

DNA methylation profiling of human chromosomes 6, 20 and 22

Florian Eckhardt¹, Joern Lewin¹, Rene Cortese¹, Vardhman K Rakyan², John Attwood², Matthias Burger¹, John Burton², Tony V Cox², Rob Davies², Thomas A Down², Carolina Haefliger¹, Roger Horton², Kevin Howe², David K Jackson², Jan Kunde^{1,3}, Christoph Koenig¹, Jennifer Liddle², David Niblett², Thomas Otto¹, Roger Pettett², Stefanie Seemann¹, Christian Thompson¹, Tony West², Jane Rogers², Alex Olek¹, Kurt Berlin¹ & Stephan Beck²

DNA methylation is the most stable type of epigenetic modification modulating the transcriptional plasticity of mammalian genomes. Using bisulfite DNA sequencing, we report high-resolution methylation profiles of human chromosomes 6, 20 and 22, providing a resource of about 1.9 million CpG methylation values derived from 12 different tissues. Analysis of six annotation categories showed that evolutionarily conserved regions are the predominant sites for differential DNA methylation and that a core region surrounding the transcriptional start site is an informative surrogate for promoter methylation. We find that 17% of the 873 analyzed genes are differentially methylated in their 5' UTRs and that about one-third of the differentially methylated 5' UTRs are inversely correlated with transcription. Despite the fact that our study controlled for factors reported to affect DNA methylation such as sex and age, we did not find any significant attributable effects. Our data suggest DNA methylation to be ontogenetically more stable than previously thought.

The completion of the Human Genome Project^{1,2} has created a basis to study how the genetic blueprint is executed at the cellular level. Many of the processes involved are governed by additional layers of epigenetic information that are not directly encoded by the DNA sequence itself but by chemical modifications of chromatin in form of DNA methylation and histone modifications, collectively also referred to as the 'epigenetic code'. Deciphering the human epigenetic code will be a daunting task, as it is encoded not in one but in many different epigenomes (for review, see refs. 3,4).

Toward this goal, a blueprint for an international human epigenome project (recently dubbed the Alliance for Human Epigenomics and Disease (AHEAD)) has been proposed⁵ that recognizes the need to integrate already ongoing epigenome projects. One of these projects, termed the Human Epigenome Project (HEP), aims to identify, catalog and interpret genome-wide DNA methylation profiles of all human genes in all major tissues⁶. In mammals, DNA methylation occurs almost exclusively within the context of CpG dinucleotides, with an estimated 80% of all CpG sites methylated. Although array-based approaches^{7–9} look promising for the future, bisulfite DNA sequencing¹⁰ remains the gold standard for high-resolution DNA methylation profiling of human epigenome(s)⁶. Using this approach, here we report the methylation profiling of human chromosomes 6, 20 and 22 in 43 samples derived from 12 different (healthy) tissues.

RESULTS

After the HEP pilot study⁶, we sought to establish DNA methylation reference profiles for three human chromosomes from a representative number of healthy human tissues and primary cells (that is, those having no known disease phenotype). We controlled for two parameters, age and sex, that could potentially influence DNA methylation. We analyzed 43 different samples derived from sperm, various primary cell types (dermal fibroblasts, dermal keratinocytes, dermal melanocytes and CD4⁺ and CD8⁺ lymphocytes) and tissues (heart muscle, skeletal muscle, liver and placenta). We pooled tissues from up to three age- and sex-matched individuals (**Supplementary Table 1** online). We cultured primary cells for no more than three passages to minimize the risk of introducing aberrant methylation. Additionally, we compared the methylation levels of selected amplicons before and after culturing and did not detect any difference in average methylation.

We designed amplicons to cover six distinct sequence categories (**Fig. 1**) based on Ensembl annotation (National Center for Biotechnology Information (NCBI) build 34). We did not include CpG islands (CGIs) as a separate category because they were present in multiple categories but analyzed them separately where indicated. In total, we analyzed 2,524 amplicons on chromosomes 6, 20 and 22 (**Table 1**) comprising coding, noncoding and evolutionarily conserved sequences that are associated with 873 genes. Taking the number of

¹Epigenomics AG, Kleine Präsidentenstrasse 1, 10178 Berlin, Germany. ²Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ³Present address: Schering AG, Müllerstr. 178, 13342 Berlin, Germany. Correspondence should be addressed to F.E. (florian.eckhardt@epigenomics.com) or S.B. (beck@sanger.ac.uk).

Received 26 April; accepted 18 September; published online 29 October 2006; doi:10.1038/ng1909

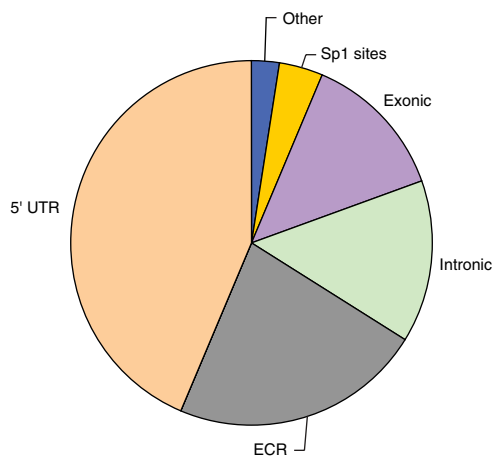


Figure 1 Type and distribution of amplicons. In total, we analyzed 2,524 amplicons from six distinct categories: 43.7% 5'-UTRs, 22.5% evolutionary conserved regions (ECR), 14.3% intronic regions, 13.3% exonic regions, 3.6% Sp1 transcription factor binding sites and 2.6% 'other'. Details of the selection criteria for each category are described in Methods.

biological (**Supplementary Table 1**) and technical (see Methods) replicates into account, we determined the methylation status of 1.88 million CpG sites. The corresponding data have been deposited into the public HEP database and can be accessed at <http://www.epigenome.org>. **Supplementary Figure 1** online shows a global view of the averaged methylation profiles of each tissue type for chromosomes 6, 20 and 22, and **Figure 2** shows a representative 1-Mb region on chromosome 22, illustrating short- and long-range amplicon coverage within the context of gene and CpG island annotation.

