Research

Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment

Franck Court,[1,15] Chiharu Tayama,[2,15] Valeria Romanelli,[1,15] Alex Martin-Trujillo,[1,15] Isabel Iglesias-Platas,[3] Kohji Okamura,[4] Naoko Sugahara,[2] Carlos Simón,[5] Harry Moore,[6] Julie V. Harness,[7] Hans Keirstead,[7] Jose Vicente Sanchez-Mut,[8] Eisuke Kaneki,[9] Pablo Lapunzina,[10] Hidenobu Soejima,[11] Norio Wake,[9] Manel Esteller,[8,12,13] Tsutomu Ogata,[14] Kenichiro Hata,[2] Kazuhiko Nakabayashi,[2,16,17] and David Monk[1,16,17]

[1–14][Author affiliations appear at the end of the paper.]

Differential methylation between the two alleles of a gene has been observed in imprinted regions, where the methylation of one allele occurs on a parent-of-origin basis, the inactive X-chromosome in females, and at those loci whose methylation is driven by genetic variants. We have extensively characterized imprinted methylation in a substantial range of normal human tissues, reciprocal genome-wide uniparental disomies, and hydatidiform moles, using a combination of whole-genome bisulfite sequencing and high-density methylation microarrays. This approach allowed us to define methylation profiles at known imprinted domains at base-pair resolution, as well as to identify 21 novel loci harboring parent-of-origin methylation, 15 of which are restricted to the placenta. We observe that the extent of imprinted differentially methylated regions (DMRs) is extremely similar between tissues, with the exception of the placenta. This extra-embryonic tissue often adopts a different methylation profile compared to somatic tissues. Further, we profiled all imprinted DMRs in sperm and embryonic stem cells derived from parthenogenetically activated oocytes, individual blastomeres, and blastocysts, in order to identify primary DMRs and reveal the extent of reprogramming during preimplantation development. Intriguingly, we find that in contrast to ubiquitous imprints, the majority of placenta-specific imprinted DMRs are unmethylated in sperm and all human embryonic stem cells. Therefore, placental-specific imprinting provides evidence for an inheritable epigenetic state that is independent of DNA methylation and the existence of a novel imprinting mechanism at these loci.

[Supplemental material is available for this article.]

Genomic imprinting is a form of epigenetic regulation that results in the expression of either the maternally or paternally inherited allele of a subset of genes (Ramowitz and Bartolomei 2011). This imprinted expression of transcripts is crucial for normal mammalian development. In humans, loss-of-imprinting of specific loci results in a number of diseases exemplified by the reciprocal growth phenotypes of the Beckwith-Wiedemann and Silver-Russell syndromes, and the behavioral disorders Angelman and Prader-Willi syndromes (Kagami et al. 2008; Buiting 2010; Choufani et al. 2010; Eggermann 2010; Kelsey 2010; Mackay and Temple 2010). In addition, aberrant imprinting also contributes to multigenic disorders associated with various complex traits and cancer (Kong et al. 2009; Monk 2010).

Imprinted loci contain differentially methylated regions (DMRs) where cytosine methylation marks one of the parental alleles, providing *cis*-acting regulatory elements that influence the allelic expression of surrounding genes. Some DMRs acquire their allelic methylation during gametogenesis, when the two parental genomes are separated, resulting from the cooperation of the de novo methyltransferase DNMT3A and its cofactor DNMT3L (Bourc'his et al. 2001; Hata et al. 2002). These primary, or germline imprinted DMRs are stably maintained throughout somatic development, surviving the epigenetic reprogramming at the oocyte-to-embryo transition (Smallwood et al. 2011; Smith et al. 2012). To confirm that an imprinted DMR functions as an imprinting control region (ICR), disruption of the imprinted expression upon genetic deletion of that DMR, either through experimental targeting in mouse or that which occurs spontaneously in humans, is required. A subset of DMRs, known as secondary DMRs, acquire methylation during development and are regulated by nearby germline DMRs in a hierarchical fashion (Coombes et al. 2003; Lopes et al. 2003; Kagami et al. 2010).

[15]These authors contributed equally to this work.
[16]These authors jointly directed this work.
[17]Corresponding authors
E-mail nakabaya-k@ncchd.go.jp
E-mail dmonk@idibell.cat
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.164913.113.

24:554–569 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/14; www.genome.org

With the advent of large-scale, base-resolution methylation technologies, it is now possible to discriminate allelic methylation dictated by sequence variants from imprinted methylation. Yet our knowledge of the total number of imprinted DMRs in humans, and their developmental dynamics, remains incomplete, hampered by genetic heterogeneity of human samples.

Here we present high-resolution mapping of human imprinted methylation. We performed whole-genome-wide bisulfite sequencing (WGBS) on leukocyte-, brain-, liver-, and placenta-derived DNA samples to identify partially methylated regions common to all tissues consistent with imprinted DMRs. We subsequently confirmed the partial methylated states in tissues using high-density methylation microarrays. The parental origin of methylation was determined by comparing microarray data for DNA samples from reciprocal genome-wide uniparental disomy (UPD) samples, in which all chromosomes are inherited from one parent (Lapunzina and Monk 2011), and androgenetic hydatidiform moles, which are created by the fertilization of an oocyte lacking a nucleus by a sperm that endoreduplicates. The use of uniparental disomies and hydatidiform moles meant that our analyses were not subjected to genotype influences, enabling us to characterize all known imprinted DMRs at base-pair resolution and to identify 21 imprinted domains, which we show are absent in mice. Lastly, we extended our analyses to determine the methylation profiles of all imprinted DMRs in sperm, stem cells derived from parthenogenetically activated metaphase-2 oocyte blastocytes (phES) (Mai et al. 2007; Harness et al. 2011), and stem cells (hES) generated from both six-cell blastomeres and the inner cell mass of blastocysts, delineating the extent of embryonic reprogramming that occurs at these loci during human development.

Results

Characterization of parent-of-origin methylation profiles in human tissues using high-resolution approaches

We combined whole-genome bisulfite sequencing with Illumina Infinium HumanMethylation 450K BeadChip arrays to generate methylation profiles. To validate this approach, we compared the DNA methylation profiles generated by each method. Methylation scores produced by the two methods are very similar when the same DNA samples were assessed by both techniques (linear regression WGBS vs. Infinium array: leukocytes $R^2 = 0.92$; brain $R^2 = 0.91$; placenta $R^2 = 0.92$) (Supplemental Fig. S1). To determine the similarity between normal biparental leukocytes and those from reciprocal genome-wide UPDs, we compared the methylation values obtained from the Infinium array. This revealed high correlations between samples, indicating that the DNAs were similar, differing only at imprinted loci (linear regression: leukocytes vs. leukocytes $R^2 = 0.95$–0.98; mean control leukocytes vs. mean pUPD $R^2 = 0.98$; mean control leukocytes vs. mUPD $R^2 = 0.98$; mUPD vs. mean pUPD $R^2 = 0.97$; F-statistics $P < 0.001$).

