ORIGINAL ARTICLE

Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes

Louise E Docherty,[1,2] Faisal I Rezwan,[1] Rebecca L Poole,[1,2] Hannah Jagoe,[1] Hannah Lake,[1] Gabrielle A Lockett,[1] Hasan Arshad,[1,3] David I Wilson,[1] John W Holloway,[1] I Karen Temple,[1,4] Deborah J G Mackay[1,2]

ABSTRACT

Background Genomic imprinting is allelic restriction of gene expression potential depending on parent of origin, maintained by epigenetic mechanisms including parent of origin-specific DNA methylation. Among approximately 70 known imprinted genes are some causing disorders affecting growth, metabolism and cancer predisposition. Some imprinting disorder patients have hypomethylation of several imprinted loci (HIL) throughout the genome and may have atypically severe clinical features. Here we used array analysis in HIL patients to define patterns of aberrant methylation throughout the genome.

Design We developed a novel informatic pipeline capable of small sample number analysis, and profiled 10 HIL patients with two clinical presentations (Beckwith–Wiedemann syndrome and neonatal diabetes) using the Illumina Infinium Human Methylation450 BeadChip array to identify candidate imprinted regions. We used robust statistical criteria to quantify DNA methylation.

Results We detected hypomethylation at known imprinted loci, and 25 further candidate imprinted regions (nine shared between patient groups) including one in the Down syndrome critical region (WRB) and another previously associated with bipolar disorder (PPIEL). Targeted analysis of three candidate regions (NHP2L1, WRB and PPIEL) showed allelic expression, methylation patterns consistent with allelic maternal methylation and frequent hypomethylation among an additional cohort of HIL patients, including six with Silver–Russell syndrome presentations and one with pseudohypoparathyroidism 1B.

Conclusions This study identified novel candidate imprinted genes, revealed remarkable epigenetic convergence among clinically divergent patients, and highlights the potential of epigenomic profiling to expand our understanding of the normal methylome and its disruption in human disease.

INTRODUCTION

Genomic imprinting is the epigenetic regulation of gene expression by parent of origin. DNA methylation at imprinting control regions (ICRs) is the most robust and widely studied epigenetic modification regulating imprinting. Genomic imprinting requires resetting of DNA methylation in the germline and its subsequent resistance to erasure during the transition from germ cell to early embryonic development.[1][2] While methylation at ICRs is ubiquitous and permanent, the effects on DNA methylation and expression of surrounding genes are dependent on other factors such as tissue and developmental stage.[3]

Many imprinted loci were identified through the developmental disorders caused by their disruption, and particularly the discovery of uniparental disomy and other genetic errors in rare human disorders of imprinting.[4][5] But the total number of imprinted genes is not known. Recent efforts to identify imprinted genes by murine transcriptome analysis yielded high numbers of transcripts with allelic bias.[6] However, this observation has been disputed and may be attributable to various technical sources of skewed allelic representation in RNA-seq data[7] and, more recently, genome-wide bisulfite sequencing has allowed direct assessment of allele-specific methylation;[8] taken together, these observations suggest that our current catalogue of imprinted genes is approaching completion, with few novel germline imprints remaining to be discovered (http://igc.otago.ac.nz).[9]

Many known imprinted genes are regulators of growth and development, and their expression at critical developmental times is functionally hemizygous. Therefore, alteration of effective copy number can cause developmental disorders.[10] To date, eight imprinting disorders (IDs) have been identified: Beckwith–Wiedemann syndrome (BWS; MIM #130659), Silver–Russell syndrome (SRS; MIM #180860), transient neonatal diabetes (TND) mellitus (MIM #601410), Prader–Willi syndrome (MIM #176270), Angelman syndrome (MIM #105830), matUPD14-like (Temple syndrome) and patUPD14-like syndromes, and pseudohypoparathyroidism 1B (PHP-1B; MIM #103580). Aetiological mechanisms of IDs include UPD, copy number variation, mutation of the expressed copy, or epimutation secondary to or independent of a predisposing genetic mutation. A subset of patients with IDs have epimutations affecting multiple imprinted loci across the genome (multi-locus methylation disorders or hypomethylation of imprinted loci (HIL)[11]). The reported rate of HIL in BWS is 38% (with ICR2

hypomethylation), 57% in TND (with *PLAGL1* hypomethylation) and 10% in SRS (with ICR1 hypomethylation).[12–14] There is no standard quantification for hypomethylation at the affected loci, though tissue mosaicism is thought to account for the variation observed between patients. In some of these disorders, a shared pattern of methylation derangement can be detected, and underlying genetic mutations have been identified;[15–18] in other cases, the cause(s) remain unknown.

In order to identify novel imprinted regions, several groups have used genome-wide methylation analyses of patients with UPD and HIL, commonly using the Infinium Human Methylation27 BeadChip array.[19–21] The potential limitations of this approach include the limited coverage of this array, and the lack of suitable bioinformatic pipelines to study large methylation changes in small study cohorts, as currently available pipelines are designed to assess modest DNA methylation changes in large study cohorts.[22–24] To address these limitations, we used the Infinium Human Methylation450 BeadChip array, and developed a new analysis pipeline capable of robust analysis of small study groups with large methylation changes.

Here, we analysed the methylomes of 10 HIL patients with two clinical presentations (five BWS and five neonatal diabetes), compared with normal controls, and identified hypomethylated regions, including three hitherto undescribed candidate imprinted regions.

