E16.5 skeletal muscle tissues. The E16.5 skeletal muscle was chosen as a reference tissue for the relative expression levels of *Dlk1* and *Gtl2* because both transcripts have been shown to be highly expressed in E14.5 to E18.5 skeletal muscle [19,20,31]. We obtained the Ct (cycle threshold) values of *Gtl2* in E5.5–7.5 tissues (23.2 to 33.5) within the Ct value range of the standard curve for *Gtl2* (20.0 to 35.3), and determined the relative expression levels of *Gtl2* in these tissues, ranging from 0.0036 (E7.5Exe) to 0.31 (E6.5Exe) relative to the average level of E16.5 skeletal muscle samples (Fig. 2A). As demonstrated by Schuster-Gossler et al. [31], we were consistently able to detect the expression of *Gtl2* in blastocysts in our replicate samples, and determined its average relative expression level to be 0.011. However, this value should be considered with caution because the Ct values of the blastocyct samples (35.9, 36.4, and 36.5) were slightly out of range. Our results demonstrate that *Gtl2* is expressed at low levels at E3.5, up-regulated transiently at E5.5 and E6.5, and down-regulated at E7.5. Variation in the relative expression levels of *Gtl2* among three replicate samples was most remarkable in extra-embryonic tissues at E6.5. Because *Gtl2* expression seems to decrease rapidly between E6.5 and E7.5 in extra-embryonic tissues, subtle differences in the developmental stage among our E6.5 samples may account for the variation in *Gtl2* expression levels.

We obtained the Ct values of *Dlk1* in E5.5–7.5 tissues (26.5 to 35.4) within the Ct value range of the standard curve for *Dlk1* (21.6 to 36.8), and determined that the relative expression levels of *Dlk1* in these tissues ranged from 0.00021 (E7.5Exe) to 0.015 (E7.5Emb). The expression of *Dlk1* was undetectable in the blastocyst samples, and was consistently low in E5.5–7.5 tissues. In embryonic tissues, *Dlk1* expression levels tended to increase as embryonic development progressed (Fig. 2A). These results represent the first quantitative measurement of *Gtl2* and *Dlk1* expression levels during early gestational stages. The transient up-regulation of *Gtl2* around E5.5 and E6.5 stages suggest a possible role of this ncRNA in the control of growth and differentiation at these developmental stages.

To complete our analysis of the allelic expression patterns of *Glt2* and *Dlk1*, we quantitatively measured the allelic expression levels of these transcripts by pyrosequencing. *Gtl2* was consistently found to be

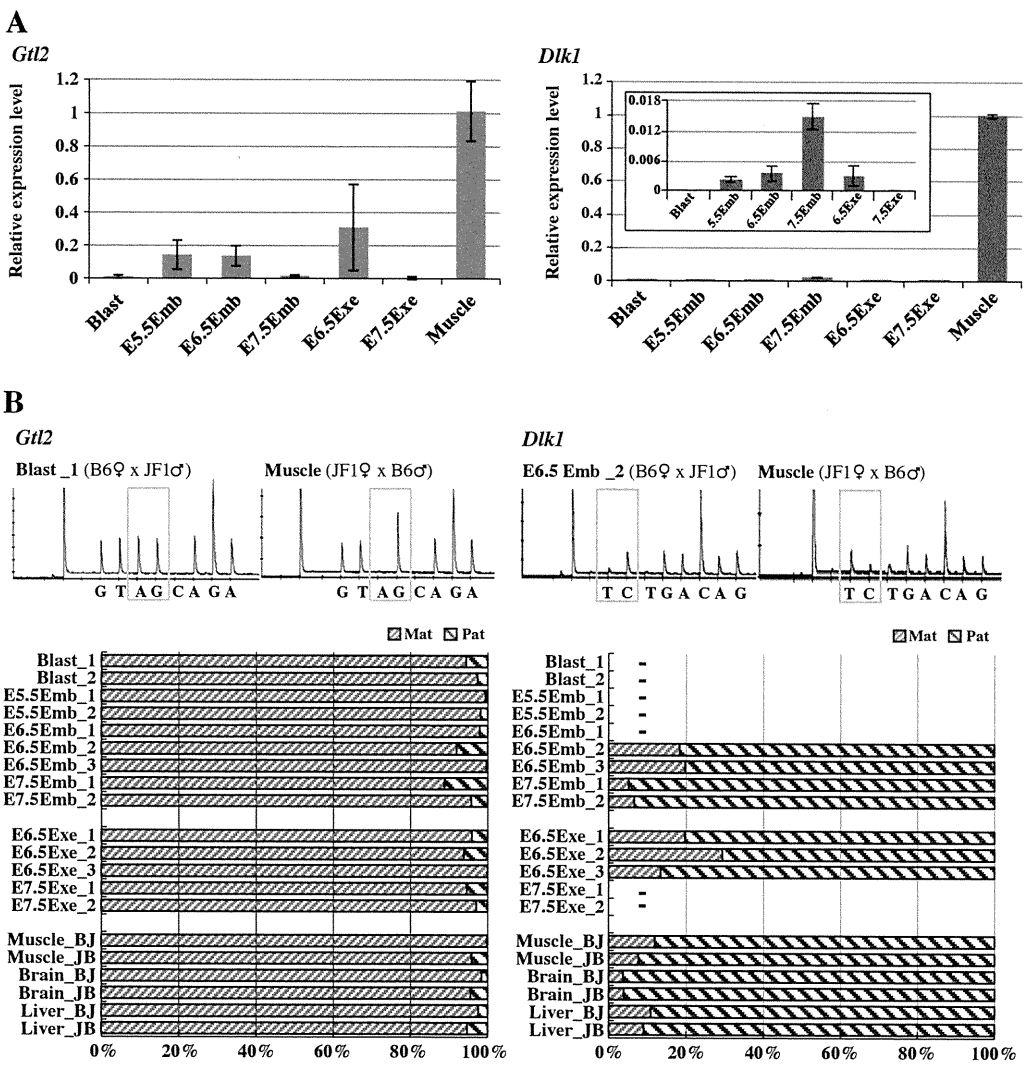

Fig. 2. Expression analysis of *Gtl2* and *Dlk1* during embryonic and extra-embryonic development. (A) Graphical representation of the relative expression levels of *Gtl2* and *Dlk1* in embryonic and extra-embryonic tissues at E3.5 to E7.5. The bars represent the mean expression levels of replicate samples relative to the mean expression level of E16.5 muscle samples (n = 4). Error bar = standard deviation (SD). The mean ± SD of each sample set for *Gtl2*: Blast, 0.011 ± 0.0094; E5.5Emb, 0.14 ± 0.088; E6.5Emb, 0.14 ± 0.061; E7.5Emb, 0.018 ± 0.0033; E6.5Exe, 0.31 ± 0.26; E7.5Exe, 0.0036 ± 0.0095. The mean ± SD for *Dlk1*: E5.5Emb, 0.0023 ± 0.00066; E6.5Emb, 0.0035 ± 0.0016; E7.5Emb, 0.015 ± 0.0025; E6.5Exe, 0.0030 ± 0.0021; E7.5Exe, 0.00021 ± 0.00015. (B) Quantitative allelic expression analysis of *Gtl2* and *Dlk1* by pyrosequencing. The top panels show examples of pyrograms. The yellow box in each pyrogram denotes the peaks at SNPs between the B6 and the JF1 strains. The alleles (B6/JF1) of SNPs are A/G for *Gtl2* and T/C for *Dlk1*. The bottom panels represent the allelic expression ratios of the paternal (Pat, blue stripe) and the maternal (Mat, red stripe) alleles. "–" in the panel for *Dlk1* indicates that the corresponding sample was not analyzed due to the absence or low expression of *Dlk1*.

maternally expressed in blastocysts (E3.5) and E5.5 embryos, as well as latter developmental stages (E6.5 and E7.5) in both embryonic and extra-embryonic tissues (Fig. 2B). These results demonstrate that the repression of *Gtl2* on the paternal allele occurs at E3.5 without methylation at its promoter (R4/R5) (Fig. 1B). Due to the relatively low levels of *Dlk1* expression, we only determined the allelic expression status of *Dlk1* in embryonic tissues at E6.5 and E7.5 and extra-embryonic tissues at E6.5. *Dlk1* was predominantly expressed from the paternal allele in E6.5 tissues, and this allelic preference was more dramatic in embryonic tissues at E7.5 (Fig. 2B). Relaxation of imprinting was more evident for *Dlk1* than *Gtl2* in the E6.5 tissues. Using a quantitative method, we successfully determined the allelic expression patterns of *Gtl2* and *Dlk1* during early gestational stages for the first time.

We also measured allelic expression levels of *Gtl2* and *Dlk1* in fetal skeletal muscle, brain, and liver tissues at E15.5/E16.5, and found that *Gtl2* and *Dlk1* were exclusively or predominantly expressed from the maternal and the paternal alleles, respectively (Fig. 2B). The ratio of *Dlk1* expression from the maternal allele (relaxation of imprinting) was higher in muscle and liver than in brain. Such allelic expression patterns replicate the data produced by da Rocha et al. [32]. We compared allelic methylation patterns at the R2 and R3 regions of IG-DMR (Fig. 1D) with the allelic expression patterns of *Dlk1* (Fig. 2B) in three tissues. However, there was no clear correlation between methylation levels at the R2/R3 regions and the extent to which *Dlk1*'s imprinting was relaxed.

3. Discussion

3.1. Establishment of allele-specific DNA methylation patterns at secondary DMRs

In this study, we demonstrated that the paternal allele of the *Gtl2*-DMR gains DNA methylation after the blastocyst stage and becomes fully methylated by the E6.5 stage in the embryonic lineage. Additionally, we determined that DNA methylation at the *Gtl2*-DMR is not a prerequisite for the imprinted expression of *Gtl2* in early development (summarized in Fig. 3 and Table 1). The timing of the establishment of parent-of-origin-dependent differential methylation patterns during post-zygotic development has previously been determined for several secondary DMRs [13,33–38] (summarized in Table 1). Although the developmental stage at which differential methylation is established differs among secondary DMRs including the *Gtl2*-DMR, it is frequently observed that the imprinted expression of the gene associated with the secondary DMR is already established prior to the gain of differential DNA methylation. This observation holds true even among DMRs which are regulated by different mechanisms.

In mid-to-late gestation fetuses that carry an insertion mutation upstream of the *Gtl2*-DMR on the paternal allele, it has been observed that the *Gtl2*-DMR loses its paternal methylation and *Gtl2* is biallelically expressed [39,40]. It has also been shown that *Cdkn1c* is biallelically expressed in E9.5 embryos that are deficient in DNMT1 activity [33]. These observations suggest that paternal methylation of the *Gtl2*- and the *Cdkn1c*-DMRs is necessary to maintain silencing of the paternal allele of these genes. However, whether DNA methylation plays a critical role in the maintenance of imprinted gene expression remains unproven for the other secondary DMRs. While Lsh, a member of the SNF2 family of chromatin remodeling proteins, is shown to be required for the proper acquisition of the paternal methylation at the *Cdkn1c*-DMR [41], information regarding such *trans*-acting factors is lacking for the majority of secondary DMRs.

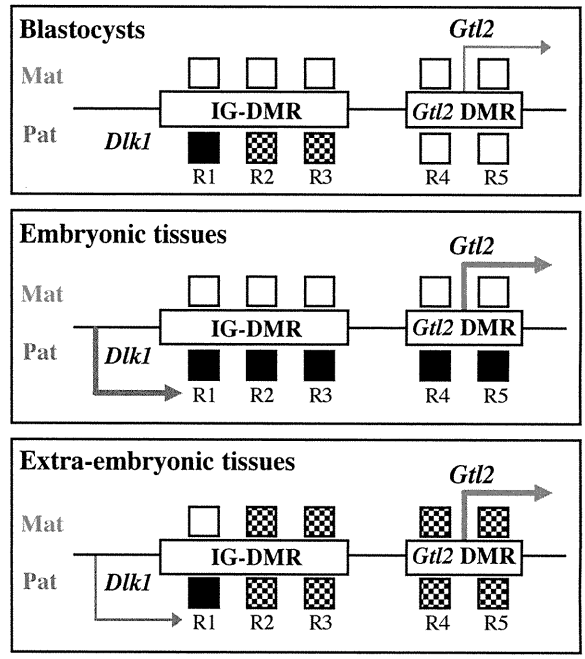

Fig. 3. Summary of the allelic DNA methylation patterns of IG-DMR and *Gtl2*-DMR and the allelic expression patterns of *Gtl2* and *Dlk1* during embryogenesis. The methylation and expression patterns in blastocysts and embryonic/extra-embryonic tissues at E6.5 and E7.5 are schematically shown. The squares above and below the DMRs symbolize the regions subjected to bisulfite sequencing analysis (R1 to R5). Open, filled, and mosaic squares indicate hypomethylated, hypermethylated, and intermediately methylated regions, respectively. Thick and thin arrows represent high and low expression levels of *Gtl2* and *Dlk1*. In blastocysts, the maternal allele-specific expression of *Gtl2* occurs without the paternal allele-specific methylation of the *Gtl2*-DMR. In embryonic tissues, the paternal allele-specific methylation of *Gtl2*-DMR is established by E6.5. Allele-specific differential methylation is well maintained in both the IG-DMR and the *Gtl2*-DMR in embryonic tissues. In extra-embryonic tissues at E6.5, imprinted expression of *Gtl2* and *Dlk1* occurs despite the loss of allele-specific differential methylation at the R2 and R3 regions in the IG-DMR and in the *Gtl2*-DMR. These allelic methylation/expression patterns suggest that the R1 region may contain a methylation-dependent element that is critical for the ICR function of the IG-DMR, and suggest that epigenetic modifications other than DNA methylation may play more critical roles in the maintenance of imprinted expression in the *Dio3-Dlk1* domain in extra-embryonic tissues than in the embryonic tissues.

3.2. Epigenetic modifications of the IG-DMR and the Gtl2-DMR during early development

In E7.5 embryo deficient in the Polycomb group (PcG) gene *Eed*, it has been shown that *Gtl2* is biallelically expressed whereas *Dlk1* is properly imprinted (paternally expressed) [42], suggesting a role for PcG complex proteins in the regulation of paternal *Gtl2* silencing. The EED-containing Polycomb Repressive Complex 2 (PRC2) catalyzes trimethylation of histone H3 lysine-27 (H3K27me3), which is a binding site for the repressive PRC1 complex [43]. It has been demonstrated by Hammoud et al. that, in human sperm, the *MEG3/GTL2*-DMR bears both active trimethylation of histone H3 lysine-4 (H3K4me3) and repressive H3K27me3 histone marks [44]. DNMT3L recognizes unmethylated H3K4 and induces de novo DNA methylation by recruitment or activation of DNMT3A2 [11,12]. The interaction of DNMT3L with H3K4 is strongly inhibited by methylation at H3K4 (the higher the degree of methylation at H3K4, the more severely the binding of DNMT3L with histone H3 N-terminal is abolished) [11]. In light of the post-zygotic acquisition of DNA methylation at the *MEG3/GTL2*-DMR and the bivalent histone pattern in this region, Hammoud et al. hypothesized that H3K4me3 may prevent DNA methylation in the sperm and early embryo, and H3K27me3 may ensure early silencing at this locus [44]. In this study, we have shown that, in the blastocyst stage, *Gtl2* is already expressed primarily from the maternal allele in the absence of DNA

Table 1
The timing of the establishment of parental-origin-dependent differential methylation at secondary DMRs and imprinted expression of associated genes during post-zygotic development.