Before we attempted to discover novel imprinted DMRs in the human genome, we wished to determine the effectiveness of the Infinium array to identify known imprinted DMRs. Loci were identified which contained at least three Infinium probes with an average minimal difference of 0.3 β-values (absolute methylation difference >30%) between reciprocal genome-wide UPD leukocyte samples, and with a prerequisite that the β-values for normal leukocytes should be between these extremes. Using these criteria, we identified 818 windows that could be merged into 145 regions harboring 576 probes incorporating 30 known DMRs within 25 imprinted domains (Table 1; Fig. 1A) (Limma linear model $P < 0.05$), and presented an intermediate methylation profile in all somatic tissues (Fig. 1B). The only imprinted DMRs not found using this approach were the IG-DMR located between *MEG3* and *DLK1* on chromosome 14, as this region does not have probes on this array platform and *IGF2*-DMR0 only contains a single probe.

Identification of new DMRs within known imprinted domains

In addition to the known imprinted DMRs, the Infinium array screen of reciprocal UPDs and tissues samples uncovered several previously unidentified DMRs located within existing imprinted domains. We discovered four maternally methylated CpG islands located between the *SNRPN* and *NDN* genes on chromosome 15, a region associated with the Angelman and Prader-Willi syndromes. The methylation profiles at the *SNRPN, NDN,* and *MAGEL2* promoters are well-established (El-Maarri et al. 2001; Sharp et al. 2010). However, little is known about the intervening ~1-Mb gene-poor region, which is likely to have arisen from an ancient duplication event, since these novel DMRs share 97.8% sequence identity with additional CpG-rich regions in the interval. We confirm the maternal methylation at these four regions using bisulfite PCR and sequencing, incorporating heterozygous SNPs in brain and leukocyte DNA (Supplemental Fig. S2A). Further analysis of this region revealed that the promoter region for *MKRN3* and *MIR4508* are also differentially methylated.

Extending our analysis to imprinted domains on other autosomes, we identified an ~600-bp interval of maternal methylation 4 kb 3′ from the *ZNF597* gene (Fig. 1C). Although the promoter of *ZNF597* is a paternally methylated bidirectional silencer presumably responsible for regulating the imprinted expression of both *ZNF597* and *NAA60* (previously known as *NAT15*), this region is unlikely to be the ICR for the domain as its methylation is somatically acquired (Nakabayashi et al. 2011). In addition, WGBS and Infinium array data sets revealed a maternally methylated DMR within intron 2 of *MEG8* within the chromosome 14 imprinted domain (Supplemental Fig. S2B). Lastly, we identify two maternally methylated regions. The first is an ~1-kb CpG island overlapping the promoter of isoform 3 of the *ZNF331* gene, and the second coincides with exon 2 of *DIRAS3* (Supplemental Fig. S2C).

Genome-wide methylation profiling identifies novel imprinted domains

To determine if there are additional imprinted DMRs in the human genome, we screened for regions of intermediate methylation common to lymphocyte, brain, and liver WGBS data sets. Using a sliding window approach that takes into account 25 consecutive CpG sites and following removal of class 1 transposable elements (LINEs, *Alu*/SINEs, and LTR elements) and satellite DNA, we identified 356 nonoverlapping, single-copy regions in pairwise comparisons of tissues, of which 63 loci were common to the all tissues ($0.25 < \text{mean} \pm 1.5\ \text{SD} < 0.75$) (Fig. 2A; Supplemental Table S1).

A screen for three consecutive partially methylated probes in leukocyte, brain, liver, kidney, and muscle Infinium data sets, with a profile consistent with parent-of-origin methylation in the reciprocal UPD leukocyte samples, identified 116 regions (Supplemental Table S1). By combining the 356 regions detected by WGBS and the 116 loci identified by the Infinium array, we identified 64 regions in common, which included all known imprinted DMRs and 17 CpG-rich sequences possessing a methylation profile consistent with imprinting. Using standard bisulfite PCR, we assessed