Distribution of methylation

In agreement with the results of the recently reported pilot study⁶, the majority of amplicons essentially showed a bimodal distribution: 27.4% of loci were unmethylated (<20% methylation), 42.4% hypermethylated (>80% methylation) and 30.2% heterogeneously methylated (20%–80% methylation). In agreement with previous studies (refs. 11–13), most of the CGIs were unmethylated (**Supplementary Fig. 2** online), and only a small fraction (9.2%) of CGIs were hypermethylated. None of the CGIs with CpG densities >10% were hypermethylated. As methylated cytosines are susceptible to spontaneous deamination¹⁴, it is conceivable that this level of CpG density might represent a threshold beyond which the mutagenic burden becomes too high for the (epi)genetic status to be stably maintained.

From the heterogeneously methylated loci, we selected 14 random amplicons and one control amplicon covering the imprinted *GNAS* locus (ref. 15) to determine if the observed heterogeneity was caused by differences between cells (mosaicism) or by parent-of-origin, allelic differences within cells (imprinting). We subcloned these amplicons and sequenced up to 20 clones. We confirmed imprinting for *GNAS* and confirmed mosaicism for the rest. One amplicon worth noting in this context mapped to the 5' UTR of *SLC22A1*, a gene located within the imprinted

cluster of *IGF2R* on chromosome 6 (refs. 16,17), but allele-specific methylation did not segregate with SNP rs1867351 (**Supplementary Fig. 3** online), thus excluding imprinting in this case. Based on this analysis, we conclude that the majority (>90%) of the observed heterogeneous methylation is caused by mosaicism, although we cannot exclude the additional possibility of heterogeneous tissue sampling.

Next, we investigated the relationship between the degree of methylation over distance (comethylation) and the difference in absolute methylation between tissues. Although we were able to establish a significant correlation for comethylation over short distances ($\leq 1,000$ bp), it deteriorated rapidly for distances >2,000 bp (**Fig. 3a**). This finding suggests that under normal circumstances (that is, cases in which disease is not present), the level of local comethylation has a shorter range compared with the long-range domains of homogeneous methylation reported in some disease situations^{18,19}. To assess the absolute differences in methylation between tissues, we carried out pairwise comparisons of all amplicons between tissues (**Fig. 3b**). Sperm clearly stood out, with the highest difference in methylation (up to 20% compared with fibroblasts and 10% compared to liver), whereas related tissues and cell types like CD4⁺ and CD8⁺ lymphocytes showed the lowest differences (~5%), consistent with their more similar gene expression profiles²⁰. This accentuates the extensive reprogramming spermatozooids undergo during gametogenesis.

Promoter methylation

Promoters are key targets for epigenetic modulation, but their exact locations remain unknown for most human genes. Therefore, we analyzed three types of 'promoter proxy' regions, including amplicons representative of the 5' UTR in general and putative transcription start sites (TSSs) and transcription factor Sp1 sites (both also part of the 5' UTR). The 5' UTR amplicons were further subdivided according to CGI content and associated gene type (known gene, new protein coding sequence (new CDS), pseudogene or new transcript), based on the annotation available from the vertebrate genome annotation (Vega) database²¹.

As expected, most (87.9%) of the CGI-containing 5' UTR amplicons were unmethylated (<20% methylated), 2.1% were hypermethylated (>80% methylated) and the remaining 10% showed heterogeneous methylation (20%–80% methylated) (**Supplementary Fig. 4** online). In contrast, almost 50% of the non-CGI containing 5' UTRs were hypermethylated and only a minority (20.2%) were unmethylated (**Supplementary Fig. 4**). When filtered for associated

Table 1 Statistical summary

| | Total | Chromosome 6 | Chromosome 20 | Chromosome 22 |
|-------------------------------------|-----------------|--------------|---------------|---------------|
| CpG islands on chromosome | 2,279 | 1,070 | 662 | 547 |
| CpG islands covered | 511 | 256 | 29 | 226 |
| CpG islands percentage covered | 22% | 24% | 4% | 41% |
| Genes covered | 873 | 383 | 89 | 401 |
| Exons covered | 853 | 454 | 23 | 376 |
| Introns covered | 920 | 465 | 118 | 337 |
| Number of tissues analyzed | 12 | | | |
| Number of samples analyzed | 43 | | | |
| Mean length of amplicon \pm s.d. | 411 \pm 77 bp | | | |
| Mean number of CpGs per amplicon | 16 \pm 10.8 | | | |
| Total number of different amplicons | 2,524 | | | |
| Number of CpGs analyzed | 1,885,003 | | | |

gene type, the percentage of unmethylated 5' UTRs was 56% for known genes, 53% for new CDSs and about 12% for new transcripts and pseudogenes (**Supplementary Fig. 4**). Methylation has been implicated before in pseudogene silencing (for example, see ref. 13),

and the methylation observed here for new transcripts indicates a similar fate for this category.

TSSs can be predicted with good specificity²² and offer higher spatial resolution than 5' UTRs. Averaging of the methylation values of