MATERIALS AND METHODS
Study population (ethics)
Peripheral blood leucocyte DNA of patients with IDs was assessed by methylation-specific PCR (msPCR) at 11 maternally methylated loci, as described (see online supplementary table S1; the majority of these patients have been previously reported in Poole *et al*[12]). Those patients with hypomethylation at loci additional to the primary locus for their presenting disorder were classified as HIL, and subgrouped using the epigenetic profiles of these 11 maternal imprinted loci. It was apparent that five patients with TND and five with BWS showed an overlapping pattern of hypomethylation: TND-HIL samples showed hypomethylation at *PLAGL1*, *DIRAS*, *IGF2R* and *IGF1R* differentially methylated regions (DMRs), with some additional overlap of hypomethylation at *MEST*, KCNQ1OT1 and *GRB10*, and BWS-HIL patients shared hypomethylation of *KCNQ1OT1*, *PLAGL1*, *IGF2R* and *MEST*, with *NESPAS* and *GNAS* hypomethylation observed in 2/5 patients. These patients were selected for further analysis to determine whether they had additional shared hypomethylation patterns.

All TND-HIL patients were negative for *ZFP57* mutations and BWS-HIL patients negative for *NLRP2* mutations. The ethical approval for the use of these samples was obtained through the study 'Imprinting Disorders Finding Out Why?', approved by Southampton and South West Hampshire Research Ethics committee 07/H0502/85 and 'Mapping clinical and molecular studies of 6q24 transient neonatal diabetes' approved by Wiltshire Research Ethics committee 08/H0104/15.

Control population
Control group 1 (N=221) and control group 2 (N=245) anonymous batch-matched healthy samples from an unrelated study were used to generate control methylation profiles for the analysis of TND-HIL and BWS-HIL cases, respectively. Control group 1 samples were mixed gender and source material, with 198 peripheral blood leucocytes DNA samples derived from cohort members and their partners and 23 cord blood leucocytes DNA samples from their offspring whereas control group

2 contained 221 peripheral blood leucocyte DNA samples from female subjects at 18 years of age from an unselected population birth cohort. Ethical approval was obtained from the Isle of Wight Local Research Ethics Committee (now named the National Research Ethics Service, NRES Committee South Central—Southampton B) for the 18 years follow-up (06/Q1701/34) and NRES Committee South Central—Hampshire B (09/H0504/129) for the third generation study.

Validation samples
Methylation array findings were validated by targeted testing of DNA and RNA samples. DNA was derived from two hydatidiform mole cell lines, peripheral blood leucocytes of 92 anonymised controls, four anonymised normal trios and 34 anonymised individuals diagnosed with Down syndrome, and patients with IDs: five TND-HIL, six BWS-HIL, seven SRS–HIL, one PHP-HIL, five *ZFP57* mutation cases presenting with TND and nine patients with hypomethylation at only one locus (two TND with *PLAGL1* hypomethylation, two BWS patients with KCNQ1OT1 hypomethylation, four SRS patients with ICR1 hypomethylation and one with UPD7mat). These samples were obtained under the same ethical approval as the study group and previously reported.[12] Nucleic acids (DNA and RNA) from human embryonic and fetal tissues were obtained with informed consent and with permission from the Southampton and South West Hampshire joint Research Ethics Committee, staged according to the Carnegie classification or foot length.

Array-based methylation analysis
1250 ng of Qubit 2.0 Fluorometer quantified DNA was bisulfite-treated using the EZ 96-DNA methylation kit (Zymo Research, California, USA), following the manufacturer's standard protocol. Genome-wide DNA methylation was assessed by The Oxford Genomics Centre using the Illumina Infinium HumanMethylation450 BeadChip (Illumina, Inc., California, USA). Arrays were processed using the manufacturer's standard protocol with multiple identical control samples assigned to each bisulfite conversion batch to assess assay variability and samples randomly distributed on microarrays to control against batch effects. The BeadChips were scanned using a BeadStation, and the methylation level (β value) calculated for each queried CpG locus using the Methylation Module of BeadStudio software.

Data preprocessing and quality control
A pipeline was developed using the Illumina methylation analysis (IMA) package within the R statistical analysis environment (http://www.r-project.org).[22] Data from five TND-HIL and five BWS-HIL samples were grouped and run in this pipeline independently. Sites were removed that contain any missing values. All samples met minimal inclusion criteria for analysis, as each sample had >75% sites with a detection p value $<1\times10^{-5}$. In all, 216 sites were removed from TND-HIL study and 106 from BWS-HIL study, as these had detected p value >0.05 in at least 75% of the sample analysed. Among these removed sites, 68 are common between the two study groups. Initial QC-plots (see online supplementary figure S1) for both of the studies showed that male and female samples clustered together via unsupervised clustering resulting from gender-specific biases in methylation level.[23 24] Therefore, probes on X and Y chromosomes were removed to discard any sex bias within the samples. The number of sites annotated by probe types that were removed by the initial quality control step is shown in online supplementary table S2. A total of 76.88% probes remained for the TND-HIL

analysis and 81.82% remained for the BWS-HIL analysis after the preprocessing.

The β-values were converted to M-values by logit transformation as M-value increases the cogency of statistical tests for differential methylation.[25] Quantile normalisation was used to normalise signal intensities for each probe and reduce inter-array variation.[26]

Illumina Human 450 K methylation array uses two different chemistries, Infinium I and II, to enhance the breadth of coverage. Infinium I uses two probes per CpG locus (both methylated and unmethylated query probes), whereas in Infinium II only one probe (either methylated or unmethylated) per CpG locus is required. To correct these differences in the results between these two chemistries, peak correction was applied.[27] No batch correction was required as all the cases and controls for individual experiments had been processed in the same batch.

Low sample number differential methylation analysis

Stringent criteria were set to select candidate imprinted sequences hypomethylated in patients, with p values adjusted using false discovery rate to ensure statistical robustness.[28] Individual CpGs were selected when hypomethylated in patients compared with controls, with an adjusted p value of $>1.33×10^{-7}$, and an M-value between +1 and −1 (equivalent to 0.26≥β≥0.7) in normal controls. Genes containing two CpGs meeting these criteria and within <2000 nucleotides were deemed to be candidate DMRs.