Table 1. Location of parent-of-origin methylation identified in this study

Known imprinted DMRs (n = 36)								Novel DMRs (n = 25)							
		Extent (WGBS)						Novel DMRs near known imprinted loci (n = 8)							
										Extent (WGBS)					
Gene locus	Chr	Start	Finish	# Infinium probes	GC content	# CpG	Methylation origin	Gene locus	Chr	Start	Finish	# Infinium probes	GC Content	# CpG	Methylation origin
DIRAS3	1	68515433	68517545	17	0.50	88	M	*DIRAS3* Ex2	1	68512505	68513486	8	0.52	39	M
ZDBF2	2	207114583	207136544	8	0.45	439	P	*MEG8*	14	101370741	101371419	1	0.66	43	M
NAP1L5	4	89618184	89619237	15	0.57	57	M	*SNRPN* intragenic CpG32	15	24346736	24347142	1	0.59	30	M
FAM50B	6	3849082	3850359	25	0.65	90	M	*SNRPN* intragenic CpG29	15	24671872	24672679	4	0.59	39	M
PLAGL1	6	144328078	144329888	16	0.58	143	M	*SNRPN* intragenic CpG30	15	24722753	24723071	1	0.66	29	M
IGF2R	6	160426558	160427561	2	0.70	74	M	*SNRPN* intragenic CpG40	15	25017924	25018886	4	0.51	67	M
GRB10	7	50848726	50851312	9	0.60	171	M	*ZNF597*	16	3481801	3482388	2	0.54	29	M
PEG10	7	94285537	94287960	53	0.60	119	M	*ZNF331*	19	54057086	54058425	4	0.66	102	M
MEST	7	130130122	130134388	55	0.54	226	M								
TRAPPC9	8	141108147	141111081	8	0.62	193	M	Novel DMRs (n = 6)							
INPP5F	10	121578046	121578727	4	0.59	52	M	*PPIEL*	1	40024626	40025540	4	0.54	39	M
H19	11	2018812	2024740	48	0.60	250	P	*WDR27*	6	170054504	170055618	2	0.56	58	M
IGF2 DMR2	11	2153991	2155112	9	0.65	63	P	*HTR5A*	7	154862719	154863382	6	0.62	55	M
IGF2 DMR0	11	2168333	2169768	1	0.62	33	P	*CXORF56 pseudogene/ ERLIN2*	8	37604992	37606088	7	0.45	37	M
KvDMR1	11	2719948	2722259	30	0.67	192	M	*WRB*	21	40757510	40758276	4	0.61	43	M
RB1	13	48892341	48895763	12	0.59	195	M	*NHP2L1*	22	42077774	42078873	8	0.54	63	M
IG-DMR	14	101275427	101278058	0	0.52	64	P								
MEG3	14	101290524	101293978	33	0.60	188	P	Known imprinted DMRs (n = 2) & novel DMRs (n = 15)							
MKRN3/ MIR4508	15	23807086	23812495	12	0.44	109	M	Placental-specific DMRs (n = 17)							
MAGEL2	15	23892425	23894029	6	0.55	51	M	*GPR1-AS*	2	207066967	207069445	3	0.49	86	M
NDN	15	23931451	23932759	8	0.65	108	M	*MCCC1*	3	182815725	182817627	13	0.54	94	M
SNRPN	15	25068564	25069481	8	0.42	19	M	*PDE4D*	5	58333774	58336554	7	0.54	145	M
SNRPN	15	25093008	25093829	4	0.49	44	M	*LIN28B*	6	105400631	105402559	8	0.45	62	M
SNRPN	15	25123027	25123905	5	0.47	45	M	*AIM1*	6	106957945	106961974	19	0.54	203	M
SNURF	15	25200004	25201976	7	0.60	113	M	*AGBL3*	7	134671024	134672011	12	0.59	74	M
IGF1R	15	99408496	99409650	7	0.51	55	M	*ZFAT*	8	135707227	135710114	3	0.60	111	M
ZNF597/ NAA60	16	3492828	3494463	11	0.54	76	P	*GLIS3*	9	4297279	4300182	9	0.63	235	M
ZNF331	19	54040510	54042212	11	0.64	125	M	*DCAF10*	9	37800140	37802937	5	0.56	157	M
PEG3	19	57348493	57353271	36	0.59	221	M	*FAM196A/DOCK1*	10	128993405	128995242	10	0.72	198	M
MCTS2P/HM13	20	30134663	30135933	9	0.48	47	M	*ZC3H12C*	11	109962727	109964784	9	0.66	198	M
BLCAP/NNAT	20	36148604	36150528	35	0.55	135	M	*N4BP2L1*	13	33000694	33002448	13	0.66	136	M
L3MBTL	20	42142365	42144040	25	0.65	84	M	*RGMA*	15	93614998	93616859	8	0.61	134	M
GNAS	20	57414039	57418612	23	0.57	257	P	*FAM20A*	17	66596155	66597643	4	0.72	162	M
NESP-AS/ GNAS-AS1	20	57425649	57428033	62	0.61	128	M	*ZNF396*	18	32956510	32957580	9	0.64	86	M
GNAS XL	20	57428905	57431463	6	0.65	200	M	*MIR512-1* cluster	19	54150515	54155608	6	0.53	216	M
GNAS Ex1A	20	57463265	57465201	38	0.67	198	M	*DNMT1*	19	10303506	10306415	10	0.55	129	M

The extent of imprinted methylation is defined by the size of the intermediately methylated region from the lymphocyte (for ubiquitous DMRs) and placenta (for placental-specific DMRs) WGBS data set.

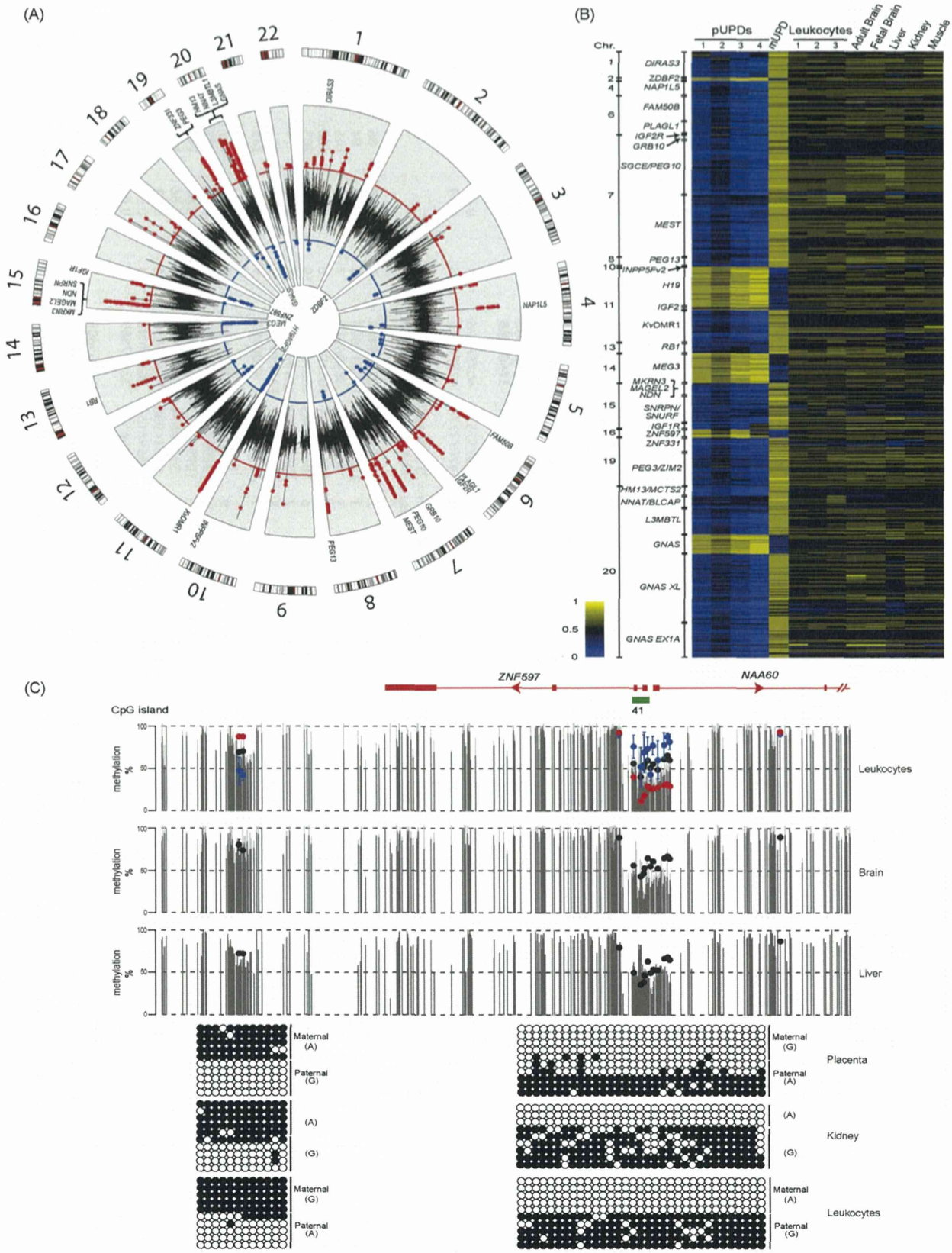

Figure 1. (Legend on next page)

all regions and verified that six regions (*PPIEL, WDR27, HTR5A, WRB, NHP2L1, ERLIN2* loci) are maternally methylated. The DMR we identify within intron 7 of *ERLIN2* appears to be a retroinsertion of the *CXorf56* pseudogene (also known as *LOC728024*) (Fig. 2B,C; Supplemental Fig. S3; Supplemental Table S1).