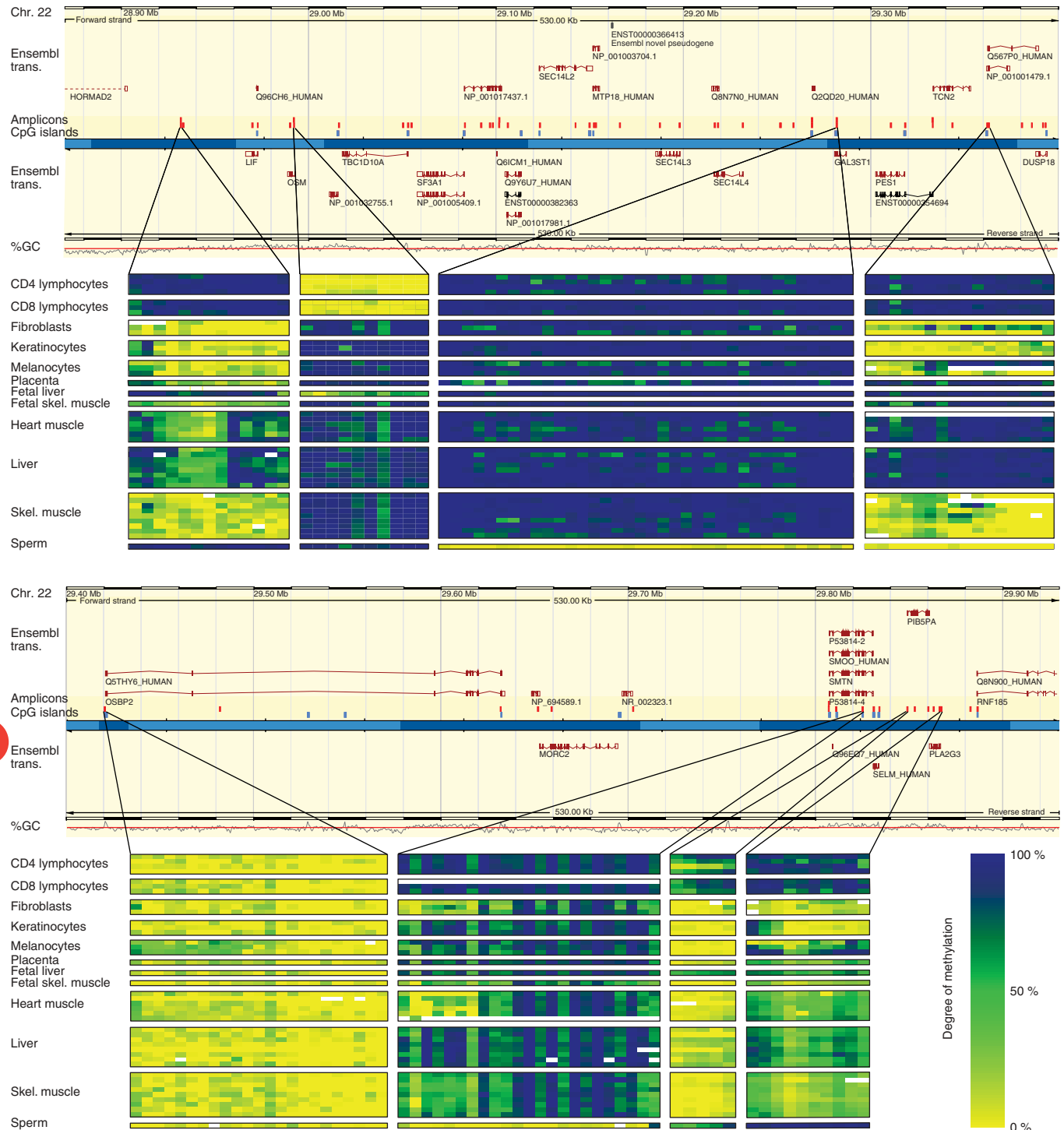


Figure 2 Amplicon coverage in the context of gene and CpG island annotation, as shown for a 1-Mb region on chromosome 22q12.2. Examples of methylation profiles are shown for eight amplicons, including examples of T-DMRs for genes of diverse functions (such as *OSM*, *NP_0010001479.1*, *SMTN* and *RNF185*) and examples of a hypermethylated CpG island (third profile from left) and an unmethylated CpG island (fifth profile from left). Rows represent different samples and are grouped according to tissue or cell type. Columns depict CpG sites, and the corresponding methylation values are indicated by color coding for each cell (blank cells indicate no data).

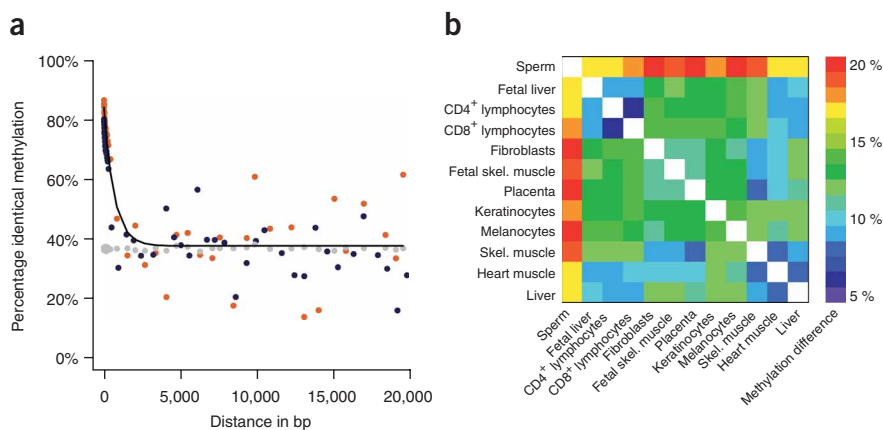


Figure 3 Correlation of DNA methylation with spatial distance and cell type. **(a)** Correlation between comethylation and spatial distance. Orange dots represent CpG methylation values aggregated and averaged over 25,000 individual measurements. Gray dots represent CpG methylation values based on resampling of random CpG positions. Blue dots indicate CpG methylation values based on resampling of amplicon positions. At distances $> 1,000$ bp, we did not detect any correlation between CpG methylation and spatial distance. **(b)** Absolute methylation differences between cell types and tissues. Absolute methylation differences of matched CpGs were determined by pairwise comparison. Differences are color-coded from blue to red indicating a 5%–20% difference in methylation, respectively.

CpGs surrounding TSSs showed an unmethylated core region of about 1,000 bp, extending symmetrically upstream and downstream of the TSS (**Fig. 4**). As unmethylated loci are generally associated with open chromatin structure (reviewed in ref. 23), the methylation status of the identified core region might reflect an open chromatin structure that extends downstream of the TSS.