Initially paired t test and one-sample t test were used for statistical analysis; however, these methods did not reveal any probes meeting our stringent criteria, probably because of the low sample number. Therefore, we explored the linear model technique, used for analysis of microarray data,[29] which models the significant part of the data and then allows the fitted coefficients to be compared in as many ways as possible. Crawford and Garthwaite proved that using a larger control group can produce significant statistical results even for a single case provided that appropriate statistical methods are applied.[30] Therefore, for both of the case groups, we used larger numbers of controls (n>200) against smaller numbers of cases (n=5). The linear model achieved convincing statistical outcomes from our pipeline, with efficient identification of known and novel hypomethylated loci for both TND-HIL and BWS-HIL case groups. Using the same criteria, only one region of hypermethylation was found in TND-HIL and four in BWS-HIL; these were not further examined as they were not relevant to this study (data not shown).

Targeted validation testing

msPCR analysis of the 11 maternally methylated loci used previously described primers and protocols.[12] msPCR primers for candidate loci NHP2L1, PPIEL and WRB are listed in online supplementary table S3.

Bisulfite sequencing

Bisulfite-specific primers were designed to amplify regions of 80–180 nt containing 7–12 CpG dinucleotides, using PyroMark software V.1.0 (Qiagen). Primer sequences are listed in online supplementary table S3. Amplicons were generated (Phusion DNA polymerase New England BioLabs) from two patients and two controls, ligated into pCR2.1 (Invitrogen); 2 μL of each ligation was transformed into chemically competent TOP10 cells (Invitrogen). Positive clones were selected on agar plates supplemented with 40 μg/mL X-gal and 100 μg/mL ampicillin. Overall, 24 white colonies were selected from each plate and

suspended in 50 μL dH_2O prior to denaturation (94°C for 5 min). An amount of 1 μL of the denatured bacterial solution was used as a PCR template for M13 primer amplification (Phusion DNA polymerase New England BioLabs). These reactions were treated with ExoSAP to degrade remaining primers, prior to sequencing with M13 forward and reverse primers. Very similar results were obtained for the two controls and the two patients; results from only one patient and one control are presented in the figures.

Restriction digest sequencing

To determine whether methylation was allele-specific or restricted by parent of origin, SNPs were analysed in proximity to DMRs in DNA from family trios. Heterozygous SNPs were identified and their inheritance determined by Sanger sequencing in DNA of off-spring and parents. To determine methylation status, 200 ng of offspring DNA was digested before amplification with restriction enzymes BstU1 or Mcrbc (New England Biolabs) according to manufacturer's instructions, as described.[31]

Expression analysis

Coding SNPs were identified within novel imprinting gene candidates WRB and NHP2L1 (rs13230 and rs8779, respectively). These were used to identify heterozygous samples collected following termination of pregnancy for a non-medical/social reason at gestational age 8–12 weeks with RNA-matched samples for a range of tissues (primers listed in online supplementary table S3). Allele-specific expression was then assessed in available heterozygous embryonic tissues.

cDNA was prepared with SuperScript III reverse transcriptase (Invitrogen) from 500 ng embryonic RNA. RT-PCR primers were designed to detect different isoforms of the candidate genes (see online supplementary table S3) and were amplified using Phusion DNA polymerase (New England BioLabs).

RESULTS
Statistical analysis of 450 K methylation array data

We developed a new analysis pipeline to detect methylation changes, with stringent selection criteria, capable of robust analysis of our small epigenetically defined groups (see Materials and methods section). The pipeline employed the linear modelling commonly used for microarray analysis and compared small patient numbers against a large control group to produce significant statistical results.[29] [30] Using stringent selection criteria, 34 hypomethylated regions were identified in the BWS-HIL cohort and 21 regions in TND-HIL (figure 1, see online supplementary tables S4 and S5).

The hypomethylated regions generated from both groups included several known imprinted genes (table 1, see online supplementary tables S4 and S5), both within and outside the 11 loci previously assessed in targeted analysis. The p values observed for known loci were proportionate to the degree of hypomethylation predicted from msPCR analysis of the patients groups. This is most clearly demonstrated at the disease-specific loci, where the lowest adjusted p value for the TND locus PLAGL1 was more significant in TND-HIL than BWS-HIL ($4.84×10^{-124}$ vs $4.39×10^{-51}$) (see online supplementary figure S2B, supplementary tables S4 and S5) and, conversely, the BWS locus KCNQ1OT1 had a lower p value in BWS-HIL than TND-HIL cohort ($4.27×10^{-68}$ vs $9.47×10^{-10}$) (see online supplementary tables S4 and S5). These p values were consistent with the degree of hypomethylation detected by targeted testing (see online supplementary table S1).