To confirm that parent-of-origin transcription occurs near these novel imprinted DMRs, we performed allelic RT-PCR in a panel of tissues with primers that discriminate major variant transcripts within each region. This revealed that the DMRs associated with *WDR27, NHP2L1,* and *CXorf56* pseudogenes regulate allelic expression in an isoform-specific fashion (Fig. 2B,C; Supplemental Fig. S3; Supplemental Table S1). We detect monoallelic expression of a short alternatively polyadenylated *ERLIN2* transcript which independently substantiates the observation that the generation of retrogenes, primarily from the X chromosome, is a common mechanism for generating imprinted loci (Wood et al. 2008; Kanber et al. 2009). Unfortunately, due to the lack of informative polymorphisms or expression in available heterozygous tissues, we could not perform allelic expression analysis for *PPIEL, HTR5A,* and *WRB.*

Histone methylation of H3K4 and DNA methylation are enriched on opposing alleles at imprinted DMRs

GC-rich sequences often coincide with enrichment of H3K4me3, which may act to protect them from de novo methylation (Thomson et al. 2010). The H3K4 demethylase KDM1B (previously known as AOF1) is required for appropriate establishment of maternal germline methylation for a subset of imprinted DMRs in mouse, suggesting that the presence of H3K4 methylation is refractory to DNA methylation deposition in the female germline (Ciccone et al. 2009). By comparing publicly available data sets for ChIP-seq for H3K4me3 and methylated DNA immunoprecipitation (meDIP-seq) from blood and brain, we observe co-enrichment of these opposing epigenetic marks at 89.5% of imprinted DMRs, consistent with differential active and repressive chromatin states on homologous chromosomes. For a limited number of informative regions, we were able to confirm H3K4me3 precipitation on the unmethylated allele (Fig. 3C). In most cases, the methylation profile of maternally methylated DMRs is more closely related to the opposing H3K4me3 profile rather than to the CpG density that classically defines CpG islands (>200 bp, GC content >50%, observed/expected ratio >0.6), with the exception of the *GNAS-XL* DMR. This maternally methylated region was thought to be a single regulatory unit; however, our WGBS and Infinium array data clearly show that it is two separate DMRs, partitioned by an ~200-bp interval of hypermethylation, with the centromeric *GNAS-AS1* (previously known as *NESP*-AS) promoter showing coenrichment for H3K4me3 and DNA methylation, while the *GNAS-XL* side lacks this permissive histone modification (Fig. 3A).

Further interrogation of this data set identified two DMRs associated with multiple promoters with a gradient effect across the CpG-rich sequences. The *GNAS/GNAS EX1A* CpG island (CpG island 320 in Fig. 3A) is unmethylated on one side, coinciding with H3K4me3, whereas the other is differentially methylated with abundant H3K4me3 and meDIP reads. This pattern was also observed in the bidirectional *HTR5A/HTR5A-AS1* promoter in brain (Fig. 3B), a tissue where these transcripts are most abundant.

Tissue-specific dynamics of imprinted DMRs

The WGBS analysis in leukocytes, brain, and liver confirmed that the extent of allelic methylation at the imprinted DMRs, as defined by the size of the intermediately methylated interval, is highly similar in these somatic tissues (Figs. 1, 4; Table 1). However, some regions were drastically different in the placenta.

By comparing the placental WGBS profile with Infinium β-values for placentae and hydatidiform moles, we observe that the DMRs associated with the maternally methylated *PEG10* and the paternally methylated *H19* are significantly larger in placenta than in somatic tissues. Using standard bisulfite PCR and sequencing, we confirm that the somatically unmethylated *SGCE* promoter, immediately adjacent to the differentially methylated *PEG10* promoter, is methylated on the maternal allele in placenta, while the maternal allele overlapping the *H19* gene body is demethylated (Fig. 4B).

In addition to identifying extended DMRs in the placenta, we also observe complex tissue-specific methylation between somatic tissues and placenta. For example, the *NNAT* and *GNAS-AS1* DMRs, which are maternally methylated in somatic tissue, exhibit hypermethylation in both placenta and hydatidiform mole. Subsequent bisulfite PCR confirmed that these regions are fully methylated in the placenta (Supplemental Fig. S4). Methylation profiling at the *MIR512-1* cluster (also known as *C19MC*)-*ZNF331* locus on chromosome 19 has previously disclosed that the promoter of the pri-miRNA for this miRNA cluster is maternally methylated in placenta, but fully methylated in somatic tissues (Noguer-Dance et al. 2010). We confirm that the *MIR512-1* DMR is unmethylated in hydatidiform moles compared to the partially methylated profile in placenta, with placental WGBS revealing that the DMR is ~5 kb in size, incorporating the promoter CpG island (CpG island 86 in Fig. 4C). However, we notice that the CpG island (CpG island 83 in Fig. 3C) associated with *ZNF331* isoform-3 is hypermethylated on both parental alleles in placenta but is a maternally methylated DMR in somatic tissues. These methylation states dictate complex allelic expression at this locus, with restricted placental-specific paternal expression of the *MIR512-1* pri-miRNA, which does not extend to the *MIR371/2* cluster, and reciprocal imprinting of *ZNF331* (Fig. 4C; Supplemental Table S2).

Novel placental-specific DMRs associated with paternally expressed transcripts

Based on the complex methylation profiles described above, we next investigated if more unknown imprinted DMRs exist solely in

Figure 1. Identification of known imprinted DMRs on the Infinium array platform. (*A*) Circular karyotype showing the difference of methylation for three consecutive probes for reciprocal UPD leukocyte samples. Red dots indicate a minimal difference of 0.3 in Infinium probe β-values (>30% absolute methylation value) for regions with maternal methylation, and blue dots indicate the same for paternal methylation. Known DMRs are indicated. (*B*) Heat map of the Infinium probes located within known imprinted DMRs in reciprocal genome-wide UPD samples and various somatic tissues. (C) WGBS and Infinium array methylation profiles of the *ZNF597* locus with bisulfite PCR confirmation of the novel maternally methylated DMR and its position in relation to the somatic paternally methylated promoter region. Vertical gray lines in the WGBS tracks represent the mean methylation value for individual CpG dinucleotides calculated from multiple data sets, with the light gray lines representing the mean + standard deviation. Infinium methylation values for normal tissues are represented by black dots, with values for the genome-wide UPDs (average pUPD in blue and mUPD in red) superimposed on the leukocyte methylation track. The error bars associated with the Infinium array probes represent the standard deviation of multiple biological samples. The PCR confirmation in placenta, kidney, and leukocyte-derived DNA was performed on heterozygous samples. Each circle represents a single CpG dinucleotide on a DNA strand. (●) Methylated cytosine, (○) unmethylated cytosine. Each row corresponds to an individual cloned sequence.