For the analysis of individual transcription factor binding sites, we selected 94 amplicons containing experimentally verified Sp1 binding sites on chromosome 22 that were previously identified in ref. 24. Of these, 46 were designated TSS-associated (within $\pm 1,000$ bp of a TSS) and 48 non-TSS associated ($> 1,000$ bp away from nearest TSS). Averaging the methylation values for each of the 94 amplicons over all 43 samples showed that 31% were hypermethylated ($> 80\%$ methylated), 25% were heterogeneously methylated (20%–80% methylated) and 44% were unmethylated ($< 20\%$ methylated), indicating that Sp1 binding might be independent of methylation. However, if we filtered amplicons for TSS association, very different ratios of hypermethylated to heterogeneously methylated to unmethylated amplicons emerged: 9%:11%:80% for TSS-associated amplicons, compared with 52%:40%:8% for non-TSS associated amplicons. Similarly, averaging over individual CpG sites showed that 76% of all TSS-associated CpGs were unmethylated compared with only 14% unmethylated, non-TSS associated CpGs (**Fig. 4**). To investigate this further, we correlated amplicon methylation with the presence or absence of a known Sp1 motif (Sp1_Q6) extracted from the TRANSFAC database and found a significant correlation ($P = 0.017$) (that is, amplicons with the 25 highest motif scores are less likely to have high methylation scores). Taken together, these findings bestow highest confidence for Sp1 binding at unmethylated and TSS-associated Sp1 sites but do not exclude the possibility of Sp1 binding at hypermethylated and/or non-TSS associated sites. In some model systems, Sp1 binding has been shown to be abolished by site-specific methylation^{25,26}, whereas in other systems, it seems to be independent of methylation^{27,28}. A direct comparison with the data from ref. 24 is not possible, as that study used cell lines, and therefore, the methylation at the respective loci might be different from the one we have observed in our samples.

Age- and sex-dependent DNA methylation

DNA methylation is influenced by a number of endogenous and exogenous parameters³. Here, we analyzed our data for potential differences associated with age and sex. For a number of different tissues (liver, skeletal muscle and heart muscle), we examined samples obtained from two age groups, one group having a mean

age of 26 ± 4 (s.d.) years and the second group having a mean age of 68 ± 8 (s.d.) years. By averaging the methylation difference of all CpGs analyzed for the two age groups, we identified a mean methylation difference of only 0.275% between these two age groups (**Fig. 5**) and a difference of 0.1% between males and females (**Fig. 5**). These differences are unlikely to be significant, as 10,000-fold resampling of the corresponding data showed similar or larger differences in these random cases (**Fig. 5**). In contrast, by comparing the average methylation between different cell types (**Fig. 5**), we detected highly significant differences between, for example, CD4⁺ lymphocytes and dermal fibroblasts (7.1%) and between skeletal muscle and liver (4.0%).

Although the above analysis of all CpGs has the power to detect global changes in average methylation levels, it might be less suitable to identify specific loci showing a correlation of methylation with age. Therefore, we reanalyzed each amplicon in our data set to identify

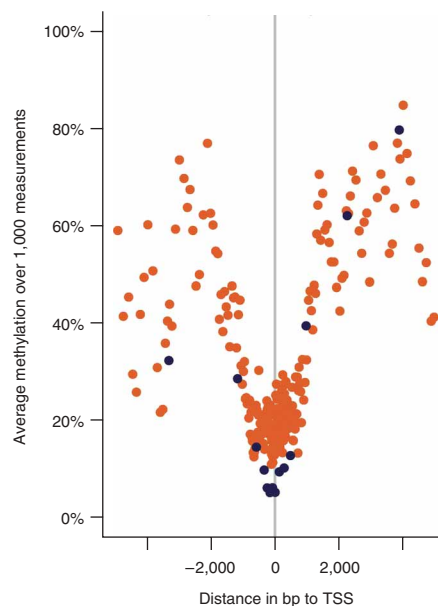
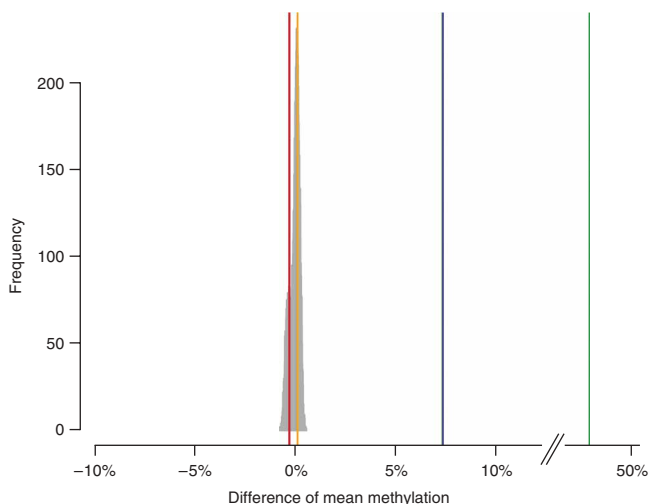


Figure 4 CpG methylation at transcription start sites (TSSs). CpG methylation values were binned (each bin containing 1,000 values), averaged and plotted according to their relative distance to the TSS (orange dots). Blue dots represent bins containing Sp1 sites identified previously in ref. 24. Centered on the TSS, a symmetric core of about 1,000 bp is unmethylated.



age-correlated differential methylation at individual loci. This approach also allowed us to detect differences $< 50\%$, but again, we did not find any statistically significant ($P < 0.05$) differential methylation.