To assess the effect of merging patient data on the ability of the pipeline to detect hypomethylation, we used SNRPN, the
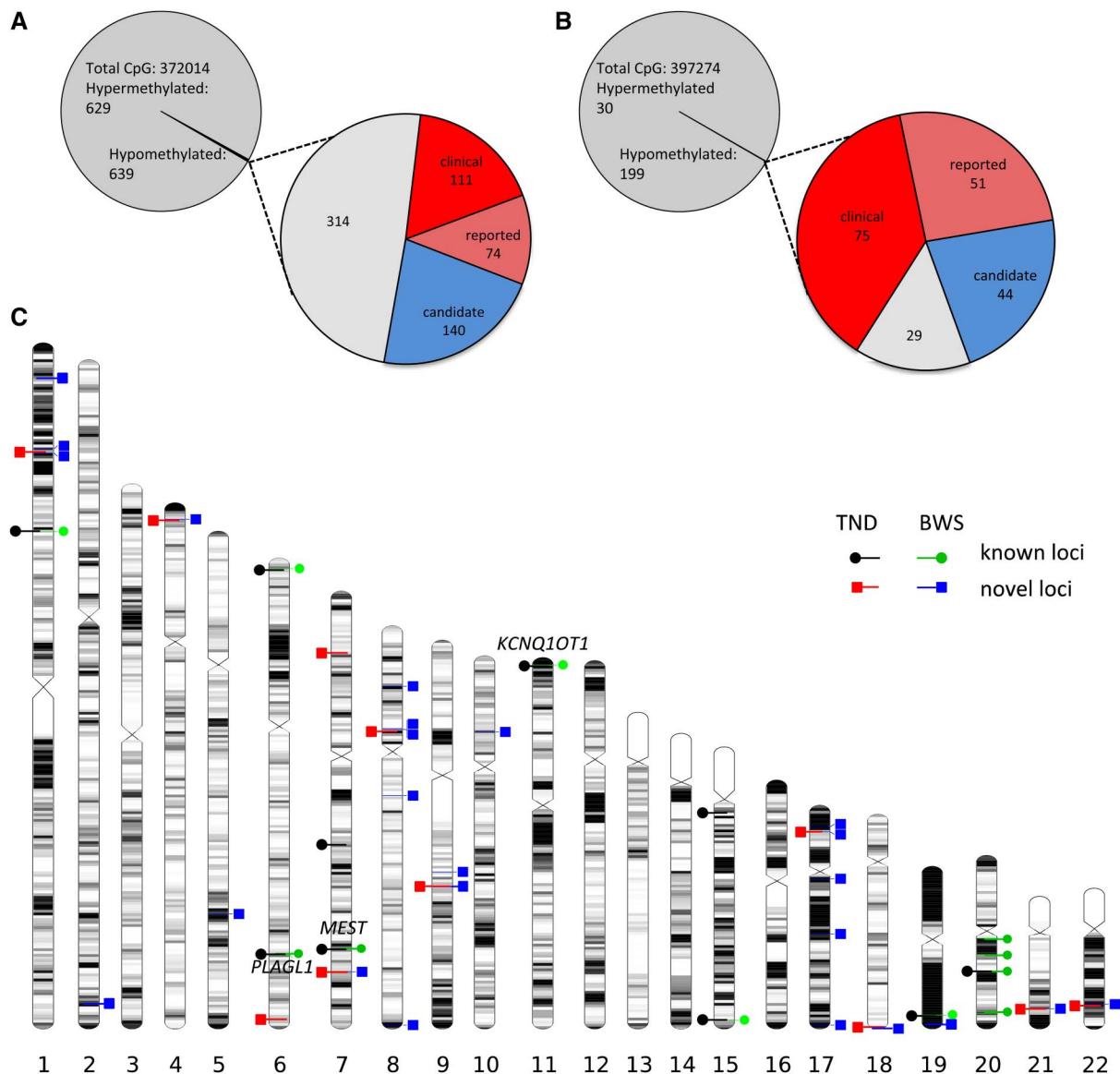
Figure 1 Distribution of known and candidate differentially methylated CpG sites in (A) Beckwith–Wiedemann syndrome (BWS) and (B) transient neonatal diabetes (TND). In each case, the pie chart to the left shows CpG sites compared between cases and controls (in grey), including those meeting criteria for differential methylation; the pie chart to the right highlights hypomethylated CpG sites, including those in known clinically-relevant loci (red), loci reported to be imprinted (pink) and loci not currently reported to be imprinted, that is, candidate loci (blue). (C) Chromosome ideogram showing the distribution across all autosomes of known and candidate differentially methylated loci. Black dots represent known imprinted genes that were shown to be hypomethylated in the TND patient group in this study; the green dots represent known imprinted genes shown to be hypomethylated in the BWS patient group in this study. Red and blue squares correspond to candidate imprinted loci in TND-HIL and BWS-HIL, respectively. The names of imprinted loci associated with imprinting disorders are displayed next to loci, in black, where they were detected as hypomethylated in patient samples.

only locus identified by msPCR in both patient groups with hypomethylation of a single patient (see online supplementary table S1). Using our criteria, hypomethylation of *SNRPN* was resolved in the TND-HIL, but not the BWS-HIL cohort where the hypomethylation was less severe (table 1, see online supplementary table S6). Thus, the pipeline was proved to resolve moderate hypomethylation in a single individual, validating the analysis of these hyper-rare patients as a group, rather than attempting analysis of single patients, which presents significant statistical challenges.

In addition to the known imprinted regions, 23 and 11 novel candidate DMRs were detected in the BWS-HIL and TND-HIL cohorts, respectively. Nine of these candidate DMRs were shared between BWS-HIL and TND-HIL patient groups (table 1). It is

noteworthy that the coverage of probes was broadly higher in known imprinted genes than novel candidates (eg, 54 in *PLAGL1*, 267 in *KCNQ1* and 73 in *MEST*, compared with 24 in *ERLIN2*, 28 in *WRB*, 23 in *NHP2L1* and 13 in *LOC728448*), reducing the likelihood of finding such novel candidates by chance.