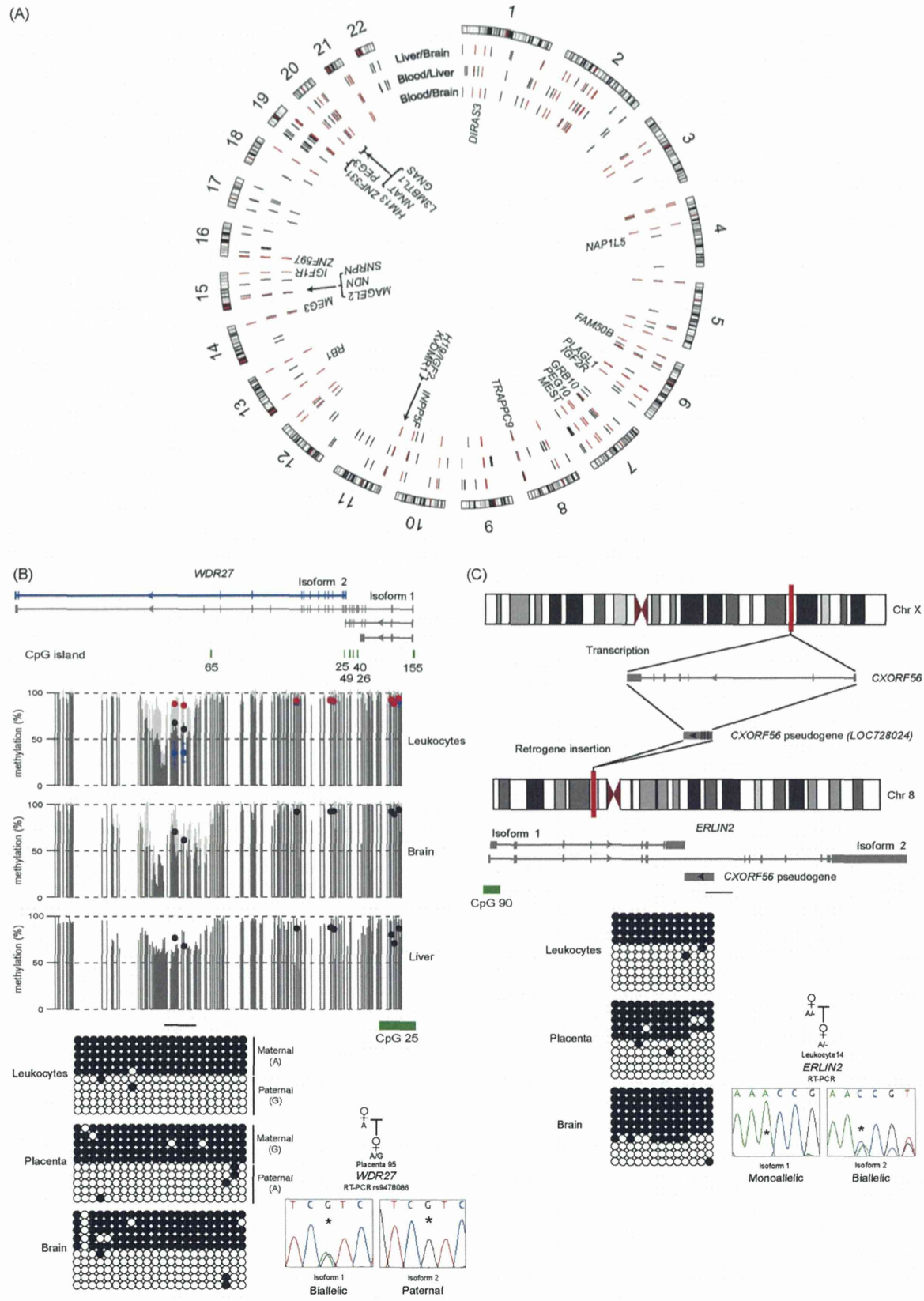

Figure 2. (Legend on next page)

placental tissues, as highlighted by the *MIR512-1* and *GPR1-AS* DMRs (Noguer-Dance et al. 2010; Kobayashi et al. 2013).

We performed a screen for partially methylated regions present solely in our placenta WGBS data set using our sliding window approach (0.25 < mean of 25 CpG ± 2 SD < 0.75). This identified 722 windows, of which 520 mapped to CpG islands. These results confirm that placental-derived DNA is significantly less methylated when compared to other tissues (Schroeder et al. 2013) and that this genome-wide lower methylation is not restricted to repetitive elements as previously described (Ehrlich et al. 1982; Fuke et al. 2004), but occurs across a large portion of the genome.

Of these partially methylated placenta domains identified by WGBS, 44 regions were ~50% methylated in placenta, with extreme methylation in hydatidiform moles using the Infinium array (average β-value for three consecutive probes >0.8 indicative of paternal methylation or <0.2 indicative of maternal methylation), and showed no evidence of allelic methylation in somatic tissues. Using standard bisulfite PCR, we assessed the allelic methylation profile of all regions in placental DNA samples. This revealed that the promoters of *N4BP2L1, DCAF10, PDE4D, FAM196A, RGMA, AGBL3, MCCC1, ZC3H12C, DNMT1, AIM1, ZNF396, FAM20A, GLIS3*, and *LIN28B* are methylated on the maternal allele (Fig. 4D; Supplemental Fig. S5; Supplemental Table S2). In addition, we identified a 2.8-kb region of intermediate methylation overlapping an alternative promoter of the paternally expressed *ZFAT* gene in the placental WGBS data set (Supplemental Fig. S5). Using allelic-specific bisulfite PCR, we confirm that the methylation is confined to the maternal chromosome at this locus. To determine if these regions of maternal methylation influence transcription, allelic RT-PCR experiments were carried out. Paternal expression of eight of these genes was verified, with biallelic expression in somatic tissues (Fig. 4D; Supplemental Fig. S5; Supplemental Table S2) consistent with recent allelic expression screens in term placenta (Yuen et al. 2011; Barbaux et al. 2012).