Similarly, we compared samples from the same age group but differing in sex to identify putative non-X chromosomal changes in methylation. Conducting both a global and candidate amplicon analysis, we did not detect any significant methylation changes associated with sex. As a positive control, we confirmed differential 5' UTR methylation of *ELK1*, an X-chromosomal gene that is differentially methylated, showing 50% and 0% methylation in female and male samples, respectively. The absence of both global and locus-specific changes in age- and sex-correlated methylation in our data set suggests that, in healthy individuals, such alterations are limited to specific loci and tissues. One caveat for all age-correlated methylation studies (including ours) is that tissue samples may be inherently more heterogeneous than primary cells because of the different cell types constituting a given tissue, which in turn determines the average level

of DNA methylation. In the present study, we pooled DNA samples in order to minimize errors introduced by heterogeneous tissue sampling. It is conceivable that some tissues (for example, those more exposed to environmental conditions, such as lung and colon) will show a stronger correlation between methylation and age. A recent study performed in monozygotic twins detected epigenetic differences in the overall content and distribution of 5-methylcytosine and histone acetylation that arose in older twins²⁹, and it is possible that age-related methylation alterations might be too subtle to be detectable on a genome-wide scale against a heterogeneous genetic background or might be undetectable because of the method used.

Differential methylation

It is believed that tissue-specific transcription is controlled, in part, by tissue-specific differentially methylated regions (T-DMRs). T-DMRs are likely to be important regulatory elements that are essential for specifying

Figure 5 Global DNA methylation, age and sex. Differences of mean methylation were determined in three tissues (heart muscle, skeletal muscle, liver) for two age groups (group 1, 26 ± 4 years; group 2, 68 ± 8 years (\pm s.d.), red line), for males and females (orange line) and for two different primary cells (CD4⁺ lymphocytes and dermal fibroblasts; blue line). As a control, tissues were resampled (10,000-fold) for both age groups, and their mean methylation differences were calculated (gray area). The same control was carried out for sex-specific differences, and similar results were obtained (data not shown). As a positive control for sex-specific methylation, an X-chromosomal gene (*ELK1*) was used that shows the expected methylation difference of about 50% (green line). Whereas the 7.1% difference between primary cells (blue line) is highly significant, the respective differences of 0.275% and 0.1% between age groups (red line) and sex (orange line) fall within the differential range observed for the control (gray area) and therefore are not significant.

tissue type identity in mammals, but we are aware of only a few, mostly CGI-associated T-DMRs in a small number of tissues (for review see ref. 30). Hierarchical clustering of our data showed that biological replicates of each tissue type clustered together (**Supplementary Fig. 5** online), indicating the presence of tissue-specific methylation profiles. Approximately 22% of the amplicons were T-DMRs ($P < 0.001$; **Supplementary Table 2** online). These were located within 5' UTRs, exons and introns of functionally diverse genes (**Fig. 2** and **Supplementary Table 2**). Within the 5' UTR, T-DMRs located within a CGI (**Supplementary Fig. 6** online) were strongly underrepresented (13% versus 87%, χ^2 test, $P < 0.001$). The comparatively low frequency of CGI-associated T-DMRs is consistent with previous reports using restriction landmark genome scanning (RLGS)^{31,32}. We also identified a number of genes (such as *JAG1*; **Supplementary Table 2**) that were differentially methylated in fetal tissues compared with their adult counterparts, emphasizing the importance of epigenetic mechanisms during mammalian development. Notably, we also found that T-DMRs were associated with both unprocessed and processed pseudogenes (such as *CMAH* and *AC000078.2-002*, respectively) and with evolutionarily conserved, non-protein coding regions (ECRs). In fact, we found that T-DMRs were strongly overrepresented in ECRs (χ^2 test, $P < 0.005$), and 30%

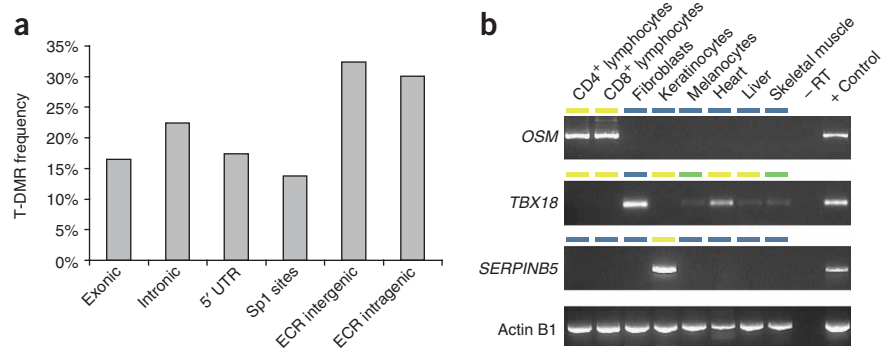


Figure 6 Analysis of T-DMRs. **(a)** Relative proportion of putative T-DMRs. Normalized for the number of amplicons in each category, the proportion of T-DMRs was highest in both intergenic and intragenic ECRs, whereas T-DMRs located within 5' UTRs had a lower frequency of occurrence. **(b)** Correlation between 5' UTR methylation and mRNA expression. Representative results are shown for two genes. We determined expression for 43 genes and one positive control, beta actin (*ACTB*) in eight tissues and cell types using RT-PCR. Total RNAs derived from mixed tissues and cell lines were used as positive controls. Differential 5' UTR methylation was inversely correlated with mRNA expression for *OSM* and *SERPINB5* (for which the inverse correlation was previously known) but not for *TBX18*. The color code depicts the degree of 5' UTR methylation for each gene (yellow = $\sim 0\%$ methylation, green = $\sim 50\%$ and blue = $\sim 100\%$).