Validation of differential methylation region candidates
Candidates were prioritised for follow-up based on prior evidence of allele-specific methylation in primary cell lines and hypomethylation in sperm (from Fang *et al*[32]) which would be consistent with maternal imprinting (this eliminated *JAKMIP1* and *GLP2R*). Further inspection highlighted three candidates (*NHP2L1*, *WRB* and *PPIEL*) where hypomethylation affected sequence contexts characteristic of imprinted genes (figures 2

Table 1 Hypomethylated regions shared between TND-HIL and BWS-HIL patients

Candidate	Chr	Gene name	CpG island	BWS			TND		
				Probe region*	No. probes†	Lowest p value‡	Probe region*	No. probes†	Lowest p value‡
Novel candidate DMRs	1	LOC728448/*PPIEL*	No	40 024 971–40 025 411	3	1.47E−18	40 024 971–40 025 232	2	3.09E−22
	4	JAKMIP1	Yes	6 107 021–6 107 339	4	2.48E−16	6 107 021–6 107 339	4	5.83E−36
	7	SVOPL	Yes	138 348 774–138 349 443	3	6.30E−41	138 348 774–138 349 443	3	7.21E−20
	9	FANCC	Yes	98 075 481–98 075 492	2	8.29E−58	98 075 481–98 075 492	2	7.28E−55
	17	GLP2R	No	9 729 250–9 729 424	3	3.33E−16	9 729 250–9 729 422	4	1.81E−23
	21	*WRB*	Yes	40 757 691–40 758 208	2	2.51E−20	40 757 691–40 758 208	4	6.71E−29
	8	LOC728024/ERLIN2	No	37 605 517–37 605 783	4	3.87E−40	37 605 359–37 605 978	6	2.69E−42'
	18	LOC100130522/PARD6G-AS1	Yes	77 905 355–77 905 947	3	1.01E−19	77 905 298–77 905 947	9	4.38E−71
	22	*NHP2L1*	Yes	42 078 217–42 078 723	6	4.08E−15	42 078 217–42 078 723	6	4.25E−54
Imprinted—not associated with ID	1	DIRAS3[43]	Yes	68 512 539–68 517 273	21	6.69E−31	68 512 539–68 517 273	20	5.45E−64
	6	*FAM50B*[20 44]	Yes	3 849 235–3 849 818	17	1.70E−18	3 849 272–3 849 818	17	1.64E−39
	15	IGF1R[45]	No	99 408 636–99 409 506	5	2.23E−15	99 408 636–99 409 957	6	1.04E−36
	19	ZNF331[46 47]	Yes	54 040 774–54 058 085	11	1.39E−40	54 040 813–54 058 085	10	9.13E−53
	20	L3MBTL[48]	Yes	42 142 417–42 143 502	13	1.32E−17	42 142 417–42 143 489	18	7.60E−25
Imprinted—associated with ID	6	*PLAGL1*	Yes	144 328 421–144 329 909	14	1.06E−55	144 328 482–144 329 909	15	1.22E−129
	7	MEST	Yes	130 130 187–130 133 110	42	6.12E−42	130 130 383–130 133 110	42	1.73E−45
	11	KCNQ1	Yes	2 715 837–2 722 258	26	1.14E−73	2 720 463–2 722 119	9	4.86E−13