Mammalian conservation of novel imprinted domains

To determine if the previously unrecognized imprinted domains are conserved throughout evolution, we assessed their allelic methylation and expression in mice, using a reciprocal cross between mouse strains. Bisulfite PCR targeting of orthologous regions failed to identify evidence of differential methylation in embryonic day E9.5–14.5 embryos or extra-embryonic tissues. Subsequent allelic RT-PCRs revealed that all murine transcripts orthologous to the novel ubiquitous and placental-specific imprinted transcripts are equally expressed from both parental alleles when detected (Supplemental Figs. S6, S7). This suggests that these new imprinted domains arose less than ~80 million years ago after the divergence of mice and humans or that selection pressures over this period have resulted in a loss of imprinted regulation of these genes in mice. It has been previously reported that imprinting in the placenta differs between human and mouse, mainly due to the lack of imprinting of genes which require repressive histone modifications for allelic silencing in humans (Monk et al. 2006). Contrary to previous reports, our results show that humans have evolved more loci subject to this form of transcriptional regulation in placenta, due to the evolutionary acquisition of loci with parent-of-origin methylation. This is endorsed by the low discovery rate of novel imprinted transcripts in RNA-seq screens of mouse placenta (Okae et al. 2011).

Differential methylation at ubiquitously imprinted loci and placental-specific domains may differ in their gametic origin

An essential step toward understanding the establishment of the germline imprint signal is to determine if the parent-of-origin methylation observed in somatic tissues is derived from the germline. Determining the methylation profiles in human gametes and during the early preimplantation stages of embryonic development is technically and ethically challenging. To circumvent these difficulties, we have used a combination of mature gametes and in vitro models to represent human gametes of both sexes and preimplantation embryos. For analysis during gametogenesis in males, we used mature sperm. We compared publicly available WGBS data sets from sperm and human embryonic stem (hES) cells that represent the inner cell mass of the blastocysts (Lister et al. 2009; Molaro et al. 2011) with our own Infinium array profiles for sperm, parthenote-derived hES cell lines (phES), and hES cell lines generated from both six-cell blastomeres (Val10B) and the inner cell mass of blastocysts (SHEF cell lines). Despite the phES cell lines having undergone reprogramming during blastocyst development, they have previously been shown to retain maternal hypermethylation at the limited imprinted loci assessed, suggesting that they are ideal surrogates for assessing the methylation profiles of imprinted DMRs in mature oocytes (Mai et al. 2007; Harness et al. 2011).

A comparison of Infinium β-values between sperm and phES cells for the human sequences orthologous to the mouse germline DMRs (Kobayashi et al. 2012) revealed that 19/22 are conserved. The novel ubiquitous DMRs we identify are also hypermethylated in phES cells and unmethylated in sperm, suggesting that the majority of imprinted DMRs, with the exception of *IGF1R*, are primarily marked in the gametes (Fig. 5A; Supplemental Fig. S8). In addition, we confirm that the IG-DMR within the chromosome 14 domain is >80%–90% methylated in the sperm WGBS data set, in line with previous reports (Geuns et al. 2007). We were particularly intrigued to observe that all placental-specific DMRs, with the exception of *ZFAT, GPR1-AS*, and *MIR512-1*, do not inherit methylation from the gametes and are devoid of methylation in hES cells (Fig. 5A). These data provide preliminary evidence to suggest that, following gametogenesis, parental alleles at some loci retain a nonequivalency that is not associated with DNA methylation.

Figure 2. Identification and characterization of allelic methylation and expression of novel imprinted loci. Circular karyotype showing the position of common regions of intermediate methylation in the leukocyte, brain, and liver WGBS data sets, as identified using a 25 CpG sliding window approach (0.25 < mean ± 1.5 SD < 0.75). Red ticks represent sites of intermediate methylation common to all tissues, whereas black ticks identify those present in only one or two pairwise comparisons. The position of known imprinted DMRs are shown. (*B*) Identification of a novel maternally methylated DMR within the *WDR27* locus by WGBS and Infinium array analysis. Vertical gray lines in the WGBS tracks represent the mean methylation value for individual CpG dinucleotides calculated from multiple data sets, with the light gray lines representing the mean + standard deviation. Infinium methylation values for normal tissues are represented by black dots, with values for the genome-wide UPDs (average pUPD in blue and mUPD in red) superimposed on the leukocyte methylation track. The DMR was confirmed using standard bisulfite PCR on heterozygous DNA samples and orchestrates paternal expression of *WDR27* isoform 2. The asterisk (*) in the sequence traces shows the position of the polymorphic base. (*C*) Imprinting of *ERLIN2* isoform 1 in leukocytes as a consequence of the retrotransposition of the X chromosome-derived *CXorf56* pseudogene into the locus.