Imprinted gene	Secondary DMR	Methylation status established (methylated allele)	Monoallelic expression observed (expressed allele)	Primary DMR (methylated allele)	References
Gtl2	*Gtl2*-DMR	E6.5 (Pat)	Blastocyst (Mat)	IG-DMR (Pat)	This study
Cdkn1c	*Cdkn1c*-DMR	E9.5 (Pat)	Morula (Mat)	KvDMR1 (Mat)	[33,34]
Igf2	*Igf2*-DMR1/DMR2	E15.5 (Pat)	Blastocyst (Pat)	*H19*-DMR (Pat)	[13,35]
H19	*H19* promoter	E6.5 (Pat)	Morula (Mat)	*H19*-DMR (Pat)	[36]
Igf2r	*Igf2r*-DMR1	From E15.5 to 4dpp[a](Pat)	E6.5 (Mat)	*Igf2r*-DMR2 (Mat)	[37,38]

[a] Days post partum.

methylation at the paternal allele of the *Gtl2*-DMR. Assuming that the bivalent histone modification pattern observed at the MEG3/GTL2-DMR in human sperm is conserved in mice and is preserved in preimplantation embryos, our results can be explained by such histone modifications. If H3K27m3 and H3K4me3 marks are present at the paternal allele of the Gtl2-DMR in the blastocyst stage, these could be attributed to the paternal repression of Gtl2 and the absence of DNA methylation at the locus, respectively. Since the *Gtl2*-DMR was found to gain DNA methylation after the blastocyst stage, the H3K4me3 mark may be erased by then. It has been shown that in E12–14 embryos, the *Gtl2*-DMR is not enriched for the H3K27me3 mark on both parental alleles [45]. Therefore, the H3K27me3 mark may also be erased during or after the establishment of DNA methylation at the *Gtl2*-DMR. DNA methylation is considered to be more critical for the paternal repression of *Gtl2* in the absence of the H3K27me3 mark.

Our results demonstrate that during early post-zygotic development, the germline-derived DNA methylation on the paternal allele of the IG-DMR is partially lost by the E3.5 stage and its methylation is restored by E5.5. Histone modification patterns on the paternal allele of the IG-DMR may be responsible for the recruitment of de novo methylation machinery between E3.5 and E5.5 stages. In the human sperm, the IG-DMR is shown to be devoid of H3K4me3 [44]. If the non-methylated status of H3K4 is maintained on the paternal allele of the mouse IG-DMR at the blastocyst stage, the DNMT3L/DNMT3A complex may possibly be recruited to direct de novo cytosine methylation to restore the methylation of the IG-DMR.

We observed a reduction of DNA methylation levels on the paternal allele of the IG-DMR in blastocysts (E3.5) compared to methylation levels in the sperm. Loss of methylation on the paternal allele of IG-DMR was also observed in extra-embryonic tissues at the E7.5 stage compared to those at E5.5. Interestingly, throughout stages E3.5 to E7.5 of development, several CpG sites within the R1 region in the IG-DMR (such as 5th, 7th, 17th, 20th and 21st CpG sites in the R1 region in Fig. 1B–D and Supplementary Fig. 1) show high levels of DNA methylation on the paternal allele. These CpG sites may have greater functional importance than other CpGs in the IG-DMR. We also observed tissue-specific relaxation of differential methylation in the R2 and R3 regions of the IG-DMR among E15.5/16.5 tissues (Fig. 1C). It remains to be elucidated whether the R2 and R3 regions have critical roles in the regulation of imprinted expression in the *Dlk1-Dio3* cluster. Creation and characterization of a knockout mouse line with a mutation or a deletion at a particular subregion of IG-DMR will help further dissect the role of IG-DMR as the ICR of the *Dlk1-Dio3* cluster.

3.3. Difference in the allelic DNA methylation patterns between the embryonic and the extra-embryonic lineages at the Gtl2-DMR and the IG-DMR

Mechanisms that regulate the imprinted expression of maternally expressed ncRNAs in the *Dlk1-Dio3* domain are suggested to differ between the embryo and the placenta [21]. However, the principal mechanisms regulating allele-specific expression in each of the two lineages are unknown. The *Gtl2*-DMR was previously shown to be partially methylated on both parental alleles in the placenta of late gestation embryos (E16.5) [21]. Here, we clearly determined that the *Gtl2*-DMR is partially methylated on both parental alleles in the placenta as early as E6.5. Furthermore, our data demonstrated that part of the IG-DMR also becomes partially methylated on both parental alleles in the placenta by E7.5. Therefore, further epigenetic profiling of these DMRs in embryonic and extra-embryonic lineages, which include DNA methylation as well as histone modifications, will help define underlying regulatory mechanisms. The differential regulation of imprinted genes between the embryo and placenta has been well-characterized for the *Kcnq1* imprinted gene cluster [34,46,47]. In 2006, Lewis et al. subjected ES and TS cells as well as their differentiated derivates to epigenetic profiling [47]. These cells, which maintain imprinted expression of the *Kcnq1* cluster of genes, were found to establish monoallelic gene expression as well as differential histone marks for placenta-specific imprinted genes during differentiation of the extra-embryonic lineage between E4.5 and E7.5 [47]. The use of such stem cells would be an effective strategy to determine the epigenetic regulatory differences in the *Dlk1-Dio3* domain between the embryonic and the extra-embryonic lineages.

3.4. Developmental functions of genes within the Dlk1-Dio3 cluster

Gtl2 and *Dlk1* are shown to be expressed in most tissues after E12.5, and to reach their highest expression levels after E15.5 in several tissues [19]. In mice, maternally transmitted deletions of the IG-DMR or the *Gtl2*-DMR result in the dysregulation of imprinted expression of the genes within the *Dlk1-Dio3* cluster [15,21,23,24]. Such mice exhibit prenatal/perinatal lethality and dysplastic phenotypes in various tissues such as skeletal muscle, bone, liver, and lung [21,23,24]. Recently, iPSCs with repressed expression of maternally expressed ncRNAs within the *Dlk1-Dio3* cluster (*Gtl2*^{off} iPSCs) were shown to contribute poorly to chimeras [25]. When *Gtl2*^{off} and *Gtl2*^{on} iPSCs were injected into tetraploid blastocysts, morphologically normal embryos (at E11.5) were obtained from both types of iPSCs. However, almost all embryos derived from the *Gtl2*^{off} iPSCs were found to be dead at E11.5 [25]. These observations indicate that genes in the *Dlk1-Dio3* cluster play important roles in cell proliferation and/or differentiation, not only at late-gestation stages, but at developmental stages earlier than E11.5. In this study, we revealed that *Gtl2* is transiently up-regulated and expressed at comparably high levels at E5.5 to E6.5 stages. On the other hand, *Gtl2* is expressed at low levels in blastocysts and E7.5 to E11.5 embryos (this study and Ref. [19]). Taken together, *Gtl2* (and the other maternally expressed ncRNAs) may be involved in the control of proliferation and differentiation of cells at early gestation stages (E5.5 to E6.5).

4. Materials and methods

4.1. Sample preparations

To obtain early embryonic and extra-embryonic tissues, female C57BL/6 (B6) (*Mus musculus musculus*) mice (Sankyo Labo Service,

Tokyo, Japan) mated with male JF1/Ms (JF1) (*Mus musculus molossinus*) mice were sacrificed at appropriate stages according to the guidelines for the care and use of laboratory animals (National Research Institute for Child Health and Development, Japan). At noon on the day the copulation plug was found was designated as embryonic day 0.5 (E0.5). Three to six samples (whole embryos or dissected tissues) at each of the developmental stages from E3.5 to E6.5 were pooled. The numbers of the samples pooled were: six blastocysts (E3.5) for each of three independent pools (Blast _1 to Blast_3); four or five whole embryos at E5.5 for three independent pools (E5.5Emb_1 to E5.5Emb_3); three or four embryonic tissues at E6.5 for three independent pools (E6.5Emb_1 to E6.5Emb_3); three or four extra-embryonic tissues at E6.5 for three independent pools (E6.5Ex_1 to E6.5Ex_3). Embryonic and extra-embryonic tissues at E7.5 were prepared from three embryos and analyzed individually (E7.5Emb_1 to E7.5Emb_3, and E7.5Ex_1 to E7.5Ex_3). These pooled and individual samples were subjected to simultaneous isolation of genomic DNA and total RNA using Allprep DNA/RNA micro or mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Fetal tissues (E16.5 skeletal muscle, E15.5 brain, and E16.5 liver) were obtained from the F1 hybrid fetuses derived from reciprocal matings between the B6 and the JF1 strains.

Spermatozoa were collected from the vas deferens of JF1 mice. Isolation of the DNA was performed as described previously [48]. Briefly, sperm genomic DNA was isolated by treatment with 1% 2-mercaptoethanol (Sigma-Aidrich, Tokyo, Japan) and proteinase K (Sigma), and followed by phenol/chloroform extraction and ethanol precipitation.

4.2. Sodium bisulfite genomic sequencing

Sodium bisulfite treatment was performed using EZ DNA methylation direct kit (Zymo Research, Orange, CA) according to the manufacturer's instructions. PCR was performed using one unit of Biotaq HS DNA polymerase (Bioline, London, UK) and primer sets as follows: 5′-TGTGTTGTGGATTTAGGTTGTAGTTTA-3′ and 5′-TAATCCCATTCCCAATCTATAAAAATA-3′ for R1 (nt 81187–81647), 5′-CCAAAACAAACCCAATAAATTCTAA-3′ and 5′-TGGTGAGTTTTGGTTAGAAAAGTGT-3′ for R2 (nt 82265–82615), 5′-CCCCCAATAACTTATAAACCATAATACT-3′ and 5′-GGATGGTAGTAGATAATTTGTTGTTTGA-3′ for R3 (nt 83273–83670), 5′-AAATCAAAATCCTTTTACCTCAACAATA-3′ and 5′-GGAAATAATTTTAATTGGTGATTGTTTT-3′ for R4 (nt 93435–93731), and 5′-AAATTTTGTAAGGAAAAGAATTTTTAGG-3′ and 5′-TTCAAAATTACTAATCAACATAAACCTC-3′ for R5 (nt 94288–94671). The thermocycling conditions were 35 to 45 cycles of 94 °C for 30 seconds (s), 55 °C for 30 s, and 72 °C for 30 s, with an initial step of 95 °C for 10 minutes (min) and a final step of 72 °C for 7 min. The amplified PCR products were cloned into pGEM T-Easy vector (Promega, Madison, WI), and sequenced by 3130xl Genetic Analyzer (Applied Biosystems, Foster city, CA). Nucleotide position (nt) and rs ID of SNPs and alleles (B6/JF1 alleles) within each region are as follows: R1, nt 81275 (rs46718958, A/G), nt 81422 (rs47741870, G/A), and nt 81610 (G/A); R2, nt 82369 (rs52043811, C/T); R3, nt 83406 (T/C), nt 83546 (rs46395233, C/T), nt 83593 (rs50881257, C/T), and nt 83631 (rs46982259, C/A); R4, nt 93671 (C/T); R5, nt 94561 (G/A). All nucleotide positions refer to the sequence of GenBank accession no. AJ320506.1. Allelic methylation patterns were assessed using two independent (pooled or individual) samples for the E3.5 to E7.5 stages, and using one each sample for the JF1 sperm and the E15.5/16.5 fetal tissues. The methylation patterns of individual clones are shown in Supplementary Fig. 1 for all regions and for all samples analyzed. Overall methylation percentage for each region (the number of methylated CpGs per number of total CpGs) was calculated for each type of tissues, and is shown in Fig. 1. When bisulfite sequencing data were available from two independent sample sets (for the tissues from the E3.5 to E7.5 stages), two datasets were combined to calculate the overall methylation percentage. The numbers of individual clones used to determine methylation percentages ranged from 5 to 35 (average 16.8).

4.3. Quantitative real-time RT-PCR

First strand cDNA synthesis was carried out by random hexamers using QuantiTect Reverse Transcription (Qiagen) according to the manufacturer's instructions. For quantitative real-time RT-PCR, amplification was performed using SYBR premix Ex taq (Takara, Kyoto, Japan) and primer sets as follows: 5′-CAGGACCCTCCAACTGTAAATC-3′ and 5′-AGGTAGGAACCTGAGCCCATTT-3′ for *Gtl2* (nt 1553–1818); 5′-CTCTTGCTCCTGCTGGCTTT-3′ and 5′-CTTGTGCTGGCAGTCCTTTC-3′ for *Dlk1* (nt 176–526); 5′-CTGCACCACCAACTGCTTAG-3′ and 5′-CCTGCTTCACCACCTTCTTG-3′ for *Gapdh*. Ct values for *Dlk1* and *Gtl2* as well as *Gapdh* as a reference were determined using the 7900HT fast real-time PCR system (Applied Biosystems). The average Ct value was calculated from the Ct values from duplicate reactions for the same cDNA sample. The relative expression levels of *Dlk1* and *Gtl2* were determined by the delta–delta Ct method. Delta Ct values were calculated using the Ct values of *Gapdh* as a reference gene. Delta–delta Ct values were calculated using the average delta Ct value of E16.5 skeletal muscle samples as a reference. To determine the range of Ct values in which quantitative accuracy is expected for *Dlk1*, *Gtl2*, and *Gapdh*, we generated a standard curve for each of the genes using an 8-fold dilution series (six dilutions) of a cDNA sample (12.5 dpc placenta). We obtained a linear standard curve showing $R^2>0.999$ with the Ct value ranging from 21.6 to 36.8 (amplification efficiency per cycle = 97.1%) for *Dlk1*, from 20.0 to 35.3 (95.6%) for *Gtl2*, and from 16.0 to 28.3 (95.9%) for *Gapdh*.