Datasets from five patients with BWS-HIL and five with TND-HIL were compared with datasets from 245 and 211 batch-matched normal controls, respectively. Probes with M-values between −1 and +1 in controls and relative hypomethylation in patients with a p value of <1.33E−7 were identified. This subset was further filtered by minimal criteria for a hypomethylated locus, that is, ≥2 hypomethylated probes spaced by <2000 nucleotides. Candidate regions that meet these criteria in both BWS-HIL and TND-HIL are listed in this table.
*Genome position of most proximal and distal probe fulfilling hypomethylation criteria.
†Number of probes within the locus fulfilling hypomethylation criteria.
‡Minimum p value among probes fulfilling hypomethylation criteria.
BWS, Beckwith–Wiedemann syndrome; DMR, differentially methylated region; HIL, hypomethylation of imprinted loci; ID, imprinting disorder; TND, transient neonatal diabetes.
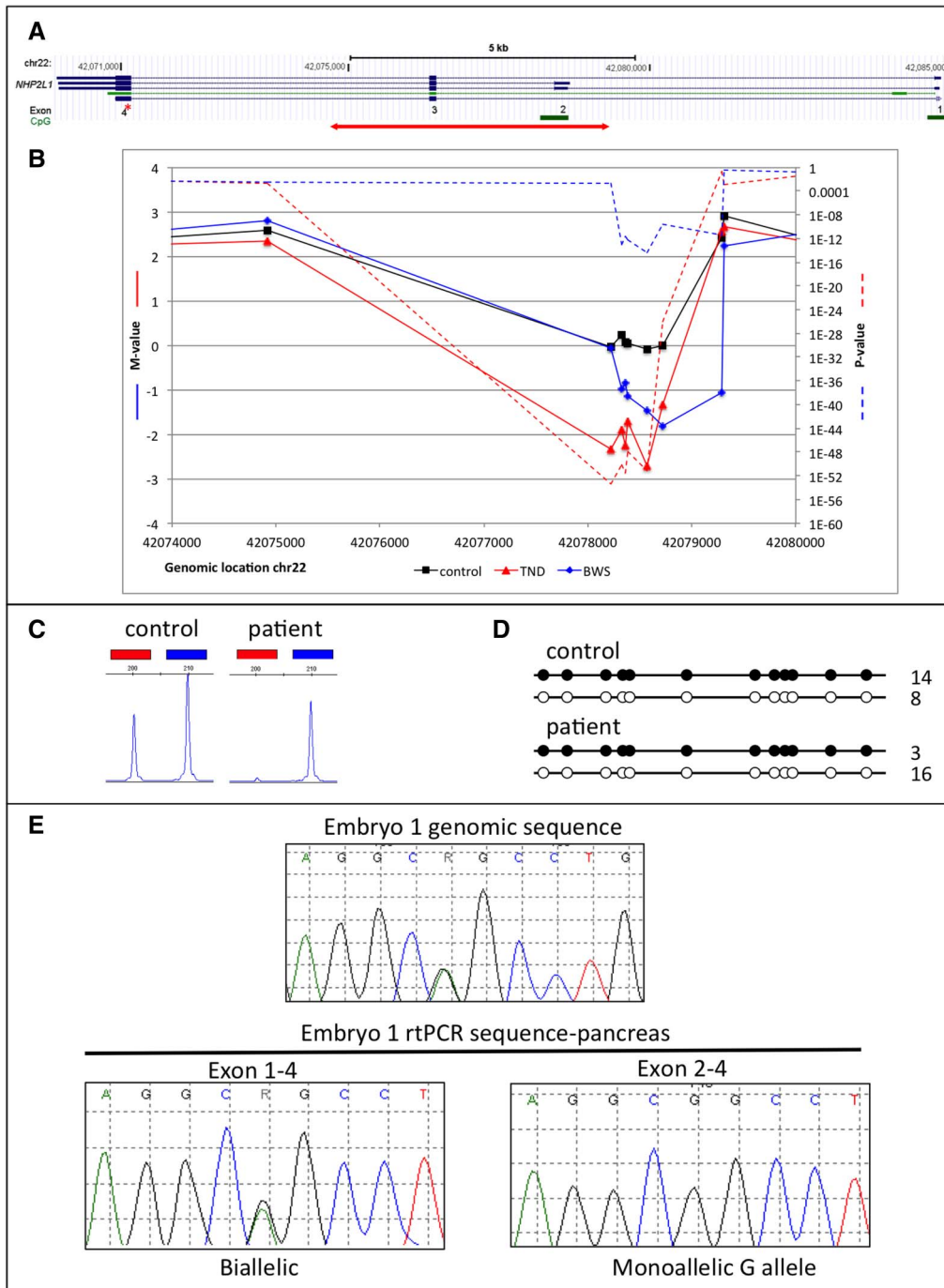
Figure 2 DNA methylation and expression analysis of *NHP2L1* in patients with Beckwith–Wiedemann syndrome (BWS) and transient neonatal diabetes (TND). (A) Screengrab from UCSC genome browser representing the *NHP2L1* gene and imprinted locus. The subregion highlighted in (B) is marked by a red double-ended arrow. Small numbers under the screengrab denote the exon numbering as used for expression analysis in (E); red asterisk indicates the position of the SNP analysed in (E). Note that *NHP2L1* is transcription from right to left with respect to genomic orientation. (B) Divergent DNA methylation between normal controls and patients, detected by methylation array. Solid lines denote M-values (left axis). Dashed lines represent p values of methylation difference between patients and controls (right axis). Black line represents normal controls; blue lines represent averaged methylation of five BWS patients; red lines represent averaged methylation of five TND patients. (C) Illustrative electropherogram from methylation-specific PCR experiment showing difference in DNA methylation between a single patient and control. Amplicons derived from methylated and unmethylated DNA are marked by red and blue lines, respectively. (D) Summary of bisulfite cloning and sequencing experiment comparing a patient with a normal control. The circles represent CpG dinucleotides within a sequence amplified after bisulfite modification, with filled and empty circles representing methylated and unmethylated DNA sequences respectively. The number to the right indicates the number of times the sequence was detected in individual clones. In no case were methylated and unmethylated CpG dinucleotides detected within a single clone. (E) Allele-specific expression analysis of *NHP2L1*. Top electropherogram represents genomic sequencing across rs8779 showing heterozygous SNP. Lower electropherograms represent sequencing of RT-PCR products from pancreatic cDNA, amplified from exons 1–4 (biallelic expression) and 2–4 (monoallelic).
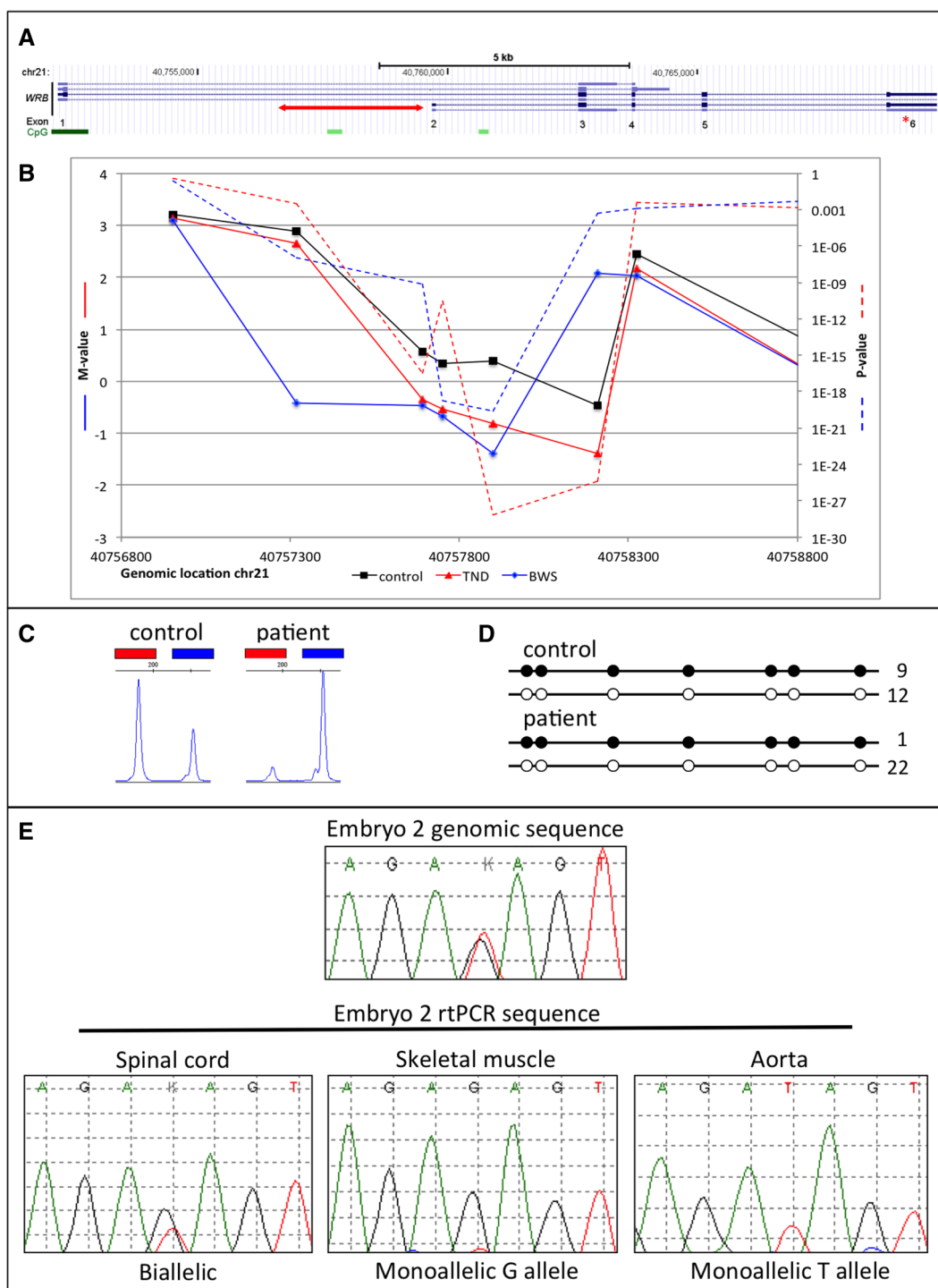
Figure 3 DNA methylation and expression analysis of *WRB* in patients with Beckwith–Wiedemann syndrome (BWS) and transient neonatal diabetes (TND). (A) Screengrab from UCSC genome browser, representing the *WRB* gene and imprinted locus. The subregion highlighted in (B) is marked by a red double-ended arrow. Small numbers under the screengrab denote the exon numbering as used for expression analysis in (E); red asterisk indicates the position of the SNP analysed in (E). (B) Divergent DNA methylation between normal controls and patients, detected by methylation array. Solid lines denote M-values (left axis). Dashed lines represent p values of methylation difference between patients and controls (right axis). Black line represents normal controls; blue lines represent averaged methylation of five BWS patients; red lines represent averaged methylation of five TND patients. (C) Illustrative electropherogram from methylation-specific PCR experiment, showing difference in DNA methylation between a single patient and control. Amplicons derived from methylated and unmethylated DNA are marked by red and blue lines, respectively. (D) Summary of bisulfite cloning and sequencing experiment comparing a patient with a normal control. The circles represent CpG dinucleotides within a sequence amplified after bisulfite modification, with filled and empty circles representing methylated and unmethylated DNA sequences, respectively. The number to the right indicates the number of times that sequence was detected in individual clones. In no case were methylated and unmethylated CpG dinucleotides detected within a single clone. (E) Allele-specific expression analysis of *WRB*. Top electropherogram represents genomic sequencing across rs1060180 showing heterozygous SNP. Lower electropherograms represent sequencing of RT-PCR amplicons in human fetal tissues as stated.

