. Genet.* **13,** 153–162 (2012).

21 Enard, W., Fassbender, A., Model, F., Adorjan, P., Paabo, S. & Olek, A. Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* **14,** R148–R149 (2004).

22 Farcas, R., Schneider, E., Frauenknecht, K., Kondova, I., Bontrop, R., Bohl, J. *et al*. Differences in DNA methylation patterns and expression of the CCRK gene in human and nonhuman primate cortices. *Mol. Biol. Evol.* **26,** 1379–1389 (2009).

23 Pai, A. A., Bell, J. T., Marioni, J. C., Pritchard, J. K. & Gilad, Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* **7,** e1001316 (2011).

24 Martin, D., Singer, M., Dhahbi, J., Mao, G., Zhang, L., Schroth, G. *et al*. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. *Genome. Res.* **21,** 4049–4057 (2011).

25 Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J. *et al*. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146,** 1029–1041 (2011).

26 McConkey, E. H. Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. *Cytogenet. Genome. Res.* **105,** 157–158 (2004).

27 Yunis, J. J., Sawyer, J. R. & Dunham, K. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208,** 1145–1148 (1980).

28 Sambrook, J. & Russell, D. W. *Molecular Cloning*. 3rd edn (Cold Spring Harbor Press, Cold Spring Harbor, NY, USA, 2011).

29 Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T. *et al*. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome. Res.* **14,** 247–266 (2004).

30 Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E. *et al*. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* **26,** 779–785 (2008).

31 Ichiyanagi, K., Li, Y., Watanabe, T., Ichiyanagi, T., Fukuda, K., Kitayama, J. *et al*. Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. *Genome. Res.* **21,** 2058–2066 (2011).

32 Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G. *et al*. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl Acad. Sci. USA* **102,** 15785–15790 (2005).

33 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25,** 402–408 (2001).

34 Schilling, E., El Chartouni, C. & Rehli, M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome. Res.* **19,** 2028–2035 (2009).

35 Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20,** 883–889 (2010).

36 Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J. *et al*. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148,** 816–831 (2012).

37 Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H. *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).

38 Millau, J. F. & Gaudreau, L. CTCF, cohesin, and histone variants: connecting the genome. *Biochem. Cell Biol.* **89,** 505–513 (2011).

39 Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137,** 1194–1211 (2009).

40 Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D. *et al*. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128,** 1231–1245 (2007).

41 Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D. *et al*. RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res.* **39,** W86–W91 (2011).

42 Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., Heuchel, R. *et al*. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.* **17,** 586–590 (2003).

43 Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144,** 16–26 (2011).

44 Grosveld, G. C. MN1, a novel player in human AML. *Blood Cells Mol. Dis.* **39,** 336–339 (2007).

45 Wisniewski, T. & Frangione, B. Molecular biology of Alzheimer's amyloid—Dutch variant. *Mol. Neurobiol.* **6,** 75–86 (1992).

46 Okamoto, K., Iwasaki, N., Nishimura, C., Doi, K., Noiri, E., Nakamura, S. *et al*. Identification of KCNJ15 as a susceptibility gene in Asian patients with type 2 diabetes mellitus. *Am. J. Hum. Genet.* **86,** 54–64 (2010).

47 Sertie, A. L., Sossi, V., Camargo, A. A., Zatz, M., Brahe, C. & Passos-Bueno, M. R. Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum. Mol. Genet.* **9,** 2051–2058 (2000).

48 Simon, M. P., Pedeutour, F., Sirvent, N., Grosgeorge, J., Minoletti, F., Coindre, J. M. *et al*. Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma. *Nat. Genet.* **15,** 95–98 (1997).

49 Zhang, X., Dowd, D. R., Moore, M. C., Kranenburg, T. A., Meester-Smoor, M. A., Zwarthoff, E. C. *et al*. Meningioma 1 is required for appropriate osteoblast proliferation, motility, differentiation, and function. *J. Biol. Chem.* **284,** 18174–18183 (2009).

50 Meester-Smoor, M. A., Vermeij, M., van Helmond, M. J., Molijn, A. C., van Wely, K. H., Hekman, A. C. *et al*. Targeted disruption of the Mn1 oncogene results in severe defects in development of membranous bones of the cranial skeleton. *Mol. Cell. Biol.* **25,** 4229–4236 (2005).

51 Cobb, S. N. The facial skeleton of the chimpanzee-human last common ancestor. *J. Anat.* **212,** 469–485 (2008).

52 Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M. *et al*. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479,** 74–79 (2011).

53 Menendez-Gonzalez, M., Perez-Pinera, P., Martinez-Rivera, M., Calatayud, M. T. & Blazquez Menes, B. APP processing and the APP-KPI domain involvement in the amyloid cascade. *Neurodegener. Dis.* **2,** 277–283 (2005).

54 Li, Q. X., Fuller, S. J., Beyreuther, K. & Masters, C. L. The amyloid precursor protein of Alzheimer disease in human brain and blood. *J. Leukoc. Biol.* **66,** 567–574 (1999).

55 Gearing, M., Rebeck, G. W., Hyman, B. T., Tigges, J. & Mirra, S. S. Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. *Proc. Natl Acad. Sci. USA* **91,** 9382–9386 (1994).

56 Rakyan, V. K., Down, T. A., Thorne, N. P., Flicek, P., Kulesha, E., Graf, S. *et al*. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* **18,** 1518–1529 (2008).

57 Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P. *et al*. The evolution of gene expression levels in mammalian organs. *Nature* **478,** 343–348 (2011).

58 Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A. *et al*. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).

59 Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Goncalves, A., Kutter, C. *et al*. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148,** 335–348 (2012).

60 Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A. *et al*. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).

61 Eckert, K. A. & Hile, S. E. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol. Carcinog.* **48,** 379–388 (2009).

62 Rockman, M. V. & Wray, G. A. Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* **19,** 1991–2004 (2002).

63 Li, Y. C., Korol, A. B., Fahima, T. & Nevo, E. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* **21,** 991–1007 (2004).

64 Kashi, Y. & King, D. G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22,** 253–259 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)