4.4. Allelic expression analysis by pyrosequencing

RT-PCR was performed using one unit of Biotaq HS DNA polymerase. The thermocycling conditions were 35 to 45 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, with an initial step of 95 °C for 10 min and a final step of 72 °C for 7 min. The primer sets used (and PCR product size in parentheses) are 5′-biotin-CAGGACCCTCCAACTGTAAATC-3′ and 5′-AGGTAGGAACCTGAGCCCATTT-3′ for *Gtl2* (266 bp), and 5′-CTCTTGCTCCTGCTGGCTTT-3′ and 5′-biotin-CTTGTGCTGGCAGTCCTTTC-3′ for *Dlk1* (94 bp). Pyrosequencing was carried out using the PSQ 96 MA system (Qiagen) and the PSQ 96 SNP Reagent kit (Qiagen) according to the manufacturer's instructions. In brief, the biotinylated PCR products were purified with the Streptavidin Sepharose HP beads (GE Healthcare, Uppsala, Sweden). The purified PCR products were washed, denatured and then annealed with a sequencing primer (5′-GGCGTCCCCGTGGCT-3′ for *Gtl2* and 5′-ATGCGACCCACCCTG-3′ for *Dlk1*). The nucleotide position and alleles (B6/JF1) of SNPs within the regions analyzed are nt 1579 of NR_027652.1 (rs46969056; A/G on the reverse strand) for *Gtl2* and nt 235 of NM_010052.4 (T/C on the forward strand) for *Dlk1*.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.05.003.

Acknowledgments

We thank Dr. Layla Parker-Katiraee for her critical reading of the manuscript. This study was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology, Japan, to KH and KN, and by and a Grant-in Aid for Scientific Research from the Japan Society for the Promotion of the Science to KN.

References

[1] A.C. Ferguson-Smith, M.A. Surani, Imprinting and the epigenetic asymmetry between parental genomes, Science 293 (2001) 1086–1089.
[2] D. Lucifero, M.R. Mann, M.S. Bartolomei, J.M. Trasler, Gene-specific timing and epigenetic memory in oocyte imprinting, Hum. Mol. Genet. 13 (2004) 839–849.

[3] H. Hiura, Y. Obata, J. Komiyama, M. Shirai, T. Kono, Oocyte growth-dependent progression of maternal imprinting in mice, Genes Cells 11 (2006) 353–361.
[4] T.L. Davis, J.M. Trasler, S.B. Moss, G.J. Yang, M.S. Bartolomei, Acquisition of the *H19* methylation imprint occurs differentially on the parental alleles during spermatogenesis, Genomics 58 (1999) 18–28.
[5] J.Y. Li, D.J. Lees-Murdock, G.L. Xu, C.P. Walsh, Timing of establishment of paternal methylation imprints in the mouse, Genomics 84 (2004) 952–960.
[6] H. Hiura, J. Komiyama, M. Shirai, Y. Obata, H. Ogawa, T. Kono, DNA methylation imprints on the IG-DMR of the *Dlk1-Gtl2* domain in mouse male germline, FEBS Lett. 581 (2007) 1255–1260.
[7] D. Bourc'his, G.L. Xu, C.S. Lin, B. Bollman, T.H. Bestor, Dnmt3L and the establishment of maternal genomic imprints, Science 294 (2001) 2536–2539.
[8] K. Hata, M. Okano, H. Lei, E. Li, Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice, Development 129 (2002) 1983–1993.
[9] M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, H. Sasaki, Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting, Nature 429 (2004) 900–903.
[10] Y. Kato, M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, H. Sasaki, Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse, Hum. Mol. Genet. 16 (2007) 2272–2280.
[11] S.K. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S.P. Lin, C.D. Allis, X. Cheng, T.H. Bestor, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, Nature 448 (2007) 714–717.
[12] D. Jia, R.Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation, Nature 449 (2007) 248–251.
[13] S. Lopes, A. Lewis, P. Hajkova, W. Dean, J. Oswald, T. Forne, A. Murrell, M. Constancia, M. Bartolomei, J. Walter, W. Reik, Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions, Hum. Mol. Genet. 12 (2003) 295–305.
[14] J.Y. Shin, G.V. Fitzpatrick, M.J. Higgins, Two distinct mechanisms of silencing by the KvDMR1 imprinting control region, EMBO J. 27 (2008) 168–178.
[15] S.P. Lin, N. Youngson, S. Takada, H. Seitz, W. Reik, M. Paulsen, J. Cavaille, A.C. Ferguson-Smith, Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12, Nat. Genet. 35 (2003) 97–102.
[16] P. Szabó, S.H. Tang, A. Rentsendorj, G.P. Pfeifer, J.R. Mann, Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function, Curr. Biol. 10 (2000) 607–610.
[17] F. Sleutels, R. Zwart, D.P. Barlow, The non-coding *Air* RNA is required for silencing autosomal imprinted genes, Nature 415 (2002) 810–813.
[18] D. Mancini-Dinardo, S.J. Steele, J.M. Levorse, R.S. Ingram, S.M. Tilghman, Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes, Genes Dev. 20 (2006) 1268–1282.
[19] S. Takada, M. Tevendale, J. Baker, P. Georgiades, E. Campbell, T. Freeman, M.H. Johnson, M. Paulsen, A.C. Ferguson-Smith, *Delta-like* and *Gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12, Curr. Biol. 10 (2000) 1135–1138.
[20] S. Takada, M. Paulsen, M. Tevendale, C.E. Tsai, G. Kelsey, B.M. Cattanach, A.C. Ferguson-Smith, Epigenetic analysis of the *Dlk1-Gtl2* imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with *Igf2-H19*, Hum. Mol. Genet. 11 (2002) 77–86.
[21] S.P. Lin, P. Coan, S.T. da Rocha, H. Seitz, J. Cavaille, P.W. Teng, S. Takada, A.C. Ferguson-Smith, Differential regulation of imprinting in the murine embryo and placenta by the *Dlk1-Dio3* imprinting control region, Development 134 (2007) 417–426.
[22] M. Kagami, M.J. O'Sullivan, A.J. Green, Y. Watabe, O. Arisaka, N. Masawa, K. Matsuoka, M. Fukami, K. Matsubara, F. Kato, A.C. Ferguson-Smith, T. Ogata, The IG-DMR and the *MEG3*-DMR at human chromosome 14q32.2: hierarchical interaction and distinct functional properties as imprinting control centers, PLoS Genet. 6 (2010) e1000992.
[23] N. Takahashi, A. Okamoto, R. Kobayashi, M. Shirai, Y. Obata, H. Ogawa, Y. Sotomaru, T. Kono, Deletion of *Gtl2*, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice, Hum. Mol. Genet. 18 (2009) 1879–1888.
[24] Y. Zhou, P. Cheunsuchon, Y. Nakayama, M.W. Lawlor, Y. Zhong, K.A. Rice, L. Zhang, X. Zhang, F.E. Gordon, H.G. Lidov, R.T. Bronson, A. Klibanski, Activation of paternally expressed genes and perinatal death caused by deletion of the *Gtl2* gene, Development 137 (2010) 2643–2652.
[25] M. Stadtfeld, E. Apostolou, H. Akutsu, A. Fukuda, P. Follett, S. Natesan, T. Kono, T. Shioda, K. Hochedlinger, Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells, Nature 465 (2010) 175–181.
[26] H. Kobayashi, C. Suda, T. Abe, Y. Kohara, T. Ikemura, H. Sasaki, Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs, Cytogenet Genome Res. 113 (2006) 130–137.
[27] W. Reik, W. Dean, J. Walter, Epigenetic reprogramming in mammalian development, Science 10 (2001) 1089–1093.
[28] H.D. Morgan, F. Santos, K. Green, W. Dean, W. Reik, Epigenetic reprogramming in mammals, Hum. Mol. Genet. 14 (2005) R47–R58.
[29] J.V. Schmidt, P.G. Matteson, B.K. Jones, X.J. Guan, S.M. Tilghman, The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted, Genes Dev. 14 (2000) 1997–2002.
[30] M. Kagami, Y. Sekita, G. Nishimura, M. Irie, F. Kato, M. Okada, S. Yamamori, H. Kishimoto, M. Nakayama, Y. Tanaka, K. Matsuoka, T. Takahashi, M. Noguchi, Y. Tanaka, K. Masumoto, T. Utsunomiya, H. Kouzan, Y. Komatsu, H. Ohashi, K. Kurosawa, K. Kosaki, A.C. Ferguson-Smith, F. Ishino, T. Ogata, Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes, Nat. Genet. 40 (2008) 237–242.
[31] K. Schuster-Gossler, P. Bilinski, T. Sado, A. Ferguson-Smith, A. Gossler, The mouse *Gtl2* gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA, Dev. Dyn. 212 (1998) 214–228.
[32] S.T. da Rocha, M. Tevendale, E. Knowles, S. Takada, M. Watkins, A.C. Ferguson-Smith, Restricted co-expression of Dlk1 and the reciprocally imprinted non-coding RNA, Gtl2: implications for cis-acting control, Dev. Biol. 306 (2007) 810–823.
[33] B. Bhogal, A. Arnaudo, A. Dymkowski, A. Best, T.L. Davis, Methylation at mouse *Cdkn1c* is acquired during postimplantation development and functions to maintain imprinted expression, Genomics 84 (2004) 961–970.
[34] D. Umlauf, Y. Goto, R. Cao, F. Cerqueira, A. Wagschal, Y. Zhang, R. Feil, Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes, Nat. Genet. 36 (2004) 1296–1300.
[35] M. Ohno, N. Aoki, H. Sasaki, Allele-specific detection of nascent transcripts by fluorescence in situ hybridization reveals temporal and culture-induced changes in *Igf2* imprinting during pre-implantation mouse development, Genes Cells 6 (2001) 249–259.
[36] H. Sasaki, A.C. Ferguson-Smith, A.S. Shum, S.C. Barton, M.A. Surani, Temporal and spatial regulation of *H19* imprinting in normal and uniparental mouse embryos, Development 121 (1995) 4195–4202.
[37] R. Stöger, P. Kubicka, C.G. Liu, T. Kafri, A. Razin, H. Cedar, D.P. Barlow, Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal, Cell 73 (1993) 61–71.
[38] W. Lerchner, D.P. Barlow, Paternal repression of the imprinted mouse *Igf2r* locus occurs during implantation and is stable in all tissues of the post-implantation mouse embryo, Mech. Dev. 61 (1997) 141–149.
[39] Y. Sekita, H. Wagatsuma, M. Irie, S. Kobayashi, T. Kohda, J. Matsuda, M. Yokoyama, A. Ogura, K. Schuster-Gossler, A. Gossler, F. Ishino, T. Kaneko-Ishino, Aberrant regulation of imprinted gene expression in *Gtl2*^{lacZ} mice, Cytogenet Genome Res. 113 (2006) 223–229.
[40] E.Y. Steshina, M.S. Carr, E.A. Glick, A. Yevtodiyenko, O.K. Appelbe, J.V. Schmidt, Loss of imprinting at the *Dlk1-Gtl2* locus caused by insertional mutagenesis in the *Gtl2* 5′ region, BMC Genet. 7 (2006) 44.
[41] T. Fan, J.P. Hagan, S.V. Kozlov, C.L. Stewart, K. Muegge, Lsh controls silencing of the imprinted *Cdkn1c* gene, Development 132 (2005) 635–644.
[42] J. Mager, N.D. Montgomery, F.P. de Villena, T. Magnuson, Genome imprinting regulated by the mouse Polycomb group protein Eed, Nat. Genet. 33 (2003) 502–507.
[43] R. Margueron, N. Justin, K. Ohno, M.L. Sharpe, J. Son, W.J. Drury III, P. Voigt, S.R. Martin, W.R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg, S.J. Gamblin, Role of the Polycomb protein EED in the propagation of repressive histone marks, Nature 461 (2009) 762–767.
[44] S.S. Hammoud, D.A. Nix, H. Zhang, J. Purwar, D.T. Carrell, B.R. Cairns, Distinctive chromatin in human sperm packages genes for embryo development, Nature 460 (2009) 473–478.
[45] M.S. Carr, A. Yevtodiyenko, C.L. Schmidt, J.V. Schmidt, Allele-specific histone modifications regulate expression of the *Dlk1-Gtl2* imprinted domain, Genomics 89 (2007) 280–290.
[46] A. Lewis, K. Mitsuya, D. Umlauf, P. Smith, W. Dean, J. Walter, M. Higgins, R. Feil, W. Reik, Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation, Nat. Genet. 36 (2004) 1291–1295.
[47] A. Lewis, K. Green, C. Dawson, L. Redrup, K.D. Huynh, J.T. Lee, M. Hemberger, W. Reik, Epigenetic dynamics of the *Kcnq1* imprinted domain in the early embryo, Development 133 (2006) 4203–4210.
[48] M. Ariel, J. McCarrey, H. Cedar, Methylation patterns of testis-specific genes, Proc. Natl. Acad. Sci. U S A 88 (1991) 2317–2321.

Human Molecular Genetics, 2011, Vol. 20, No. 16 **3188–3197**
doi:10.1093/hmg/ddr224
Advance Access published on May 18, 2011

Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes†

Kazuhiko Nakabayashi[1,‡,¶], Alex Martin Trujillo[3,¶], Chiharu Tayama[1], Cristina Camprubi[3], Wataru Yoshida[1], Pablo Lapunzina[4], Aurora Sanchez[5], Hidenobu Soejima[6], Hiroyuki Aburatani[7], Genta Nagae[7], Tsutomu Ogata[2], Kenichiro Hata[1] and David Monk[3,*,‡]

[1]Department of Maternal-Fetal Biology and [2]Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [3]Cancer Epigenetic and Biology Program (PEBC), Institut d'Investigació Biomedica de Bellvitge (IDIBELL), Hospital Duran i Reynals, Barcelona, Spain, [4]Instituto de Genética Médica y Molecular (INGEMM), CIBERER, IDIPAZ-Hospital Universitario La Paz, Universidad Autónoma de Madrid, Madrid, Spain, [5]Servei de Bioquímica i Genètica Molecular, CIBER de Enfermedades Raras, and Institut d'Investigacions Biomèdiques August Pi i Sunyer, Hospital Clínic, Barcelona, Spain, [6]Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga 849-8501, Japan and [7]Genome Science Division, Research Center for Advanced Science and Technology, the University of Tokyo, Tokyo 153-8904, Japan

Received April 11, 2011; Revised and Accepted May 13, 2011

Nuclear transfer experiments undertaken in the mid-80's revealed that both maternal and paternal genomes are necessary for normal development. This is due to genomic imprinting, an epigenetic mechanism that results in parent-of-origin monoallelic expression of genes regulated by germline-derived allelic methylation. To date, ~100 imprinted transcripts have been identified in mouse, with approximately two-thirds showing conservation in humans. It is currently unknown how many imprinted genes are present in humans, and to what extent these transcripts exhibit human-specific imprinted expression. This is mainly due to the fact that the majority of screens for imprinted genes have been undertaken in mouse, with subsequent analysis of the human orthologues. Utilizing extremely rare reciprocal genome-wide uniparental disomy samples presenting with Beckwith–Wiedemann and Silver–Russell syndrome-like phenotypes, we analyzed ~0.1% of CpG dinculeotides present in the human genome for imprinted differentially methylated regions (DMRs) using the Illumina Infinium methylation27 BeadChip microarray. This approach identified 15 imprinted DMRs associated with characterized imprinted domains, and confirmed the maternal methylation of the *RB1* DMR. In addition, we discovered two novel DMRs, first, one maternally methylated region overlapping the *FAM50B* promoter CpG island, which results in paternal expression of this retrotransposon. Secondly, we found a paternally methylated, bidirectional repressor located between maternally expressed *ZNF597* and *NAT15* genes. These three genes are biallelically expressed in mice due to lack of differential methylation, suggesting that these genes have become imprinted after the divergence of mouse and humans.