and 3, see online supplementary figure S3A). msPCR on a panel of 96 anonymised normal control samples showed methylation levels at all three loci to be stable in the normal population (SD *NHP2L1*=0.18, *WRB*=0.23 and *PPIEL*=0.22: data not shown). Analysis of complete hydatidiform mole (no methylation at maternally imprinted loci) showed complete hypomethylation in all three loci (data not shown).

DNA methylation at the candidate loci was then confirmed by msPCR in four of the five test HIL patients in each cohort (figures 2C and 3C; see online supplementary figure S3C; online supplementary table S1). For the two other patients, insufficient DNA remained for further analysis). All showed hypomethylation of at least one candidate locus: 2/4 TND-HIL patients were hypomethylated at all 3 loci, while 3/4 BWS-HIL and 1/4 TND-HIL patients showed hypomethylation at 2–3 loci. We then explored the methylation of these loci in DNA from further ID patients, including those with and without HIL, and those with hypomethylation of maternal and paternal DNA. Four of five additional TND-HIL patients and five of six additional BWS-HIL patients had hypomethylation at one or more loci, thus validating these as regions frequently affected by hypomethylation in TND-HIL and BWS-HIL patients (see online supplementary table S1). Less expected was the observation that *NHP2L1*, *WRB* and *PPIEL* candidate DMRs also showed hypomethylation in SRS-HIL patients (6/7, 4/7 and 1/7, respectively) and *WRB* hypomethylation in 1/1 PHP-HIL patient. No hypomethylation was observed at any of the loci in five patients with *ZFP57* mutations nor in nine patients with an ID affecting only one locus. This suggested that hypomethylation at these loci was restricted to HIL patients, rather than being widespread among ID patients.

Additionally, *WRB* methylation was analysed in 34 anonymised DNA samples from individuals diagnosed with Down syndrome. In all, 31 samples showed partial hypermethylation in a ratio consistent with the presence of one additional methylated allele of *WRB*; two showed partial hypomethylation consistent with one additional unmethylated allele of *WRB*; and one showed methylation equivalent to normal controls (see online supplementary figure S4). We were unable to confirm the parental origin of the additional chromosome 21 for these patients. However, given that 95% of trisomy 21 is of maternal origin,[33] we infer that this ratio of apparent hypermethylation and hypomethylation, at 31:2 Down syndrome patients (94%:6%), is consistent with DNA methylation being present on the maternal allele of *WRB*.