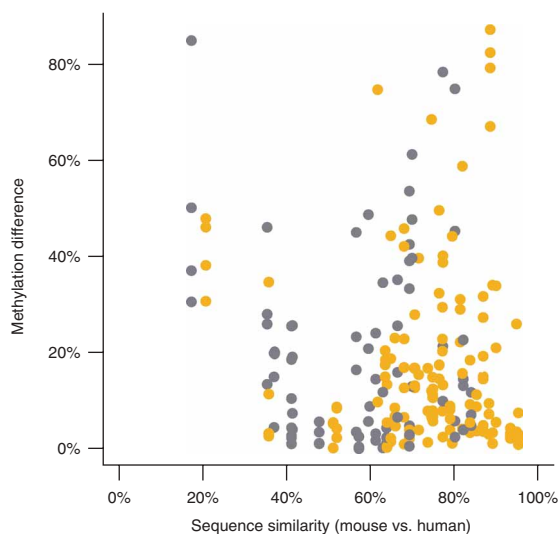


Figure 7 Conservation of methylation between human and mouse orthologous amplicons. We analyzed 59 orthologous amplicons (37 ECRs (yellow) and 22 5' UTRs (gray)) in four tissues (skin, skeletal muscle, heart muscle and liver) from both species. Methylation of the majority (69.4%) of ECR and 5' UTR amplicons differed by <20%, indicating significant conservation. Both hypermethylated and unmethylated amplicons showed a similar degree of methylation conservation (data not shown).

of all examined ECRs were T-DMRs compared with a T-DMR frequency of 17% in 5' UTRs and exons (Fig. 6a). Some of the T-DMR ECRs were located up to 100 kb away from the nearest annotated gene, consistent with putative long-range regulatory effects associated with enhancer or silencer function; however, this could also indicate the presence of as yet unknown genes. These findings support the notion that T-DMRs may have a functional role beyond the mere control of transcription via promoter methylation. For instance, comparative analysis of the mouse *IL4* locus identified two ECRs that undergo differential methylation during differentiation from naive CD4 to T_H1 and T_H2 cells and can act as enhancers for *IL4* expression (reviewed in ref. 33).

Transcriptional silencing by promoter methylation is one of the major mechanisms for tumor suppressor gene silencing and neoplastic transformation³⁴. Few genes have been found to be regulated by promoter methylation in healthy tissues³⁵; one example is *SERPINB5* (ref. 36), in which 5' UTR methylation correlates with the silencing of mRNA expression. We randomly selected 43 genes associated with 5' UTR T-DMRs and ten genes that contained T-DMRs within the gene, and we determined mRNA expression by RT-PCR. Of the 5' UTR T-DMRs, the methylation state did not correlate with mRNA expression levels for 63% of the genes and inversely correlated for 37% (examples of both possible situations are shown in Fig. 6b). Notably, genes without a CGI in their respective 5' UTRs (such as *oncostatin (OSM)*; Figs. 2 and 6b) also showed an inverse correlation, indicating that genes with a low CpG density might be subject to transcriptional regulation via DNA methylation as well. None of the T-DMRs located within genes showed a correlation with expression of the cognate mRNA. These observations suggest that in some cases, differential 5' UTR methylation might have only a permissive role, such as establishing an open chromatin conformation. In this model, additional factors that drive transcription, such as transcription factors or histone modifications, would be missing. Alternatively, the examined T-DMRs might not be located in the region that regulates transcription.

Conservation of DNA methylation

The conservation of DNA sequences between species is well studied, but much less is known about cross-species conservation of DNA methylation. To determine whether DNA methylation is conserved between species, and if so, to what degree, we compared the methylation profiles of 59 orthologous amplicons (as far as can be ascertained by conserved synteny and sequence similarity) in four human and mouse tissues (skin, liver, heart muscle and skeletal muscle). The amplicons were located either within 5' UTRs or within ECRs. The majority (69.4%) of profiles were conserved (differing by less than 20%) in both amplicon categories (Fig. 7); for example, in both species, we observed methylation of about 90% in the 5' UTR of *RIN2* in liver, whereas other tissues were consistently unmethylated. Only 4.3% of the orthologous loci differed by more than 60%, indicating that these amplicons were differentially hypermethylated or unmethylated in the two species. One such example is the 5' UTR amplicon of gene *ZC3H12D*, which was approximately 60% methylated in human tissue and unmethylated in the corresponding mouse tissues. Based on this analysis, we extrapolate that about 70% of orthologous loci between human and mouse may have conserved DNA methylation profiles (differing by <20%). This finding adds further evidence to the concept that many epigenetic states may be evolutionarily conserved between mammals. A recent study has already shown that epigenetic histone modifications are strongly conserved between human and mouse, even though many of the corresponding sites are not conserved at the DNA level³⁷.

DISCUSSION

The generation of a DNA methylation reference map of the human genome represents an important contribution towards the elucidation of the human epigenetic code. The present study gives new insights on how DNA methylation contributes to the epigenetic plasticity of the human genome and demonstrates that large-scale and quantifiable DNA methylation analysis at single-base pair resolution is possible using the sequencing infrastructure established for the Human Genome Project. Similar to the ENCODE³⁸ and HAPMAP³⁹ resources, the availability of a high-resolution DNA methylation resource adds another layer of information to the annotation and understanding of chromatin, which defines the functional state of the human genome. The HEP and other epigenome projects will be invaluable for the discovery of new epigenetic diagnostics and drugs⁴⁰, the monitoring of drug efficacy⁴¹ and the development of a truly integrated (epi)genetic approach⁴² to common disease.

METHODS

Cell and tissue samples. Tissue samples were obtained from one of the following sources: Asterand, Pathlore, Tissue Transformation Technologies, Northwest Andrology, National Disease Research Interchange and Biocat. Only anonymized samples were used, and ethical approval was obtained for the study from Ärztekammer Berlin and the Cambridge Local Research Ethics Committee. Contamination by blood cells is estimated to be low, as blood-specific methylation profiles were not detected in the tissues. Human primary cells were obtained from Cascade Biologics, Cell Applications, Analytical Biological Services, Cambrex Bio Science and the Deutsches Institut für Zell- und Gewebersatz. Dermal fibroblasts, keratinocytes and melanocytes were cultured according to the supplier's recommendations up to a maximum of three passages, reducing the risk of aberrant methylation due to extended culturing. As an additional control, we compared the average methylation of selected amplicons obtained from dermal fibroblasts, keratinocytes and melanocytes with the methylation of the same loci in additional human skin samples. We did not detect any significant deviation between the methylation of the primary cells and tissues, indicating that cell culturing for a limited

number of passages does not change DNA methylation. CD4⁺ T lymphocytes were isolated from fresh whole blood by depletion of CD4⁺ monocytes followed by negative selection. CD8⁺ cells were isolated from fresh whole blood by positive selection. Subsequent FACS analysis confirmed a purity of CD4⁺ CD8⁺ T lymphocytes of >90%. In some cases, DNA samples were pooled according to the sex and age of the donors. All genders were confirmed by sex-specific PCR.