*To whom correspondence should be addressed. Tel: +34 932607500 ext. 7128; Fax: +34 2607219; Email: dmonk@idibell.cat
†Methylation array data: the data from the Illumna Infinium Human Methylation27 BeadChip microarray has been deposited with GEO database, accession number GSE28525.
‡Co-corresponding author. Tel: +81-3-3416-0181; Fax: +81-3-3417-2864; Email: knakabayashi@nch.go.jp
¶These authors contributed equally to this work.

INTRODUCTION

Genomic imprinting is an epigenetic process in which one allele is repressed, resulting in parent-of-origin specific monoallelic expression (1). To date, around 100 imprinted transcripts have been identified in mouse, including protein coding genes, long non-coding RNAs (ncRNA) and microRNAs. Approximately two-thirds show conserved imprinting status between mouse and humans, while some show imprinting restricted to humans (http://igc.otago.ac.nz/home.html).

Genomic imprinting is regulated by epigenetic modifications such as DNA methylation, along with repressive histone modifications that are transmitted through the gametes from the parental germlines (1). Many imprinted regions contain differentially methylated regions (DMRs) that exhibit parent-of-origin-dependent DNA methylation. Of the 21 known DMRs in mouse, a subset have been shown to function as *cis*-acting imprinting control regions (ICRs) orchestrating the monoallelic expression of genes over more than 100 kbp away (2). The establishment of imprinted methylation in both the maternal and paternal germlines requires the *de novo* DNA methyltransferae Dnmt3a and its related protein Dnmt3L (3,4). Maintenance of these DMRs is stable throughout somatic development and is regulated by Dnmt1 and Uhrf1 during DNA replication (5,6).

The identification of novel imprinted genes is important as it is becoming increasing evident that alterations in the fine-tuning of imprinted gene expression can influence a number of complex diseases such as obesity, diabetes, neurological diseases and cancer (7–9), in addition to the well-defined imprinting syndromes associated with severe disruption of imprinted domains.

The identification of imprinted genes has traditionally been performed in mouse owing to the ease of embryo and genetic manipulations, and has utilized gynogenetic and androgenetic embryos, or mice harboring regions of uniparental disomy (UPD), where two copies of an entire chromosome or chromosomal region is inherited from only one parent (reviewed in 10). These embryos have then been used in expression screen-based approaches such as subtractive hybridization, differential display or expression array hybridization (11,12). However, these screens are not deemed comprehensive, as imprinted gene expression can be both tissue- and developmental-stage specific. Previously, sophisticated screens have detected allelic differences in DNA methylation at imprinted DMRs present in all somatic tissues, irrespective of temporal and spatial expression. Techniques such as restriction landmark genomic screening, methylation-sensitive representation difference analysis (Me-RDA) and methylated DNA immunoprecipitation (MeDIP) have identified regions of allelic DNA methylation associated with chromosomal regions controlling several imprinted genes in mice (13–15) and humans (16,17).

In order to identify novel imprinted genes in humans, we have performed a quantitative genome-wide methylation screen comparing the methylomes of three-genome-wide paternal UPD (pUPD) samples identified with Beckwith–Wiedemann-like phenotypes and one genome-wide maternal UPD (mUPD) Silver–Russell-like syndrome case (18–21) with the methylomes of six normal somatic tissues. The genome-wide UPD samples were all mosaic, and we utilized DNA extracted from leukocytes as these presented with lowest level of the biparental cell line. The DNA methylation profiles of these samples only differ at imprinted DMRs, since they are all derived from leukocytes, making them ideal to screen for novel imprinted loci. We utilized the Illumina Infinium Human Methylation27 BeadChip microarray and were able to identify 15 imprinted DMRs associated with known imprinted transcripts, and confirm the allelic methylation within intron 2 of the *RB1* gene (22).

By comparing the methylation profiles of six somatic tissues and the genome-wide UPD cases, we identified a novel paternally methylated DMR which acts as a directional silencer resulting in the maternal expression of *ZNF597* (also known as *FLJ33071*) and *NAT15* on chromosome 16, and a maternally methylated DMRs encompassing the promoter region of the *FAM50B* retrotransposon on chromosome 6, which is paternally expressed in human tissues. Interestingly, the CpG islands of the mouse orthologues of *ZNF597*, *NAT15* and *FAM50B* are all unmethylated, resulting in biallelic expression in mid-gestation embryonic tissues.

RESULTS

Defining a hemimethylated data set

Almost all imprinted domains contain at least one region of allelic DNA methylation which is thought to regulate imprinting *in cis* (1). In order to identify new imprinted genes in humans, we performed a methyation screen of six different normal somatic tissues derived from the three germinal layers (placenta, leukocytes, brain, muscle, fat, buccal cells) and compared the data set with the methylation profiles from reciprocal genome-wide UPD samples. Genomic DNA was modified by sodium bisulfite treatment and hybridized to the Illumina Infinium Human Methylation27 platform. This array covers 27 578 CpG dinucleotides associated with 14 000 human genes. To identify novel imprinted DMRs, we took advantage of the fact that these CpG-rich sequences have a methylation profile of ~50% in all somatic tissues. We identified 78 CpG probes associated with 15 known imprinted DMRs on the array (average methylation 52%, SD 11.7) (Supplementary Material, Fig. S1). To define a range in which novel imprinted DMRs should lie, we used the mean for the known imprinted DMR ± 1.5 SD (range 34.4–69.6). After applying these defined cutoffs, we identified 3212 CpG probes for which the mean methylation value for all normal tissues was within this range. To rule out the possibility that a mean of ~52% was caused by extreme values of hyper- and hypomethylation as a result of tissue-specific methylation, we only assessed those within 1.8 times SD distance from the methylation average. This step ensures that the ~52% methylation value is representative of all tissues. Using these criteria, we reduced the data set to 1836 CpG probes, which were in addition to 72 probes mapping to known imprinted DMRs.

Determining the allelic methylation using genome-wide UPDs

To identify novel imprinted DMRs within the above hemimethylated data set outlined earlier, we compared the tissue methylation profiles to those obtained for the samples with genome-wide UPD. Of the 1836 CpG probes, only 14 gave methylation profiles consistent with an imprinted profile (Supplementary Material, Fig. S2). We subsequently mapped the exact location of the candidate CpGs using the genomic sequence of the unconverted DNA probes in the BLAT search tool (UCSC Genome Bioinformatics http://genome.ucsc.edu/). These 14 CpG probes were located close to nine autosomal genes, *RB1* (5), *FAM50B* (2), *ZNF597* (1), *TRPC3* (1), *SYCE1* (2), *TSP50* (1), *SORD* (1) and *ZBTB16* (1). We identified five independent probes located throughout CpG 85 (the CpG island identifier on the UCSC genome browser, build GRCh37/hg19) of the recently identified *RB1* imprinted gene on chromosome 13. These probes were unmethylated with average β-values of 0.21, 0.17 and 0.18 in the three genome-wide pUPD samples but hypermethylated, having an average β-value 0.88, in the genome-wide mUPD sample (a complete unmethylated CpG has a β-value of 0, and a fully methylated dinucleotide being 1). Using bisulphite PCR incorporating the single-nucleotide polymorphism (SNP) rs2804094 and sequencing of individual DNA strands, we were able to confirm that this 1.2 kb CpG island is a maternally methylated DMR in placenta, leukocyte and kidney-derived DNA and unmethylated in sperm (Supplementary Material, Fig. S3).

We identified one probe was located close to CpG 55 of the *TRPC3* gene on human chromosome 4 that was suggestive of a maternally methylated DMR. Subsequent allelic bisulphite PCR encompassing the SNP rs13121031 revealed that this region was subject to SNP-associated methylation and not parent-of-origin methylation (data not shown). The CpG islands within the promoters of *ZBTB16*, *TSP50* and *SORD* each had one probe that was suggestive of imprinted methylation, however allele-specific bisulphite PCR analysis revealed that these regions had a mosaic methylated profile (data not shown).

Two probes mapping to CpG 124 of *SYCE1/SPRNP1* on chromosome 10 also had a methylation profile consistent with an imprinted DMR. However, these probes were unable to discriminate *SYCE1* from *SPRN*, a second region that shared 93% homology. Due to the difficulty in designing bisulphite PCR primers that could specifically target *SYCE1*, we were unable to validate our initial observations.

The *ZNF597/NAT15* CpG island is a paternally methylated DMR

To date, only seven paternally methylated DMRs have been identified, the somatic DMRs at the *NESP, IGF2-P0* and *MEG3/GTL2* promoters, the germline *H19* differentially methylated domain (DMD), *Rasgrf1* DMD, IG-DMR and *ZDBF2* DMR (15,23–26). The *RASGRF1* is not imprinted in humans due to lack of the DNA repeat elements that are involved in establishing germline methylation (27). We identify two CpG probes, one mapping to CpG 41 between the promoters of *ZNF597* and *NAT15*, the other 500 bp away, in a region flanking CpG 41. Both probes were hypermethylated in the three genome-wide pUPD samples (β-values of 0.83, 0.42, 0.75) and hypomethylated (*β*-value of 0.08) in the genome-wide mUPD sample. Using bisulphite PCR and subsequent sequencing of heterozygous DNA samples for the SNP rs2270499, we were able to confirm that the methylation was solely on the paternally derived allele in placenta, leukocyte and kidney (Fig. 1). This is consistent with the previous report that *ZNF597* is maternally expressed in human leukocytes (28). Bisulphite PCR and sequencing of sperm DNA revealed that this region lack methylation, indicating that CpG41 is not a germline DMR. Using allele-specific RT–PCR that incorporated coding SNPs within exon 3, we observed maternal expression in brain ($n = 1$) and placenta ($n = 3$), and confirmed imprinting in leukocytes ($n = 2$).

The gene encoding *N*-acetyltransferase 15, *NAT15*, is encoded by two different transcripts (Fig. 1A). To determine whether *NAT15* is also subject to genomic imprinting, we performed allelic RT–PCR using PCR primers that could discriminate each isoform. We find that *NAT15* isoform 1 is maternally expressed in both placenta ($n = 5$) and leukocytes ($n = 1$), whereas isoform 2 is biallelically expressed ($n = 4$) which is consistent with CpG 101 being unmethylated (Fig. 1, data not shown).

FAM50B DMR shows graduated methylation

We identified two probes mapping to a 1.7 kbp CpG island within the *FAM50B* promoter. These probes were hypermethylated in the genome-wide mUPD (average *β*-values of 0.86), but hypomethylation in the three pUPD samples mUPD (*β*-value of 0.23, 0.39, 0.31). Allelic bisulphite sequencing showed that the methylation profile of CpG 143 differs between the 5′ and 3′ ends. The 5′ region flanking the SNP rs2239713, overlapping the *FAM50B* promoter, is a maternally methylated DMR in placenta-, leukocyte- and kidney-derived DNA, while the 3′ region near rs34635612 is fully methylated on both parental alleles. Despite this methylation gradient, the *FAM50B* gene is paternally expressed in placenta ($n = 6$) (Fig. 2).

The absence of allelic methylation at the mouse orthologues *of ZNF597, NAT15* and *FAM50B* is associated with biallelic expression

To determine whether the allelic expression of the novel imprinted transcripts was conserved in mouse, we investigated the allele-specific expression using RT–PCR amplification across transcribed SNPs. Mouse tissues were derived from interspecies crosses at both embryonic day E9.5 and post-natal day 1. The *Fam50b* gene has two isoforms with alternative first exons. We could only detect expression in testis, which was derived from both parental alleles. Exon 2 of Fam50b corresponds to an X-chromosome-derived retrogene and overlaps a methylated CpG island.

The *Nat15* and *Znf597* genes share two different promoter CpG islands, CpG 35 and CpG 87 that are orthologous to the *ZNF597* DMR and the *NAT15* isoform 2 promoters, respectively. In mouse, both of these regions are unmethylated. Both *Nat15* isoforms are predominantly expressed in