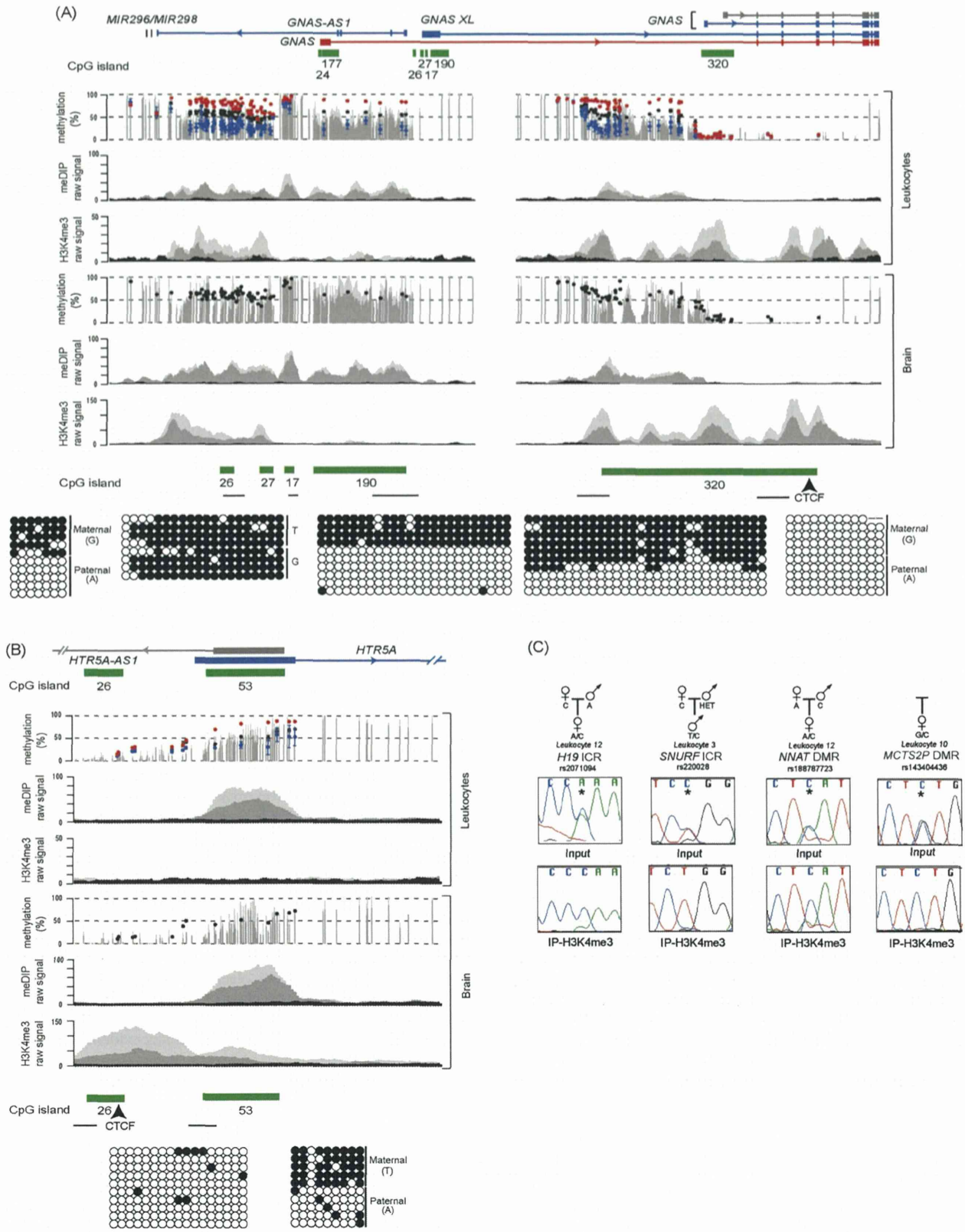

Figure 3. H3K4me3 chromatin profile and DNA methylation at imprinted loci. (*A*) Map of the human *GNAS* locus on chromosome 20 with the H3K4me3 and meDIP signatures in brain and leukocytes at the DMRs identified in the WGBS and Infinium array analysis. Infinium methylation values for normal leukocytes (black dots), with values for the genome-wide pUPD (blue) and mUPD (red) superimposed on the leukocyte WGBS track. Similarly, Infinium methylation values for two normal brain samples are shown as black and gray dots. The light and dark gray peaks in the meDIP and ChIP-seq panels represent two independent biological replicates compared to input (black peaks). The bars *under* the CpG islands, as identified in the UCSC Genome Browser, show the location of the bisulfite PCR amplicons. (*B*) The gradient DMR identified at the *HTR5A* promoter. The samples used for the WGBS, Infinium array, and ChIP are the same as in *A*. The independent methylation pattern on either side of the bidirectional promoter interval was confirmed using standard bisulfite PCR and sequencing. (*C*) Allelic ChIP for H3K4me3 reveals predominant enrichment of this histone modification on the unmethylated allele of the *H19* ICR, *SNURF* ICR, *NNAT*, and *MCTS2P* DMRs. The asterisk (*) in the sequence traces shows the position of the polymorphic base.

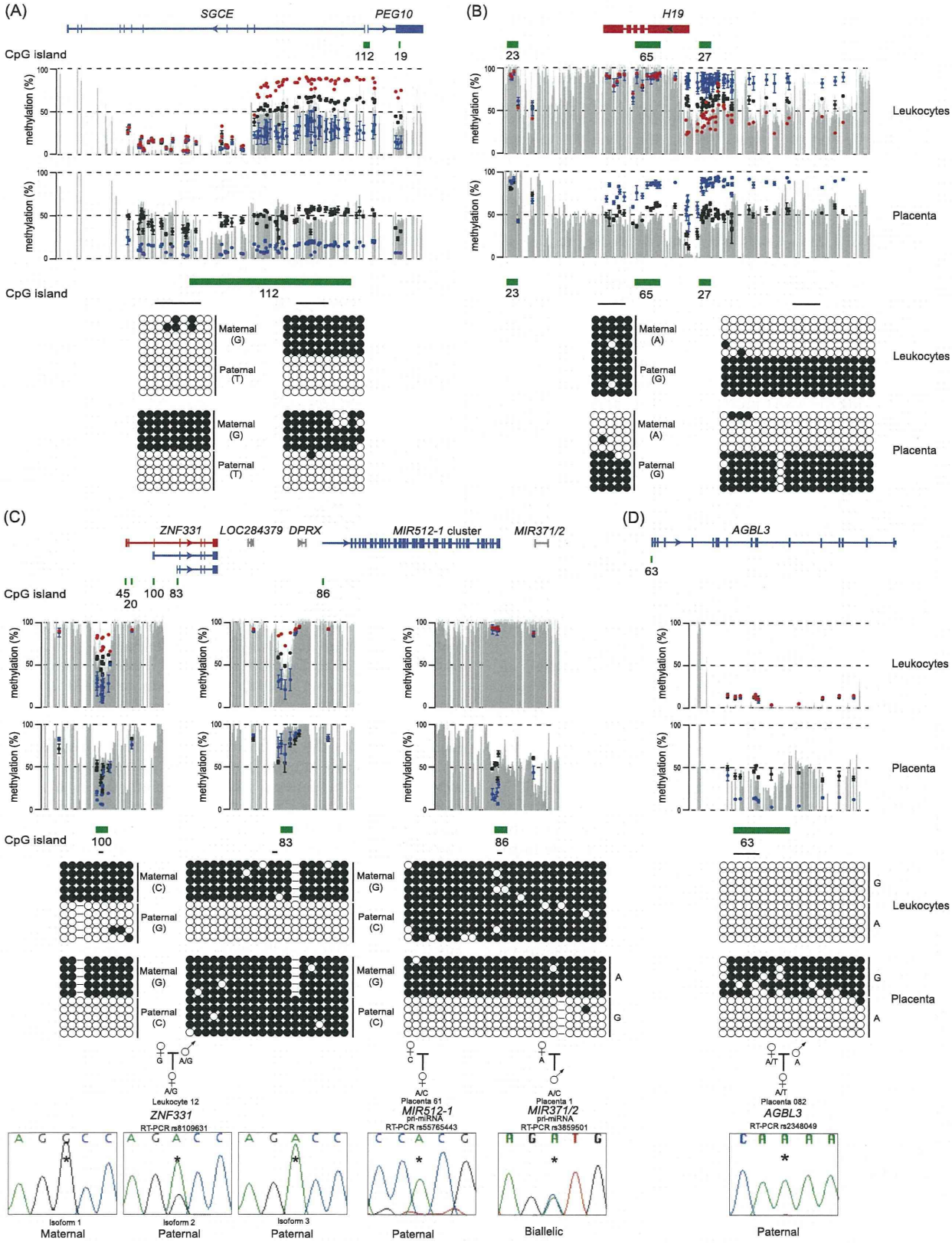

Figure 4. (Legend on next page)

Therefore, in the female germline, as represented by the phES cells, a subset of imprinted loci retain their identity in the absence of methylation, suggesting that additional epigenetic mechanisms mark these regions for maternal methylation during trophoblast differentiation (Fig. 5A).