Amplicon selection and classification. Amplicons were selected and classified into six categories (5' UTR, exonic, intronic, ECR, Sp1 and 'other') based on Ensembl^{22,43} (NCBI build 34) annotation. 5' UTR amplicons overlapped by at least 200 bp with (or within) a core region from 2,000 bp upstream to 500 bp downstream of the TSS. In cases where multiple sites were annotated per gene, the first annotated TSS was used. Exonic amplicons were those in which >50%, and at least 200 bp, of the amplicon overlapped with an annotated exon. Intronic amplicons were those in which >50%, and at least 200 bp, of the amplicon overlapped with an annotated intron. Amplicons classified as ECRs had at least four CpGs and ≥70% DNA sequence similarity between mouse and human noncoding sequences, for at least 100 bp. Out of 3,249 ECRs identified on chromosome 20, we selected 290 intergenic and 206 intronic (496 in total) ECRs. Amplicons classified as Sp1 overlapped with putative Sp1 sites identified by ChIP-chip analysis²⁴. Amplicons classified as 'other' were not located within a gene or 5' UTR and did not belong to any other category. CGIs were classified based on the criteria in ref. 44, except that they had to have a minimum length of 400 bp rather than 200 bp, as longer CGIs are less frequently associated with *Alu* repeats⁴⁵.

DNA extraction, PCR amplification and sequencing. DNA was extracted using the Qiagen DNA Genomic-Tip Kit according to the manufacturer's recommendations. After quantification, DNA was bisulfite-converted as previously described⁴⁶. Bisulfite-specific primers with a minimum length of 18 bp were designed using modified Primer-3 software (<http://frodo.wi.mit.edu/primer3/>). The target sequence of the designed primers did not contain any CpGs, allowing amplification of both unmethylated and hypermethylated DNAs. All primers were tested for their ability to yield high-quality sequences. Primers that gave rise to an amplicon of the expected size using non-bisulfite treated DNA as a template were discarded, thus ensuring the specificity for bisulfite-converted DNA. Primers were also tested for specificity by electronic PCR. DNA amplification was set up in 96-well plates using an automated pipeline as described previously⁶. PCR amplicons were quality controlled by agarose gel electrophoresis, rearranged into 384-well plates for high-throughput processing, cleaned up using ExoSAP-IT (USB) to remove any excess nucleotides and primers and sequenced directly in the forward and reverse directions. Some PCR amplicons were subcloned into the pGEM vector (Promega), and up to 20 clones were picked for sequencing. Sequencing was performed on ABI 3730 capillary sequencers using a 1:32 dilution of ABI Prism BigDye terminator V3.1 sequencing chemistry after hot start (96 °C for 30 s) thermocycling (44 cycles of 92 °C for 5 s, 50 °C for 5 s and 60 °C for 120 s) and ethanol precipitation. PCR fragments were sequenced using the same PCR amplification primers. Trace files and methylation signals at a given CpG site were quantified (estimated sensitivity: >20% difference in methylation) using ESME software as previously described⁴⁷. The bisulfite sequencing approach chosen here allows measurement of DNA methylation with high reproducibility and accuracy, as independent measurements are derived from both the sense and antisense strands of a PCR amplicon ($R = 0.87$; $N = 557,837$). In addition, about 4.1% of the amplicons were subjected to independent PCR amplification and sequencing. These technical replicates also showed high correlation ($R = 0.9$; $N = 15,655$). Furthermore, the signal is independent of the position of the measured CpG within the amplicon, which is supported by high correlation between measurements of the same CpGs in overlapping amplicons ($R = 0.85$; $N = 91,528$).

RNA extraction and RT-PCR. Aliquots of the same samples of the human melanocytes, keratinocytes, fibroblasts and CD4⁺ and CD8⁺ cells that were used for methylation analysis were used for RNA analysis. Primary cell cultures of human melanocytes, keratinocytes and dermal fibroblasts cells were harvested (after a maximum of three passages) and stored at -80 °C until RNA isolation.

Isolated RNA samples from heart, liver and skeletal muscle were purchased from Ambion and stored at -80 °C until used for reverse transcription. Total RNA was isolated using the Qiagen RNeasy kit followed by cDNA synthesis using the Qiagen Omniscript RT Kit with random hexamers. PCR (30–40 cycles of 92 °C for 1 min, 55–63 °C (depending on assay) for 1 min and 72 °C for 1 min) was performed using the HotStartTaq DNA Polymerase Kit (Qiagen) with 3 μl of the prepared cDNA and gene-specific primers. All kits were used according to the manufacturer's recommendations. PCR products were analyzed by electrophoresis on 2.5% agarose gels. Universal RNA was obtained from Biocat and total RNA isolated from brain and sperm from Stratagene.

Analysis and statistical methods. Methylation profiles were calculated as described previously⁶ and are available from the HEP database and browser at <http://www.epigenome.org>. Kruskal-Wallis tests were used to determine differential methylation between tissues (T-DMRs), measuring the proportion of uncorrected *P* values <0.001 for all CpGs. As this test is insensitive to samples that were measured in only a single sample, such as sperm and placenta, the obtained number of T-DMRs is unlikely to be overstated owing to putative aberrant methylation within these samples. Some T-DMRs were experimentally validated by sequencing independent DNA samples. Comparisons between two groups (separated by age or sex) were performed using Wilcoxon tests.