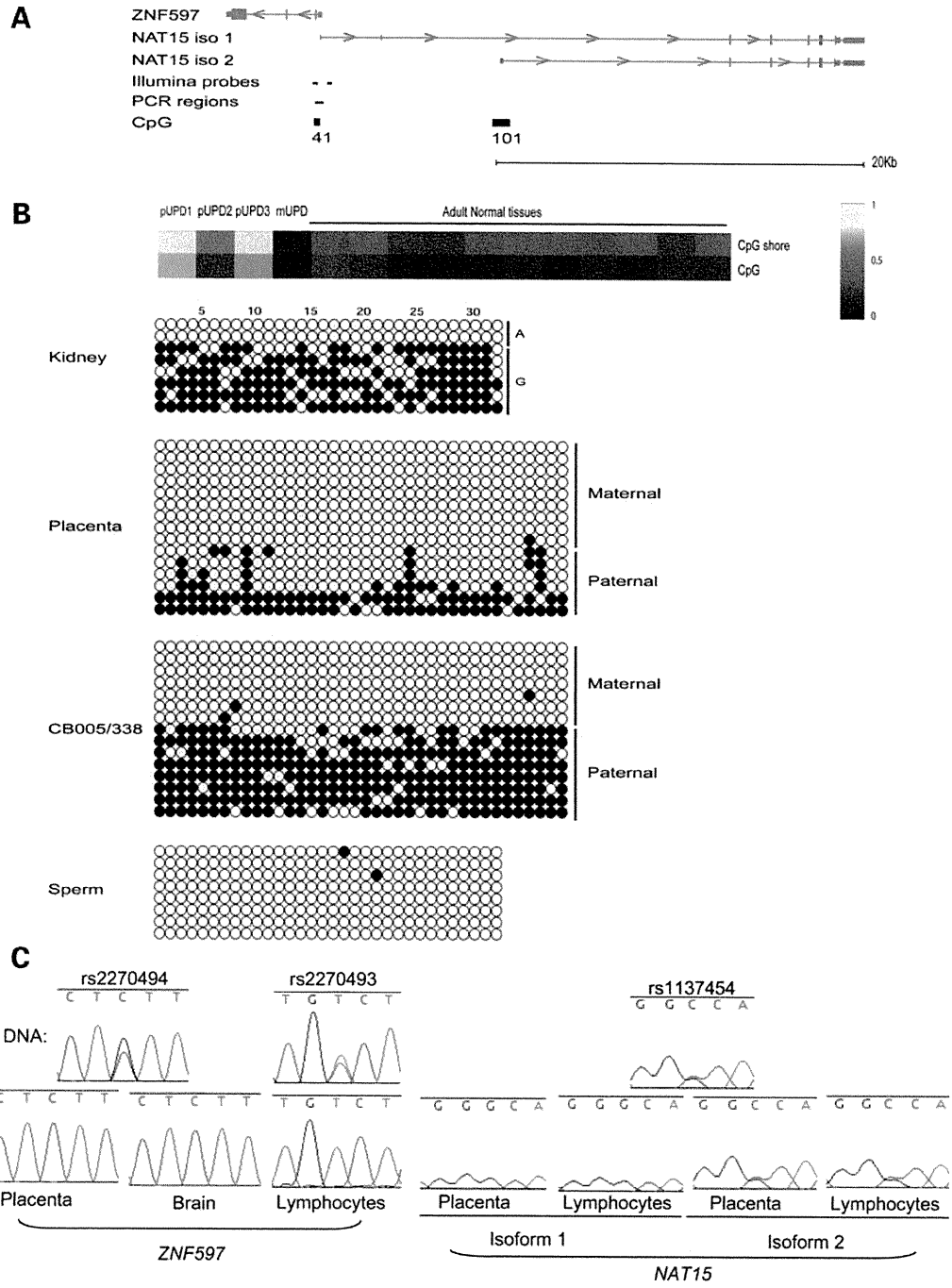

Figure 1. (**A**) Map of the *ZNF597-NAT15* locus on human chromosome 16, showing the location of the various transcripts, CpG islands, Illumina probes and bisulphite PCR regions (red transcripts are maternally expressed, blue paternally expressed and grey are expressed from both parental alleles. Arrows represent the direction of transcription) (not drawn to scale). (**B**) Heat map of the Infinium HumanMethylation27 BeadChIP for the *ZNF597* CpG probes (cg24333473 in CpG island; cg14654875 in CpG shore), with confirmation of allelic methylation in kidney, placenta and cord blood derived DNA. Each circle represents a single CpG dinucleotide and the strand, a methylated cytosine (filled circle) or an unmethylated cytosine (open circle). The same region was analyzed in sperm-derived DNA. (**C**) The sequence traces show allelic expression for the *ZNF597* and *NAT15* genes.

brain and testis, which is equally derived from both parental alleles. The variants of *Znf597* were expressed in E9.5 whole embryo, yolk sac and placenta, and in individual tissues later in development. Allelic expression analysis revealed that these transcripts were not imprinted, with equal expression from both parental chromosomes (Fig. 3).

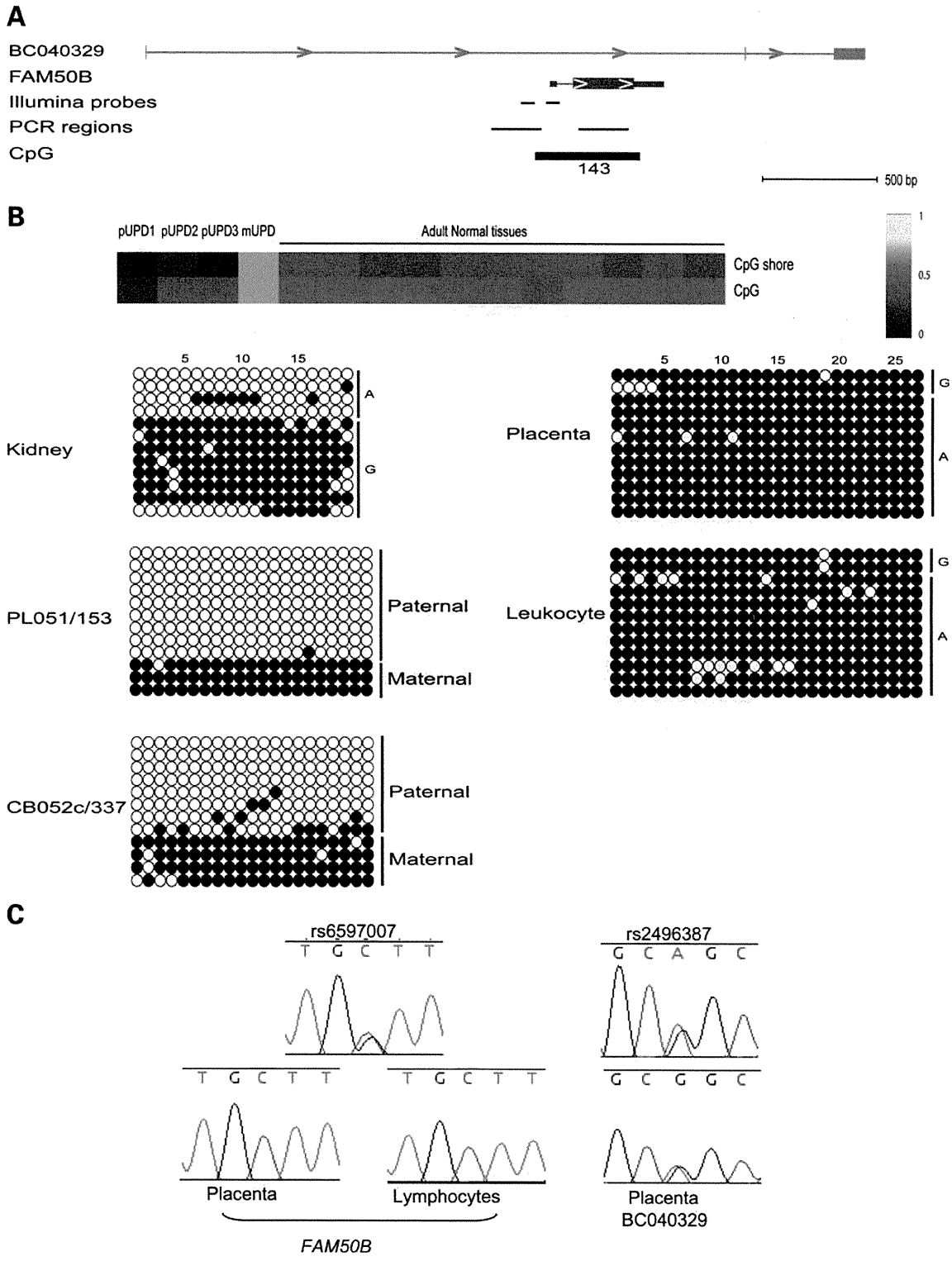

Figure 2. (**A**) A map of the *FAM50B/BC040329* locus, with the location of the of the CpG island (not to scale). (**B**) Heat map for CpG probes mapping to the *FAM50B* promoter (cg01570885; cg03202897) and the subsequent analysis of allelic methylation in various tissues. The methylation profiles on the left are from the 5′ CpG island region, while those on the right are from the 3′ region. (**C**) The allelic expression of *FAM50B* and the host gene in term placenta and leukocytes.

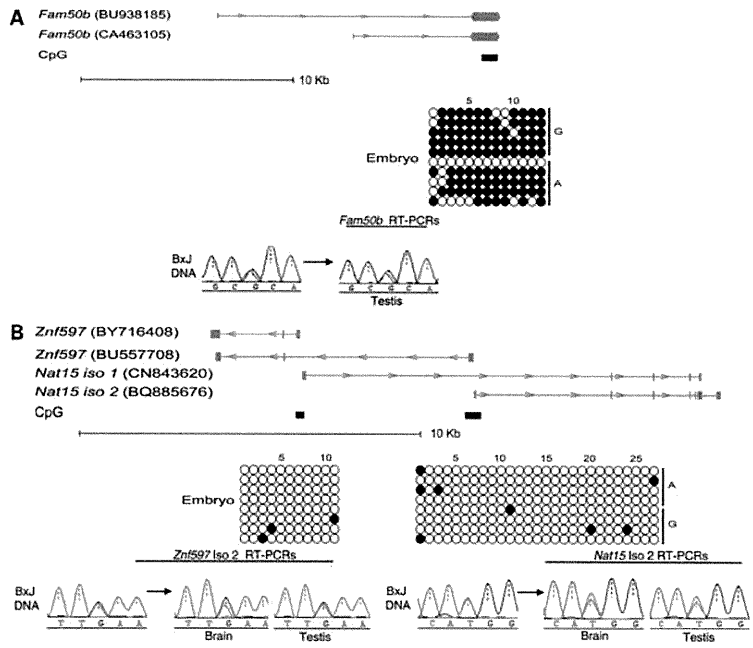

Figure 3. Schematic maps of the *Fam50b* (**A**) and *Znf597/Nat15* genes (**B**), with the location of the alternative promoter regions. The methylation status of the orthologous CpG islands associated with each domain was examined in embryo-derived DNA. The allelic expression of each gene in various mouse tissues from reciprocal mouse crosses. For clarity, only the expression in B6 × JF1 tissues is shown.

DISCUSSION

Identification of new human imprinted genes requires screening human samples

Most screens for new imprinted genes are undertaken in mouse with subsequent confirmation of the imprinting status of the human orthologues. Despite the success, this approach will not identify imprinted loci specifically imprinted in humans. To date, very few imprinted genes are human-specific, however, these rare transcripts do exist as highlighted by the paternally expressed *L3MBTL, C19MC* and *RB1* genes (22,29,30). Using DNA from Beckwith–Wiedemann and Silver–Russell-like phenotypes with reciprocal genome-wide UPDs, we have performed a comprehensive screen of ~0.1% of the human methylome. Despite the extensive coverage of Illumina Infinium Human Methylation27 BeadChip microarray, we identified very few novel imprinted loci. However, it must be noted that paternal germline DMRs are not associated with CpG islands, and therefore maybe remote from gene promoters and promoter CpG islands present on the array.

The predicted number of imprinted genes varies with estimates from 200–2000 transcripts in mouse, with one transcriptome-wide analysis, using the ultra sensitive RNA-seq technology, identifying over 1000 transcripts in brain with parent-of-origin expression bias (31). Recent studies have predicted and experimentally verified imprinted genes based on sequence and epigenetic characteristics. For example, human imprinted regions significantly lack short interspersed transposable elements in comparison with the rest of the genome and are associated with CpG islands (32,33). Using a bioinformatics approach, Luedi *et al.* (34) predicted 156 imprinted genes in humans based on similarity with known imprinted transcripts, confirming the maternal expression of *KCNK9*. In addition, the paternally expressed *MCTS2* gene was identified through a hypothesis-driven search for intronic X-chromosome-derived retrotransposons that are associated with CpG island promoters (35). Interestingly, *FAM50B* is also an imprinted X-chromosome-derived retrogene gene and was correctly identified by Luedi *et al.* (34) during their computational screening and the imprinting status recently confirmed (36).

We wished to identify additional imprinted loci based on data generated in previously published analyses. We have compared our hemimethylated data set against the 156 bio-informatically predicted imprinted genes and the 82 candidates predicted due to unequal representation of alleles in public EST libraries and expression genotype arrays (37,38). We found that fifteen out of one hundred and fifty-six and nine out of eighty-two, respectively, were present in our data set. However, none of these additional genes had a methylation profile consistent with an imprinted DMR, highlighting the high false-positive rates of bioinformatic predictions (Supplementary Material, Fig. S4). From our observations, we predict that the majority of human DMRs overlapping promoters have been identified. Following analysis of more than 14 000 genes, we identified only two new imprinted DMRs. Extrapolating this trend to the 34 702 annotated RefSeq genes, we predict that there will be around five additional unidentified DMRs in the human genome, resulting in a total of ~35.

Parent-of-origin DNA methylation is not the only epigenetic signature associated with imprinted DMRs (reviewed in 9). Recently, a chromatin signature has been shown to mark imprinted DMRs; with trimethylation of lysine 9 of histone

H3 (H3K9me3) and trimethylation of lysine 20 of histone H4 (H4K20me3) associated with the DNA methylated allele (39), while the unmethylated allele is enriched for the transcriptionally permissive Lysine 4 methylation of histone H3 (H3K4me2/3) (40). The combination of differential DNA methylation between sperm and somatic tissues and an overlapping H3K9me3 and H3K4me3 signature has recently been used to identify 11 new candidate DMRs in mouse (41). With the availability of human ChIP-seq derived genome-wide data sets for most histone modifications (42,43), it would be interesting to determine if this histone signature recognized in mouse can be used to identify novel human imprinted DMRs. Interrogation of the NHLBI ChIP-seq data set (http://dir.nhlbi.nih.gov/papers/lmi/epigenomes/hgtcell.aspx) revealed that the *RB1*, *ZNF597* and *FAM50B* DMRs are enriched for both H3K4me3 and H3K9me3, with the later two regions harboring functional CTCF binding sites (data not shown).

The regulation of imprinted domains on human chromosomes 13 and 16

The *RB1* DMR has previously been proposed to contain the promoter of the paternally expressed *E2B-RB1* isoform (22). We were unable to identify coding SNPs within the *RB1* gene that would allow us to determine the allelic expression in our cohort of tissues. However, we were able to show that the *LPAR6* gene, encoding lysophosphatidic acid receptor 6 located in intron 16 of *RB1* is biallelically expressed, suggesting that the *RB1* DMR does not influence the expression of this gene (Supplementary Material, Fig. S3).

The maternal expression of *ZNF597* has previously been shown in leukocytes (28). Here, we show that the *ZNF597* DMR acts as a bidirectional silencer, which orchestrates the paternal silencing of *ZNF597* and *NAT15*. This organization is reminiscent of *PEG10-SGCE* domain on human 7p22 (44). We did not observe methylation in DNA isolated from mature sperm, which suggests that this region acquires methylation during early somatic development (Fig. 1). All known somatic DMR are associated with nearby germline DMRs, which regulate the methylation in a hierarchical fashion (23,45,46), implying a yet to be identified germline DMR is situated within the vicinity of the *ZNF597* gene.

The maternally expressed *NAT15* is a highly conserved protein coding gene with two alternative first exons, with only isoform I subject to imprinting. In addition, there is evidence from EST libraries of an ncRNA (genbank: DA387972) that originates from the *NAT15* isoform 1 promoter and continues past the exon–intron splice site to produce a ~550 bp transcript. Unfortunately, we were unable to detect expression of this transcript in our tissue set, so we could not determine if this ncRNA is imprinted.

FAM50B is an imprinted retrogene

Sequence analysis revealed that the *FAM50B* transcript (previously named *X5L*) is a retrotransposon that originated from *FAM50A/XAP5* within Xq28. Unlike other classical retrogenes, this gene has an intron in the 5′ UTR in both humans and mouse, which has no counterpart in its parental gene. It is likely that the intron was inserted after retroposition, possibly during recruitment of a functional promoter region (47). Interestingly, several other imprinted genes have been shown to originate from retrotransposition from the X-chromosome genes (35,48). *FAM50B* is ubiquitously expressed, and is inserted within the intron of a host transcript *BC040329*, which is predominantly expressed in testis with low detection in brain and placenta (data not shown). This host gene is biallelically expressed in placenta ($n = 7$) (Fig. 2), of which two samples exhibited imprinted expression of *FAM50B*.