For the majority of DMRs for which allelic methylation was observed in the somatic tissues (80%), the genomic interval showing methylation differences between sperm and phES cells is larger than the allelic DMRs in hES cells and somatic tissues (Fig. 5B,C). In the case of maternally methylated DMRs, we observe that these regions are flanked by fully methylated intervals in both gametes, and that these DMRs are observed as regions devoid of methylation in the sperm genome. Interrogation of ChIP-seq data sets for nucleosomes containing the histone modifications revealed that the majority of unmethylated DMR regions in sperm are enriched for H3K4me3 containing nucleosome fractions. Our analysis indicates that the size of the unmethylated region in sperm is therefore associated with nucleosome occupancy, rather than protamines. Notably, the maternally methylated germline DMR overlapping the *NNAT* promoter is ~4 kb, as defined by full methylation in phES cell and the H3K4me3 enriched DNA unmethylated region in sperm. This region contracts to an ~1.5-kb region of maternal methylation after preimplantation reprogramming as represented by blastocyst-derived hES cells and somatic tissue profiles (Fig. 5C; Supplemental Fig. S9A showing the contraction at the *NAP1L5* locus). Such resizings are also observed in mouse (Tomizawa et al. 2011), suggesting that imprinted DMRs are not totally protected from genome-wide demethylation during the oocyte to embryo transition. We speculate that the larger regions of differential methylation dictated by the gametes, in combination with protective factors, ensure that they survive reprogramming.

In addition, we also observe other subtle differences in germline-derived methylation profiles. For example, the two sides of the *GNAS-XL* DMR that we show to have independent H3K4 methylation profiles from each other behave differently in the germline, with the *GNAS-AS1* side being a somatic DMR only, but the *GNAS-XL* side being methylated in phES cells and hypomethylated in sperm (Supplemental Fig. S4). Lastly, we identified a dynamic relocalization of methylation at the *FAM50B* DMR during preimplantation development. The 1.2-kb promoter of this imprinted retrogene is methylated on the maternal allele in somatic tissues but is completely unmethylated in phES cells and hES cells derived from six-cell embryos, and has been shown to be unmethylated in sperm (Nakabayashi et al. 2011). However, we do find that allelic methylation is conferred during preimplantation development, at a point between the six-cell stage and blastocyst development. In fact, the ~1-kb regions flanking the promoter (labeled 1 and 3 in Supplemental Fig. S9B) show strongly opposing methylation profiles, with the sperm being unmethylated and phES cells methylated, which then become fully methylated on both alleles immediately after fertilization, leaving allelic methylation over the promoter itself.

Discussion

Differentially methylated regions between the parental alleles are essential for genomic imprinting and development. In this study, we have performed a comprehensive survey of methylation in various human tissues, uncovering all known imprinted DMRs as well as 21 novel loci, which we demonstrate wherever possible regulate imprinted transcription. Our present work demonstrates that the human genome contains a significantly larger number of regions of parent-of-origin methylation than previously thought. The identification of imprinted domains has traditionally been performed in mouse by utilizing gynogenetic and androgenetic embryos, mice harboring regions of uniparental disomies, or highly polymorphic inbred strains (Cooper and Constância 2010). These embryos have been subjected to expression-based screens, including RNA-seq (Gregg et al. 2010; Okae et al. 2011), and genome-wide methylation techniques (Hayashizaki et al. 1994; Kelsey et al. 1999; Hiura et al. 2010). By relying on the confirmation of the evolutionarily conserved expression of the human orthologs, imprinted genes specific to higher primates and humans would have been missed. We have utilized high-throughput bisulfite analysis from in vitro models of gametes and early embryos, and somatic and placental DNA, to characterize the developmental dynamics of imprinted methylation coupled with allelic expression analysis of nearby transcripts. This analysis reveals that 30 regions of parentally inherited differential methylation are observed in humans but not mice. Conversely, we also show that the DMRs associated with *Cdkn1c*, *Rasgrf1*, the *Igf2r* promoter, *Impact*, *Slc38a4*, and *Zrsr1* (previously known as *U2af1-rs1*) imprinted transcripts in mouse do not exhibit allelic methylation in humans (Xie et al. 2012).

Recently, a novel mechanism has been described in which differences in germline methylation can give rise to tissue-specific DMRs in mouse (Proudhon et al. 2012). The *Cdh15* DMR inherits methylation from the oocyte and maintains this parental allelic methylation during in utero development and in adults, with the exception that the paternal allele gains methylation in various brain regions. Therefore, the intragenic *Cdh15* DMR is conserved during adulthood, but in a tissue-specific manner. In humans, the *CDH15* locus does not exhibit allelic DNA methylation in any tissue (data not shown), suggesting that this tissue-specific methylation profile might be limited to mice. We cannot rule out the existence of temporally regulated tissue-specific imprinted DMR in humans, since our samples were derived from adults, and therefore any imprinted DMRs specific to the fetal period would be missed.

Our study reveals the power of combining WGBS and Infinium HumanMethylation450 BeadChip arrays to identify novel imprinted DMRs. We have previously combined reciprocal genome-wide UPD samples and the Infinium HumanMethylation27

Figure 4. The methylation profiles of imprinted loci in placenta compared to somatic tissues. The placenta- and leukocyte-derived WGBS and Infinium array profiles at the (*A*) *PEG10* and (*B*) *H19* loci. Infinium methylation values for normal leukocytes (black dots), with values for the genome-wide pUPD (blue) and mUPD (red) superimposed on the leukocyte WGBS track. Similarly, Infinium methylation values for normal placenta (black dots) and hydatidiform mole (blue dots) are overlaid on the placental WGBS track. The error bars associated with the Infinium array probes represent the standard deviation of multiple biological samples. Bisulfite PCR analysis was used to confirm the tissue-specific methylation profiles. (*C*) Complex tissue-specific allelic methylation and expression patterns at the *ZNF331-MIR512* cluster locus on chromosome 19. The *ZNF331* sequence traces represent the RT-PCR products from leukocytes, whereas both the *MIR512-1* cluster and *MIR371/2* are from placenta. (*D*) A placental-specific imprinted DMR identified using the placenta-derived WGBS and Infinium array data sets. The methylation profiles were confirmed using standard bisulfite PCR on heterozygous DNA samples with allelic RT-PCR performed on placental biopsies. The results confirm that the region of maternal methylation overlapping the *AGBL3* promoter dictates paternal expression of this gene in placenta.