For the analysis of comethylation, median methylation values were used over all technical replicates to minimize any skewing effect because of possible outliers. In addition, we excluded all CpGs for which the methylation values derived from the forward and reverse reads of the same amplicon differed by >10%. Based on this criterion, 38% of CpGs were excluded from the analysis. As only one DNA strand was analyzed after bisulfite conversion, no assessment of hemimethylation was possible in this case. Methylation changes were calculated based on the absolute methylation differences between CpG pairs of identical samples. To minimize a bias introduced by the amplicon selection, the analysis was performed using both individual CpGs (window size, 20,000 bp) and CpGs of the same amplicons. Comethylation of CpGs was described as a function of similar methylation levels over distance (in bp).

For scatter plots, equal numbers of measurements were binned and ranked by numerical order of the *x* axis values, representing means of *x* and *y* data. For box plots and histograms, data were binned according to the intervals indicated on the *x* axis containing different numbers of measurements.

URLs. Data described in the manuscript and the software used for the analysis of all loci are freely available at <http://www.epigenome.org>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank E. Calautti for his advice on culturing of keratinocytes, A. Meyerhans for critical reading of the manuscript, J. Maass for her help obtaining tissue samples and K. Fischer for his support providing genomic annotations. E.E. thanks Y.-S. Kim for many discussions. V.K.R. was supported by a C.J. Martin Fellowship from the National Health and Medical Research Council of Australia. J.A., J.B., T.C., R.D., T.A.D., R.H., K.H., D.K.J., J.L., D.N., R.P., T.W., J.R. and S.B. were supported by the Wellcome Trust.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245–254 (2003).
4. Murrell, A., Rakyant, V.K. & Beck, S. From genome to epigenome. *Hum. Mol. Genet.* **14**, R3–R10 (2005).
5. Jones, P.A. & Martienssen, R. A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. *Cancer Res.* **65**, 11241–11246 (2005).

6. Rakyan, V.K. *et al.* DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* **2**, 2170–2182 (2004).
7. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
8. Schumacher, A. *et al.* Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**, 528–542 (2006).
9. Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: The HELP assay. *Genome Res.* **16**, 1046–1055 (2006).
10. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831 (1992).
11. Strichman-Almashanu, L.Z. *et al.* A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res.* **12**, 543–554 (2002).
12. Smiraglia, D.J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10**, 1413–1419 (2001).
13. Grunau, C., Hindermann, W. & Rosenthal, A. Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. *Hum. Mol. Genet.* **9**, 2651–2663 (2000).
14. Duncan, B.K. & Miller, J.H. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560–561 (1980).
15. Hayward, B.E. *et al.* The human *GNAS1* gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc. Natl. Acad. Sci. USA* **95**, 10038–10043 (1998).
16. Kalscheuer, V.M., Mariman, E.C., Schepens, M.T., Rehder, H. & Ropers, H.H. The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat. Genet.* **5**, 74–78 (1993).
17. Verhaagh, S., Schweifer, N., Barlow, D.P. & Zwart, R. Cloning of the mouse and human solute carrier 22a3 (*Slc22a3/SLC22A3*) identifies a conserved cluster of three organic cation transporters on mouse chromosome 17 and human 6q26-q27. *Genomics* **55**, 209–218 (1999).
18. Xu, G.L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
19. Frigola, J. *et al.* Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat. Genet.* **38**, 540–549 (2006).
20. Zeng, W. *et al.* Transcript profile of CD4⁺ and CD8⁺ T cells from the bone marrow of acquired aplastic anemia patients. *Exp. Hematol.* **32**, 806–814 (2004).
21. Ashurst, J.L. *et al.* The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.* **33**, D459–D465 (2005).
22. Down, T.A. & Hubbard, T.J.P. Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res.* **12**, 458–461 (2002).
23. Fuks, F. DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.* **15**, 490–495 (2005).
24. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
25. Mancini, D.N., Singh, S.M., Archer, T.K. & Rodenhiser, D.I. Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. *Oncogene* **18**, 4108–4119 (1999).
26. Clark, S.J., Harrison, J. & Molloy, P.L. Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene* **195**, 67–71 (1997).
27. Holler, M., Westin, G., Jiricny, J. & Schaffner, W. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* **2**, 1127–1135 (1988).
28. Harrington, M.A., Jones, P.A., Imagawa, M. & Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* **85**, 2066–2070 (1988).
29. Fraga, M.F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA* **102**, 10604–10609 (2005).
30. Shiota, K. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. *Cytogenet. Genome Res.* **105**, 325–334 (2004).
31. Costello, J.F., Smiraglia, D.J. & Plass, C. Restriction landmark genome scanning. *Methods* **27**, 144–149 (2002).
32. Shiota, K. *et al.* Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells* **7**, 961–969 (2002).
33. Ansel, K.M., Djuretic, I., Tanasa, B. & Rao, A. Regulation of Th2 differentiation and I14 locus accessibility. *Annu. Rev. Immunol.* **24**, 607–656 (2006).
34. Jones, P.A. & Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
35. Song, F. *et al.* Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc. Natl. Acad. Sci. USA* **102**, 3336–3341 (2005).
36. Futscher, B.W. *et al.* Role for DNA methylation in the control of cell type specific maspin expression. *Nat. Genet.* **31**, 175–179 (2002).
37. Bernstein, B.E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
38. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
39. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
40. Yoo, C.B. & Jones, P.A. Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.* **5**, 37–50 (2006).
41. Widschwendter, M. *et al.* Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res.* **64**, 3807–3813 (2004).
42. Bjornsson, H.T., Fallin, M.D. & Feinberg, A.P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* **20**, 350–358 (2004).
43. Curwen, V. *et al.* The Ensembl automatic gene annotation system. *Genome Res.* **14**, 942–950 (2004).
44. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
45. Takai, D. & Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* **99**, 3740–3745 (2002).
46. Berlin, K., Ballhause, M. & Cardon, K. Improved bisulfite conversion of DNA. Patent PCT/WO/2005/038051 (2005).
47. Lewin, J., Schmitt, A.O., Adorjan, P., Hildmann, T. & Piepenbrock, C. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics* **20**, 3005–3012 (2004).