Discrepancy between imprinted DMR methylation screens

The quantitative methylation values obtained using the Illumina Infinium platform makes it suitable for comparing reference and test samples. This approach has previously been used to screen for imprinted DMRs using paternally derived androgenetic complete hydatidiform moles versus maternally derived mature cystic ovarian teratomas and in patients with maternal hypomethylation syndrome (24,49). In both cases, the genetic material analyzed is not ideally suited for comprehensive screening for novel imprinted loci. This is because it is currently unknown to what extent the DNA methylation profile is altered in ovarian teratomas, and any differences maybe due to the uniparental nature of the sample or tumorgenic changes, and candidates obtained from comparisons with complete hydatidiform moles may simply reflect tissue-specific differences. This is highlighted by the fact that of the 95 candidate probes identified by Choufani *et al.* (49), 68 overlapped with our hemimethylated data set (Supplementary Material, Fig. S4) with only *ZNF597* DMR being identified in both screens. These authors also suggest that *AXL*-promoter region is a DMR, but this was not identified using our genome-wide UPDs, and bisulphite PCR and sequencing of our samples revealed a non-allelic mosaic methylation profile (Supplementary Material, Fig. S5). In addition, the methylation profiles obtained from comparing normal and maternal hypomethylation samples will only facilitate the identification of a subset of imprinted DMRs, since *ZFP57* mutations do not effect the maintenance of all maternally methylated imprinted DMRs equally (50,51).

Functional relevance of the new imprinted domains

Very little is known about the role of *FAM50B*, *ZNF597* and *NAT15*, with no previous publications describing functional studies. The three new imprinted regions we identify all map to chromosomes for which recurrent chromosomal UPDs have been reported. With the exception of pUPD and the overexpression of *PLAGL1/HYMAI* in Transient Neonatal Diabetes Mellitus, the UPDs for these chromosomes are not associated with obvious developmental phenotypes and most cases were identified because of the unmasking of mutant recessive alleles (reviewed in 52,53).

CONCLUSIONS

Our study has assisted in defining a comprehensive catalog of human imprinted genes. The use of extremely rare reciprocal

genome-wide UPD samples in unbiased methylation screens such as bisulphite genome sequencing will aid the identification of additional imprinted loci, which will facilitate study of genetic diseases associated with aberrant imprinting. The general trend until now has been that, while imprinted genes play an important role in fetal development and behavior, evolutionary forces dictated by the genetic conflict have allowed for a lack of conserved imprinting between mouse and humans (54). However, our screen has identified new human-specific imprinted transcripts, all of which have conserved gene orthologues in many taxa. These genes have selected imprinting as a mechanism of transcriptional regulation in humans despite the risk of being functional hemizygous.

MATERIALS AND METHODS

The human reciprocal genome-wide UPD samples

Genomic DNA isolated from three previously described Beckwith–Weidemann syndrome-like cases (16–18) and one Silver–Russell syndrome-like patient (19) was used in this study. Each of these cases had undergone extensive molecular characterization to confirm genome-wide UPD status and level of mosaicism. We used DNA isolated from leukocytes as these samples had minimal mosaicism of a biparental cell line. The genome-wide BWS samples had 9, 11 and 15% biparental contribution, whereas the genome-wide SRS sample had 16%.

Human tissues

Two independent tissue collections were used in this study. All tissues were collected after obtaining informed consent. The Spanish collection was from the Hospital St Joan De Deu tissue cohort (Barcelona, Spain). Normal peripheral blood was collected from adult volunteers aged between 19 and 60 years old. A selection of normal brain samples was obtained from BrainNet Europe/Barcelona Brain Bank. The Japanese tissues were collected at the National Center for Child Health and Development (Tokyo, Japan) and at the Saga University Hospital.

DNA was extracted using either the standard phenol/chloroform extraction method or the QIAamp DNA Blood Midi Kit (Qiagen). RNA was extracted using either Trizol (Invitrogen) or Sepasol®-RNA I Super G (Nacalai Tesque) and cDNA synthesis was carried out as previously described (54). Ethical approval for this study was granted by the Institutional Review Boards at the National Center for Child Health and Development and Saga University and Hospital St Joan De Deu Ethics Committee (Study number 35/07) and IDIBELL (PR006/08).

Cell lines and mouse crosses

Wild-type mouse embryos and placentas were produced by crossing C57BL/6 with *Mus musculus molosinus* (JF1) mice. C57BL/6 (B6) mice were purchased from Sankyo Labo Service Corporation, Inc. (Tokyo, Japan) and JF1/Ms (JF1) mice were obtained from the Genetics Strains Research Center at the National Institute of Genetics, Japan. All animal husbandry and breeding was approved and licensed by the National Research Institute for Child Health and Development, Japan (Approved number A2010–002).

Illumina Infinium methylation27 BeadChip microarray analysis

Approximately 1 μg DNA from the reciprocal genome-wide UPDs, placenta, leukocytes, brain, muscle, fat, buccal cells was subjected to sodium bisulphite treatment and purified using the EZ GOLD methylation kit (ZYMO, Orange, CA, USA). This DNA was then hybridized to the Illumna Infinium Human Methylation27 BeadChip microarray either at the Centro Nacional de Investigaciones Oncologicas (Madrid, Spain) or Genome Science Division, Research Center for Advanced Science and Technology (University of Tokyo, Japan) using Illumina-supplied reagents and protocols. The loci included on this array and the technologies behind the platform have been described previously (55). Before analyzing the methylation data, we excluded possible sources of technical biases that could alter the results. We discarded 109 probes because they had a false-positive rate >0.1. We also excluded 261 probes because of the lack of signal in one of the 11 DNA samples analyzed. Lastly, prior to screening for novel imprinted DMRs, we excluded all X chromosome CpG sites. Therefore, in total we analyzed 26 152 probes in all DNA samples. All hierarchical clustering and β-value evaluation was performed using the Cluster Analysis tool of the BeadStudio software (version 3).

Allelic methylation analysis

A panel of placenta-, leukocyte-, brain- and kidney-derived DNAs were genotyped to identify heterozygous samples. These DNA were converted using the EZ GOLD methylation kit. Approximately 100 ng of converted DNA was used for each bisulphite PCR. Bisulphite-specific primers (Supplementary Material, Table S1) which incorporate the SNPs were used with Hotstar Taq polymerase (Qiagen, West Sussex, UK). Amplifications were performed using either 45 cycles or a nested PCR using 35 cycles for each round. The subsequent PCR products were cloned into pGEM-T Easy vector (Promega) for subsequent sequencing.

Allelic expression analysis

Genotypes on DNA were obtained for exonic SNPs identified in the UCSC browser (NCBI36/hg18, Assembly 2006) by PCR. Sequences were interrogated using Sequencher v4.6 (Gene Codes Corporation, MI) to distinguish informative heterozygote samples. Informative samples were analyzed by RT–PCR. All primers, with the exception of those targeting *FAM50B*, are intron-crossing and incorporated the heterozygous SNP in the resulting amplicon (Supplementary Material, Table S1). RT–PCRs were performed using cycle numbers determined to be within the exponential phase of the PCR, which varied for each gene, but was between 32 and 40 cycles.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Isabel Iglesias Platas for supplying DNA/cDNA from the Hospital Sant Joan de Deu placenta cohort and Professor Isidro Ferrer of the Barcelona Brain Bank for supplying human brain specimens. DNAs from normal adult tissues were a kind gift from Manel Esteller. We also thank Hiroko Meguro for technical assistance with the methylation array. We are especially grateful to Jose Martin-Subero for help and advice with the methylation array data analysis.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Spanish Ministerio de Educación y Ciencia (SAF2008–1578 to D.M.); the Asociación Española Contra el Cancer (to D.M.); Fundació La Marató de TV3 (101130 to D.M. and P.L.); the Japan Society for the Promotion of the Science (to K.N., T.O., K.H.); the National Center for Child Health and Development of Japan (Grant 20C-1 to K.N., and Grant 22C-7 to K.H.). D.M. is a Ramon y Cajal research fellow (RYC-04548).

REFERENCES

1. Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.*, **2**, 21–32.
2. Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G. and Sasaki, H. (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development, 138*, 811–820.
3. Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536–2539.
4. Hata, K., Okana, M., Lei, H. and Li, E. (2002) Dnmt3L cooperates with Dnmt3 family of de novo methyltransferases to establish maternal imprints in mice. *Development*, **129**, 1983–1993.
5. Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R. and Sasaki, H. (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.*, **22**, 1607–1616.
6. Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K. *et al.* (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908–912.
7. Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A. *et al.* (2009) Parental origin of sequence variants associated with complex diseases. *Nature*, **462**, 868–874.
8. Davies, W., Isles, A.R. and Wilkinson, L.S. (2005) Imprinted gene expression in the brain. *Neurosci. Biobehav. Rev.*, **29**, 421–430.
9. Monk, D. (2010) Deciphering the cancer imprintome. *Brief Funct. Genomics*, **9**, 329–339.
10. Henckel, A. and Arnaud, P. (2010) Genome-wide identification of new imprinted genes. *Brief Funct. Genomics*, **9**, 304–314.
11. Kaneko-Ishino, T., Kuroiwa, Y., Miyoshi, N., Kohda, T., Suzuki, R., Yokoyama, M., Viville, S., Barton, S.C., Ishio, F. and Surani, M.A. (1995) Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat Genet.*, **11**, 52–59.
12. Kuzmin, A., Han, Z., Golding, M.C., Mann, M.R., Latham, K.E. and Varmuza, S. (2008) The PcG gene Sfmbt2 is paternally expressed in extraembryonic tissues. *Gene Expr. Patterns*, **8**, 107–116.
13. Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M., Komatsubara, H. *et al.* (1994) Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nat. Genet.*, **6**, 33–40.
14. Kelsey, G., Bodle, D., Miller, H.J., Beechey, C.V., Coombes, C., Peters, J. and Williamson, C.M. (1999) Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. *Genomics*, **62**, 129–138.
15. Hiura, H., Sugawara, A., Ogawa, H., John, R.M., Miyauchi, N., Miyanari, Y., Horiike, T., Li, Y., Yaegashi, N., Sasaki, H., Kono, T. *et al.* (2010) A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic Acids Res.*, **38**, 4929–4945.
16. Hayward, B.E., Kamiya, M., Strain, L., Moran, V., Campbell, R., Hayashizaki, Y. and Bonthron, D.T. (1998) The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc. Natl Acad. Sci. USA*, **95**, 10038–10043.
17. Kamiya, M., Judson, H., Okazaki, Y., Kusakabe, M., Muramatsu, M., Takada, S., Takagi, N., Arima, T., Wake, N., Kamimura, K., Satomura, K., Hermann, R. *et al.* (2000) The cell cycle control gene ZAC/PLAGL1 is imprinted–a strong candidate gene for transient neonatal diabetes. *Hum. Mol. Genet.*, **9**, 453–460.
18. Morales, C., Soler, A., Badenas, C., Rodríguez-Revenga, L., Nadal, A., Martínez, J.M., Mademont-Soler, I., Borrell, A., Milà, M. and Sánchez, A. (2009) Reproductive consequences of genome-wide paternal uniparental disomy mosaicism: description of two cases with different mechanisms of origin and pregnancy outcomes. *Fertil. Steril.*, **92**, 393.e5–e9.
19. Yamazawa, K., Nakabayashi, K., Matsuoka, K., Masubara, K., Hata, K., Horikawa, R. and Ogata, T. (2011) Androgenetic/biparental mosaicism in a girl with Beckwith-Wiedemann syndrome-like and upd(14)pat-like phenotypes. *J. Hum. Genet.*, **56**, 91–93.
20. Romanelli, V., Nevado, J., Fraga, M., Trujillo, A.M., Mori, M.Á., Fernández, L., de Nanclares, G.P., Martínez-Glez, V., Pita, G., Meneses, H. *et al.* (2010) Constitutional mosaic genome-wide uniparental disomy due to diploidisation: an unusual cancer-predisposing mechanism. *J. Med. Genet.*, **48**, 212–216.
21. Yamazawa, K., Nakabayashi, K., Kagami, M., Sato, T., Saitoh, S., Horikawa, R., Hizuka, N. and Ogata, T. (2010) Parthenogenetic chimaerism/mosaicism with a Silver-Russell syndrome-like phenotype. *J. Med. Genet.*, **47**, 782–785.
22. Kanber, D., Berulava, T., Ammerpohl, O., Mitter, D., Richter, J., Siebert, R., Horsthemke, B., Lohman, D. and Buiting, K. (2009) The human retinoblastoma gene is imprinted. *PLoS Genet.*, **12**, e1000790.
23. Takada, S., Paulsen, M., Tevendale, M., Tsai, C.E., Kelsey, G., Cattanach, B.M. and Ferguson-Smith, A.C. (2002) Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum. Mol. Genet.*, **11**, 77–86.
24. Yoon, B.J., Herman, H., Sikora, A., Smith, L.T., Plass, C. and Soloway, P.D. (2002) Regulation of DNA methylation of Rasgrf1. *Nat. Genet.*, **30**, 92–96.
25. Reik, W., Brown, K.W., Slatter, R.E., Sartori, P., Elliott, M. and Maher, E.R. (1994) Allelic methylation of H19 and IGF2 in the Beckwith-Wiedemann syndrome. *Hum. Mol. Genet.*, **3**, 1297–1301.
26. Murrell, A., Ito, Y., Verde, G., Huddleston, J., Woodfine, K., Silengo, M.C., Spreafico, F., Perotti, D., De Crescenzo, A., Sparago, A. *et al.* (2008) Distinct methylation changes at the IGF2-H19 locus in congenital growth disorders and cancer. *PLoS One*, **26**, e1849.
27. Pearsall, R.S., Plass, C., Romano, M.A., Garrick, M.D., Shibata, H., Hayashizaki, Y. and Held, W.A. (1999) A direct repeat sequence at the Rasgrf1 locus and imprinted expression. *Genomics*, **55**, 194–201.
28. Pant, P.V., Tao, H., Beilharz, E.J., Ballinger, D.G., Cox, D.R. and Frazer, K.A. (2006) Analysis of allelic differential expression in human white blood cells. *Genome Res.*, **16**, 331–339.
29. Li, J., Bench, A.J., Vassiliou, G.S., Fourouclas, N., Ferguson-Smith, A.C. and Green, A.R. (2005) Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome deleted in human myeloid malignancies. *Proc. Natl Acad. Sci. USA*, **101**, 7341–7346.

30. Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefevre, A., Coullin, P., Moore, G.E. and Cavaille, L. (2010) The primate-specific mircoRNA gene cluster (C19MC) is imprinted in the placenta. *Hum. Mol. Genet.*, **19**, 3566–3582.
31. Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D. and Dulac, C. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, **329**, 643–648.
32. Greally, J.M. (2002) Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl Acad. Sci. USA*, **99**, 327–332.
33. Ke, X., Thomas, N.S., Robinson, D.O. and Collins, A. (2002) A novel approach for identifying candidate imprinted genes through sequence analysis of imprinted and control genes. *Hum. Genet.*, **111**, 511–520.
34. Luedi, P.P., Dietrich, F.S., Weidman, J.R., Bosko, J.M., Jirtle, R.L. and Hartemink, A.J. (2007) Computational and experimental identification of novel human imprinted genes. *Gen. Res.*, **17**, 1723–1730.
35. Wood, A.J., Roberts, R.G., Monk, D., Moore, G.E., Schulz, R. and Oakey, R.J. (2007) A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet.*, **3**, e20.
36. Zhang, A., Skaar, D.A., Li, Y., Huang, D., Price, T.M., Murphy, S.K. and Jirtle, R.L. (2011) Novel retrotransposed imprinted locus identified at human 6p25. *Nucleic Acids Res.* **[Epub ahead of print]**, doi:10.1093/nar/gkr108.
37. Seoighe, C., Nembaware, V. and Scheffler, K. (2006) Maximum likelihood inference of imprinting and allele-specific expression from EST data. *Bioinformatics*, **22**, 3032–3039.
38. Pollard, K.S., Serre, D., Wang, X., Tao, H., Grundberg, E., Hudson, T.J., Clark, A.G. and Frazer, K. (2008) A genome-wide approach to identifying novel imprinted-genes. *Hum. Genet.*, **122**, 625–634.
39. Henckel, A., Nakabayashi, K., Sanz, L.A., Feil, R., Hata, K. and Arnaud, P. (2009) Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals. *Hum. Mol. Genet.*, **18**, 3375–3383.
40. Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A.M., Esteller, M. and Feil, R. (2002) Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J.*, **21**, 6560–6570.
41. Dindot, S.V., Person, R., Striven, M., Garcia, R. and Beaudet, A.L. (2009) Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res.*, **19**, 1374–1383.
42. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
43. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.
44. Monk, D., Wagschal, A., Arnaud, P., Muller, P.S., Parker-Katiraee, L., Bourc'his, D., Scherer, S.W., Feil, R., Stanier, P. and Moore, G.E. (2008) Comparative analysis of the human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. *Genome Res.*, **18**, 1270–1281.
45. Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forné, T., Murrell, A., Constância, M., Bartolomei, M., Walter, J. *et al.* (2003) Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum. Mol. Genet.*, **12**, 295–305.
46. Coombes, C., Arnaud, P., Gordon, E., Dean, W., Coar, E.A., Williamson, C.M., Feil, R., Peters, J. and Kelsey, G. (2003) Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Mol. Cell. Biol.*, **23**, 5475–5488.
47. Sedlacek, Z., Münstermann, E., Dhorne-Pollet, S., Otto, C., Bock, D., Schütz, G. and Poustka, A. (1999) Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of an ancient functional retroposon differentially expressed in testis. *Genomics*, **61**, 125–132.
48. McCole, R.B. and Oakey, R.J. (2008) Unwitting hosts fall victim to imprinting. *Epigenetics*, **3**, 258–260.
49. Choufani, S., Shapiro, J.S., Susiarjo, M., Butcher, D.T., Grafodatskaya, D., Lou, Y., Ferreira, J.C., Pinto, D., Scherer, S.W., Shaffer, L.G. *et al.* (2011) A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. *Genome Res.*, **21**, 465–476.
50. Mackay, D.J., Callaway, J.L.A., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H.V., Goodship, J.A., Haemers, A.P. *et al.* (2008) Hyomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat. Genet.*, **40**, 949–951.
51. Li, X., Ito, M., Zhou, F., Yougson, N., Zuo, X., Leder, P. and Ferguson-Smith, A.C. (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev. Cell*, **15**, 547–557.
52. Lapunzina, P. and Monk, D. (2011) The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol. Cell*, in press.
53. Kotzot, D. and Utermann, G. (2005) Uniparental disomy (UPD) other than 15: phenotypes and bibliography updated. *Am. J. Med. Genet. A.*, **136**, 287–305.
54. Monk, D., Arnaud, P., Apostolidou, S., Hills, F.A., Kelsey, G., Stanier, P., Feil, R. and Moore, G.E. (2006) Limited evolutionary conservation of imprinting in the human placenta. *Proc. Natl Acad. Sci. USA*, **103**, 6623–6628.
55. Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K. (2009) Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics*, **1**, 177–200.

Human Molecular Genetics, 2011, Vol. 20, No. 14 **2710–2721**
doi:10.1093/hmg/ddr170
Advance Access published on April 19, 2011

Tissue-specific demethylation in CpG-poor promoters during cellular differentiation

Genta Nagae[1], Takayuki Isagawa[1], Nobuaki Shiraki[3], Takanori Fujita[1], Shogo Yamamoto[1], Shuichi Tsutsumi[1], Aya Nonaka[1], Sayaka Yoshiba[1], Keisuke Matsusaka[1,2], Yutaka Midorikawa[1], Shumpei Ishikawa[1,2], Hidenobu Soejima[4], Masashi Fukayama[2], Hirofumi Suemori[5], Norio Nakatsuji[6], Shoen Kume[3] and Hiroyuki Aburatani[1,*]

[1]Genome Science Division, Research Center for Advanced Science and Technology and [2]Department of Pathology, University of Tokyo, Tokyo, Japan, [3]Department of Stem Cell Biology, Institute of Molecular Embryology and Genetics, University of Kumamoto, Kumamoto, Japan, [4]Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan, [5]Laboratory of Embryonic Stem Cell Research, Stem Cell Research Center, Institute for Frontier Medical Sciences and [6]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto, Japan

Received December 13, 2010; Revised and Accepted April 14, 2011

Epigenetic regulation is essential in determining cellular phenotypes during differentiation. Although tissue-specific DNA methylation has been studied, the significance of methylation variance for tissue phenotypes remains unresolved, especially for CpG-poor promoters. Here, we comprehensively studied methylation levels of 27 578 CpG sites among 21 human normal tissues from 12 anatomically different regions using an epigenotyping beadarray system. Remarkable changes in tissue-specific DNA methylation were observed within CpG-poor promoters but not CpG-rich promoters. Of note, tissue-specific hypomethylation is accompanied by an increase in gene expression, which gives rise to specialized cellular functions. The hypomethylated regions were significantly enriched with recognition motifs for transcription factors that regulate cell-type-specific differentiation. To investigate the dynamics of hypomethylation events, we analyzed methylation levels of the entire *APOA1* gene locus during *in vitro* differentiation of embryonic stem cells toward the hepatic lineage. A decrease in methylation was observed after day 13, coinciding with alpha-fetoprotein detection, in the vicinity of its transcription start sites (TSSs), and extends up to ~200 bp region encompassing the TSS at day 21, equivalent to the hepatoblastic stage. This decrease is even more pronounced in the adult liver, where the entire *APOA1* gene locus is hypomethylated. Furthermore, when we compared the methylation status of induced pluripotent stem (iPS) cells with their parental cell, IMR-90, we found that fibroblast-specific hypomethylation is restored to a fully methylated state in iPS cells after reprogramming. These results illuminate tissue-specific methylation dynamics in CpG-poor promoters and provide more comprehensive views on spatiotemporal gene regulation in terminal differentiation.

INTRODUCTION

In a series of differentiation processes during embryogenesis, a wide variety of cells are generated and organized in a spatiotemporal manner. They acquire distinctive patterns of gene expression to execute specialized cellular functions. In mammals, this gene specification is tightly regulated by multiple levels of epigenetic systems such as DNA methylation, histone modification, chromatin remodeling and noncoding RNA guidance (1,2). Mammalian cells coordinately regulate the complex transcriptional networks, which are essential for establishment of cellular programming and maintenance of given cellular phenotypes.

*To whom correspondence should be addressed at: Genome Science Division, Research Center for Advanced Science and Technology, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan. Tel: +81 354525352; Fax: +81 354525355; Email: haburata-tky@umin.ac.jp

DNA methylation has a strong impact on transcriptional repression. Because covalent modification of DNA itself is chemically stable when compared with other epigenetic marks, methylation-mediated repression is thought to be an effective mechanism to maintain long-lasting cell memories. Indeed, it plays pivotal roles in fundamental biological processes, including genome imprinting, retrotransposon silencing, X chromosome inactivation and tissue-specific gene expression (1). The lethality due to selective ablation of DNA methyltransferase with global loss of 5-methylcytosine also provides solid evidence for its significance in mammalian embryogenesis (3,4). Embryonic stem (ES) cells deficient in maintenance methyltransferase, *Dnmt1*, are viable, but die when induced to differentiate (5), demonstrating *Dnmt1* is essential for the dynamic epigenetic changes in cellular differentiation.

For many years, tissue-specific differentially methylated regions (tDMRs) have been of great interest (6–9). In somatic tissues, which include terminally differentiated cells, significant methylation variance between cells have been reported (6,8–11). Although recent technological advances in methylation profiling have broadened our understanding of the human methylome, we are still far from a comprehensive map required for deeper understanding of developmental epigenomics. That is partly because, due to technological limitations, most of earlier studies on human tDMRs have focused on CpG island promoters (12,13). In general, housekeeping genes, which constitutively express across many tissues, have such CpG-rich promoters. However, more than half the genes which have a tissue-specific pattern of expression have CpG-poor promoters (14). Therefore, it is important to analyze CpG-poor promoters in addition to CpG-rich promoters to elucidate regulatory changes of methylation during terminal differentiation.

Recent large-scaled methylation analyses of human normal tissues have revealed that methylation variance can be identified outside CpG islands and at CpG-poor promoters (6). The significance of methylation in the marginal regions of CpG islands (so-called CpG shore methylation) has been also proposed (15). In addition, tissue-specific binding of RNA polymerase II is often observed in CpG-poor promoters (16). These observations point to a significant role for epigenetic dynamics in CpG-poor promoters for terminal differentiation.

There are some difficulties in analyzing methylation levels in human tissue samples with accuracy. Cell populations in human tissues are not homogenous but rather are composed of a heterogeneous cell population which originates from different lineages. Because measurements of methylation are derived from these different methylomes of component cells, large differences in methylation between different cell types can be obscured. It is necessary to evaluate the methylation status quantitatively, rather than just qualitatively, to allow any comparison of methylation profiles between different samples. This requires good assay reproducibility to detect the more subtle methylation differences. In this study, we performed genome-wide promoter methylation analysis of human normal tissues using an epigenotyping beadarray system, which allows methylated CpG quantification in CpG-poor promoters as well as in CpG-rich promoters (17). We utilized inclusive probe sets for tissue-specific hypermethylation and hypomethylation, which occur mainly in CpG-poor promoters. Of note, we found that tissue-specific hypomethylation is well correlated with gene expression profiles that underlie tissue phenotypes. Around these cell-type-specific hypomethylated regions, binding motifs of particular transcription factors are remarkably enriched. These results suggest that a combination of tissue-specific promoter hypomethylation and selective binding of transcription factors is deeply involved in targeting specific genes during terminal differentiation. In addition, we demonstrated spreading of hypomethylation in CpG-poor promoters by *in vitro* cellular differentiation. The restoration of the fibroblast-specific hypomethylation was also observed during cellular reprogramming into induced pluripotent stem (iPS) cells. These results emphasize the importance of methylation dynamics in CpG-poor regions for multilayered epigenetic regulation in mammalian embryogenesis.

RESULTS

Genome-wide methylation analysis of human normal tissue reveals methylation variances in CpG-poor promoters

To develop a better understanding of methylation diversity among human normal tissues, we performed promoter methylation analysis of 21 human normal tissue samples from 12 anatomically different regions (Supplementary Material, Table S1). A HumanMethylation27 BeadChip® (Illumina, Inc) was used to quantify the methylation level of 27 578 CpG sites harboring 14 475 Refseq promoter regions (17).

First, we evaluated the accuracy and sensitivity of the assay using the modified DNA samples as methylation controls (0, 25, 50, 75 and 100% of methylation). The observed values of the methylated CpG ratio for the control samples were well correlated with the expected ratio of methylated CpGs (Supplementary Material, Fig. S1). Thus, methylation changes are quantitatively detectable using this system.

Next, we analyzed inter-individual methylation differences. The comparison plots of autosomal probes using biological duplicates of nine human tissues (brain, oral mucosa, lung, stomach, colon, liver, peripheral blood, kidney and skeletal muscle) showed good correlation between each pair (Pearson correlation coefficient; $r > 0.97$) (Supplementary Material, Fig. S2). For X-linked genes, most promoters on the inactivated allele are methylated in female cells. As expected, 0% methylation in male cells and ~50% methylation in female cells are accurately reported by the system (Supplementary Material, Fig. S2).

In this epigenotyping beadarray, most probes are designed to bind at and around the promoter regions, which are from 1.5 kb upstream to 1 kb downstream of transcription start sites (TSSs) of Refseq genes (Supplementary Material, Fig. S3). On the basis of the classification by the local CpG observed to expected ratio (CpG o/e) and the GC content ratio (GCR) around the probe, probes are divided into three groups: high-CpG density probes (HCG; CpG o/e > 0.75, GCR > 0.55), low-CpG density probes (LCG; CpG o/e $<$ 0.48) and intermediate-CpG density probes (ICG; neither HCG nor LCG) (Supplementary Material, Fig. S4). The promoter methylation status is strongly affected by local CpG density. Most probes in relatively CpG-rich regions (HCG

and ICG) showed hypomethylation in all tissues, while more than half the probes in CpG-poor regions (LCG) are fully methylated (Supplementary Material, Fig. S5). To clarify the relationship between local CpG density and the methylation breadth among human normal tissues, we examined the tissue spatiality of hypermethylation (defined as methylation level more than 0.5) with regard to every autosomal probe ($n = 26\,486$). As shown in Figure 1, the methylation status of the LCG probes is highly variable among different tissues, whereas most CpG-rich probes are ubiquitously unmethylated. Therefore, the majority of intra-individual differences are observed in CpG-poor promoters.

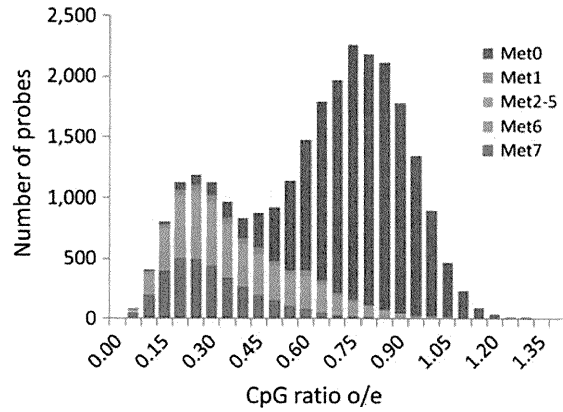

Figure 1. Methylation breadth among human normal tissues. The density histogram represents distribution of the CpG o/e around the probes with respect to the frequency of methylation among seven tissues. First, we generate the histogram of the numbers of probes with regard to CpG ratio o/e. Then, the methylation value of each probe is roughly divided into hypermethylation (0.0–1.0) or hypomethylation (0.0–0.5). Finally, each bar was partitioned by the number of tissues showing hypermethylation. For example, red bars (Met7) indicate all seven tissues are methylated, and blue bars (Met0) indicate no tissues are methylated.

Identification of tissue-specific hypermethylated and hypomethylated regions

To identify the tissue-specific differential gene methylation, we compared the methylation profiles of seven representative tissues. These were the brain and oral mucosa from the ectodermal lineage, the colon and liver from the endodermal lineage, the peripheral blood and skeletal muscle from the mesodermal lineage and the testis. First, we ranked the 26 486 autosomal probes in order of difference of the methylation level between the one tissue and the average of the other tissues. In case that tissue-specific hypomethylation or hypermethylation are sorted by the absolute values of the difference of the methylation level (more than 0.25 or less than −0.25), the number of distinctive gene sets varies widely (Supplementary Material, Fig. S6). There are more tissue-specific hypermethylated genes in the brain, liver, blood and testis than in other tissues. As for the hypomethylation, a large number of genes are selected in the testis and oral mucosa by this criterion. To evaluate the specific differential methylation equally among human tissues, we selected the top 250 probes of tissue-specific hypomethylation and hypermethylation for each tissue (Supplementary Material, Table S2). The methylation panel clearly shows specific hypomethylation as well as hypermethylation among seven tissues (Fig. 2A and B). We validated the methylation levels of the distinctive genes using the MassARRAY system. These methylation levels were consistent with the microarray data (Supplementary Material, Fig. S7). With respect to CpG density, most tissue-specific hypomethylated sites (80–90%) are associated with CpG-poor promoters (Fig. 2A and B). A notable exception was the testis, as testis-specific hypomethylated genes are associated with CpG-rich promoters. In CpG-rich promoter regions (HCP and ICP, $n = 18\,481$), ~900 regions ($5.10 \pm 0.48\%$) were found to be densely hypermethylated (mCpG > 70%) in somatic tissues. In the testis, only 286 regions (1.55%) were methylated (Supplementary Material, Fig. S5). These results are in agreement with earlier systematic screens that found the major fraction of tDMRs corresponding to CpG island methylation are either sperm- or testis-specific (6,9,10).

Variable hypomethylation patterns are associated with tissue-specific gene functions, gene expression patterns and selective binding of transcription factors

To characterize the gene function related to tissue-specific hypomethylation, we examined the enrichment of the specific gene ontology (GO) biological process categories in the top 250 hypomethylated gene sets. As shown in Table 1, the tissue-specific hypomethylated genes are closely related to cell-type-specific functions. For example, oral mucosa-specific hypomethylated genes show over-representation of genes related to ectoderm or epidermis development. In the gene set of liver-specific hypomethylation, we found several protein families synthesized by hepatocytes, such as serpin peptidase inhibitors and complement factors. Thus, these gene sets show over-representation of genes associated with acute inflammatory response. In the blood set, we found genes related to immune response, a key role for white blood cells. Genes related to the reproductive process are the major targets for CpG methylation in somatic cells besides sperm and its progenitor cells in the testis.

In contrast, we could not find any meaningful functional association between gene sets which undergo tissue-specific hypermethylation (Supplementary Material, Table S3). Although previous reports have identified a substantial number of confirmed sets of tissue-specific hypermethylation, it has been difficult to associate these with the tissue phenotype. *De novo* hypermethylation in differentiated cells might be often induced independently of functional specification.

While dense methylation of the CpG island promoter deeply contributes to gene silencing in pathological conditions such as cancer (18,19), the influence of sparse methylation in CpG-poor promoters on gene expression still remains controversial. CpG-poor promoters preferentially display the TATA box and numerous transcription factor-binding motifs around the TSS (20). Combinations of transcription factor binding in regulatory elements are involved in targeting gene expression. Thus, we analyzed expression levels of representative gene sets of tissue-specific hypomethylation and hypermethylation across seven human tissues (Fig. 3). The average expression level of hypomethylated genes is significantly higher than that of hypermethylated genes in a tissue-

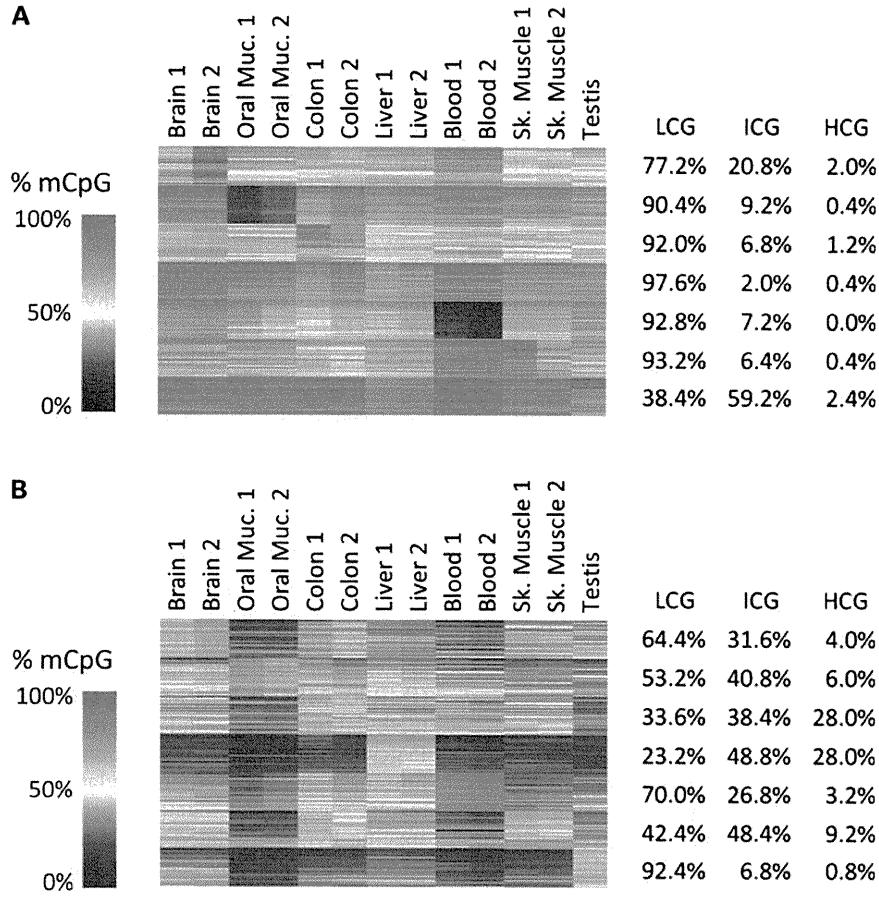

Figure 2. Methylation panel for tissue-specific differential methylation across seven human normal tissues. Left panels indicate the methylation levels of probe sets selected as tissue-specific hypomethylation (**A**) and as tissue-specific hypermethylation (**B**) among seven human tissues. Each row represents a CpG locus (250 for each tissue) and each column represents a tissue sample. The color scale bar at the left side shows the percentage of the methylation level (0–100%). The percentages of the LCG, ICG and HCG in a given probe set are represented at the right side.

specific manner. In contrast, tissue-specific hypermethylated genes are suppressed among all tissues. These results indicate that hypomethylation in the CpG-poor promoters identified here underlie tissue-specific expression in a given cell type.

Local epigenetic modification and recruitment of transcription factors are a fundamental part of the system for appropriate transcriptional regulation (21). We performed enrichment analysis of 746 recognition motifs for transcription factors to examine the relationship between cis-regulatory elements of promoters and tissue-specific hypomethylation. As shown in Figure 4, some matrices are significantly enriched (Z-score $>$ 8.0) in the hypomethylated regions. In oral-mucosa-specific hypomethylated regions, the binding motifs of p53 family genes are highly enriched. p63, the master regulator of keratinocyte differentiation, has similar DNA-binding domains to p53 and half of p63-bound regions in the squamous cell carcinoma cell line have p53 consensus motifs (22). In liver-specific hypomethylated regions, the matrices for the C4 zinc finger domain of the PPAR family (PPARA, PPARG and RXRs) and the NR2F family (HNF4A) are enriched compared with the background sequences. In the blood set, the matrices for the ETS domain of ETS factors (ETS1, ETS2, ELF2, ELK1) and the Runt domain of AML factors (RUNX1) are enriched. Similarly, MyoD-binding motifs are enriched in skeletal muscle-specific hypomethylated regions. In contrast, we could not find significant enrichment of transcription factor-binding motifs in tissue-specific hypermethylated regions (data not shown). Although the molecular mechanism of *de novo* hypermethylation and hypomethylation remains unknown, it is suggested that selective binding of transcription factors are at least significantly associated with regional hypomethylation during terminal differentiation.

Dynamic changes of CpG-poor promoter methylation during *in vitro* differentiation and cellular reprogramming

Although tissue-specific hypomethylation in CpG-poor promoters are closely related to gene specification for the tissue phenotype, when and how these variable methylation statuses are established remain unknown. To elucidate the methylation changes during cellular differentiation, we performed clustering analysis of human somatic tissue and normal cells

Table 1. GO analysis of tissue-specific hypomethylation

Tissue	GO term (biological process)	No. of genes	*P*-value
Brain	Nervous system development	28	4.22E − 05
	Multicellular organismal development	50	4.33E − 04
	Developmental process	53	4.64E − 04
Oral mucosa	Ectoderm development	20	4.89E − 14
	Epidermis development	18	2.19E − 12
	Tissue development	24	2.72E − 07
Colon	Defense response	31	3.68E − 11
	Response to stress	49	3.42E − 09
	Response to stimulus	71	7.31E − 07
Liver	Acute inflammatory response	20	9.00E − 19
	Response to wounding	34	7.22E − 16
	Response to external stimulus	43	3.72E − 15
Blood	Immune system process	55	6.86E − 24
	Immune response	42	2.53E − 19
	Defense response	33	2.62E − 13
Skeletal muscle	Muscle contraction	19	6.02E − 14
	Muscle system process	19	2.70E − 13
	Striated muscle contraction	9	2.64E − 08
Testis	Reproductive process in a multicellular organism	16	1.82E − 04
	Multicellular organism reproduction	16	1.82E − 04
	Gamete generation	14	2.40E − 04

including human ES cells, iPS cells and primary fibroblast cells using tissue-specific hypomethylation sites (Fig. 5). The heatmap shows distinct methylation patterns between the pluripotent cells and somatic tissues composed of the terminally differentiated cells. Seven human ES cell lines and two iPS cell lines show similar methylation patterns. Intriguingly, most genes representing specific hypomethylation in differentiated cells are densely methylated in both ES cells and iPS cells, raising the possibility that the default state of low CpG promoters in the embryonic stage is totally methylated and erasure of methylation may occur during terminal differentiation in a cell-type-specific manner.

We next compared the methylation status of the adult human liver and the fetal liver. Liver-specific hypomethylated genes are heavily methylated in KhES3, a human ES cell line, but are hypomethylated in the adult liver tissue (Fig. 6B and C). In the fetal liver, the methylation level of these genes shows a mild decrease in these regions. Bisulfite sequencing also revealed the partial hypomethylation of *ITIH3* and *APOA1* promoters (Supplementary Material, Fig S8A and B).

To further analyze the demethylation dynamics during hepatic differentiation, we analyzed methylation during *in vitro* differentiation toward hepatic lineages (23). On day 7, the cells began to express an endoderm marker, *SOX17* (Fig. 6A). *AFP* expression was detected on day 13 and *ALB* expression was detected on day 21. The methylation status of liver-specific hypomethylated genes showed a slight decrease during hepatic differentiation (Fig. 6B and C). Indeed, bisulfite sequencing of the *APOA1* promoter region demonstrated that CpG sites in this promoter region are fully hypermethylated in KhES3 and gradually become demethylated during *in vitro* differentiation (Supplementary Material, Fig. S8B). Demethylated regions are observed only in the vicinity of *APOA1* TSSs at day 21 of differentiation, and spread over 1 kb beyond the *APOA1* TSS in adult liver tissues. Sparse non-CpG methylation is observed in KhES3 and lost at day 21 of differentiation and also in adult liver tissues. This demethylation in non-CpG sites in KhES3 is also observed in the promoter region of *CD6* in adult blood and of *STMN4* in the adult brain (Supplementary Material, Fig. S9).

We then analyzed further the methylation status over the entire *APOA1* gene locus to determine the extent of demethylation events (Fig. 6D). Demethylation starts from the vicinity of *APOA1* TSSs at day 13 and extends to 200 bp around the TSS on day 21. Hypomethylated regions in human liver tissues spread over the *APOA1* region, from TSSs to the CpG island of the 3′ end and the further downstream region, suggesting the correlation of extensive demethylation with the stable expression of specific gene sets and cell fate determination.

Epigenetic reprogramming using defined factors enables terminally differentiated cells to gain pluripotency (24). Re-expression of pluripotency genes associated with these promoters, which are methylated in differentiated somatic cells, is important for iPS cell generation (25). The heatmap shows that the four human primary fibroblast cell lines (IMR90, MRC-9, KMS-6 and TIG-103) share specific hypomethylation. After cellular reprogramming into iPS cells, the IMR90 cells show restoration of methylation in these fibroblast-specific hypomethylated sites (Fig. 5). These results suggest that regaining promoter methylation in tissue-specific hypomethylated genes, as well as erasure of methylation in pluripotency genes, is important for this process.

DISCUSSION

In this study, we analyzed inclusive gene sets for tissue-specific hypomethylation and hypermethylation among human normal tissues. Of note, the former gene subsets are remarkably associated with cellular functions characterizing the tissue phenotypes. Although we have examined the limited sites of promoter regions, we reveal here that these hypomethylated genes display tissue-specific patterns of gene expression and specific enrichment of transcription factor recognition motifs in their promoters. This indicates the methylation changes in these regulatory regions might have functional roles in spatiotemporal transcriptional control. Furthermore, the hypomethylation panel showed an unexpected dense methylation pattern in pluripotent stem cells and regional hypomethylation in differentiated cells, suggesting this type of tDMRs might be a consequence of methylation erasure or a dilution process.

To date, the exploration of tDMRs was performed on the premise that stepwise addition of promoter methylation contributes to cell fate determination during early embryogenesis (26,27). It has been widely accepted that the genomic DNA of the embryo, which has pluripotency to differentiate into multiple lineages, is initially unmethylated and subsequent accumulations of hypermethylation in CpG island promoters are important for lineage restriction by reinforcing transcriptional repression of the unnecessary genes (28). Although