Parent of origin-specific methylation were investigated at *NHP2L1* and *PPIEL* candidate DMRs using methylation-specific restriction digest and sequencing. These results were consistent with maternal inheritance of the methylated allele at both candidate DMRs (see online supplementary figures S5 and S6). To further demonstrate that DNA methylation was discrete, that is, concentrated on one allele rather than homogeneously distributed, we performed bisulfite cloning and sequencing of *NHP2L1*, *WRB* and *PPIEL* DMRs. Amplicons from each candidate region were cloned and sequenced in two controls and two patients identified by msPCR as having hypomethylation. This confirmed the presence of fully-methylated and fully-unmethylated amplicons in controls, and relative hypomethylation in patient samples for all three candidate regions (figures 2D and 3D; see online supplementary figure S2D).

Validation of allele-specific expression

To determine whether the hypomethylation observed at the three candidate DMRs correlated with allele-specific expression

of the associated genes, we analysed expression of transcripts in human foetal nucleic acids. We identified informative SNPs in *NHP2L1* and *WRB* in the genomic DNA of 8–12 week embryos (we could not identify informative coding SNPs in *PPIEL*). Matched RNA from multiple tissues was reverse-transcribed and amplified by RT-PCR using isoform-specific primers.

For *NHP2L1*, monoallelic expression was observed for exon 2–4 specific transcripts and biallelic expression for exon 1–4 specific transcripts (figure 2E) in all tested tissues for four embryos (data not shown). Biallelic expression of *WRB* was observed in the majority of tissues tested with both exon 1–6 and 2–6 specific transcripts. However, sporadic monoallelic expression was observed with opposing allelic expression in the skeletal muscle and aorta of a single embryo (exon 1–6 specific primers: figure 3E), and monoallelic expression in 1/3 adrenal tissues assayed (exon 2–6 specific primers; data not shown).

DISCUSSION

The data presented here demonstrate the successful use of whole genome methylation array technology to explore the methylome in two rare epigenetically defined cohorts of patients with IDs characterised by HIL.

Our small cohort size necessitated the development of a new pipeline capable of robust analysis of small group sizes. While other statistical analyses could not significantly detect hypomethylated loci, the linear model we applied in the pipeline, with the stringent criteria, detected differential methylation robustly. These loci were validated by the evidence from the prior partial epigenetic profiling of our patient groups and low p values. Moreover, these p values were proportionate to the degree of hypomethylation predicted from the known patient epimutations. This allowed us to use the pipeline confidently to predict novel imprinted regions.

Consistent with the aim of this study, novel candidate DMRs were identified that share several attributes of imprinted genes. From the nine candidate DMRs identified, follow-up of three candidates did not validate hypomethylation in the patients analysed by 450 K methylation array. These loci showed hypomethylation in additional TND-HIL and BWS-HIL patients, but not in patients with hypomethylation restricted to one primary locus or in normal controls. Hypomethylation of all loci in individuals with SRS-HIL and *WRB* in a PHP-HIL patient expanded the range of patients observed to have hypomethylation at these regions. Additionally, allele-specific methylation and parent-specific methylation analysis was consistent with monoallelic methylation of maternal origin for all three candidate DMRs, with *NHP2L1* and *WRB* showing evidence of allele-specific expression.

It is noteworthy that patterns of hypomethylation were shared between HIL patients with divergent clinical presentations. This is a surprising observation, but consistent with a shared cause of their syndromic presentation. It has become apparent in recent years that IDs with common phenotypes are associated with multiple imprinted genes (eg, H19 and *KCNQ1OT1* in BWS, and H19 and chr7 in SRS: refs[34 35]). It is also apparent that some patients with HIL have clinical features inconsistent with their epigenotype.[14 36 37] There may be several reasons for this phenotype–epigenotype divergence, but the most likely is somatic mosaicism, which is common among IDs and strongly modifies clinical presentation. It is therefore possible that common underlying causes, including environmental insults, primary epimutations and trans-acting mutations, may cause HIL disorders with highly variable phenotypic features. Comprehensive epigenetic profiling may be required to

stratify HIL patients with common epimutation patterns and seek subtle clinical overlaps. Such stratification may support exome analysis for common genetic causes, and moreover identify further epimutations that may account for some of their additional clinical features. It may also be informative to compare epigenotype patterns among patients of different genetic aetiologies. In this regard, it is interesting that an epigenetic analysis of a patient whose mother had an *NLRP7* mutation showed very limited overlap of affected imprinted genes (*FAM50B*) alone with our patients, but some shared hypomethylation of non-imprinted genes which may inform differences in clinical presentation.[38]

Of the three candidate imprinted loci described here, none has a well-defined role in either normal physiology or a disease process. *NHP2L1* is a nuclear protein which plays a role in pre-mRNA splicing as a component of the U4/U6-U5 tri-snRNP[39] and shows evidence of allele-specific methylation.[32] Little is known about the function of *PPIEL* (pseudogene of peptidylprolyl isomerase E) but aberrant DNA methylation at *PPIEL* has previously been associated with bipolar disorder with a reported strong inverse correlation between gene expression and DNA methylation levels of *PPIEL*.[40] *WRB* encodes a basic nuclear protein of unknown function and maps to the region associated with congenital heart disease in Down syndrome.[41 42] The clinical relevance of these loci, if any, is unknown. It is possible that these genes, or any of the others identified as hypomethylated in our study, could be associated additional clinical disorders beyond the eight IDs currently known in clinical genetics. Cardiac disorders have been reported in 9% of a TND cohort,[13] and it is possible that analysis of further patients will reveal whether the involvement of this locus is of clinical significance.

There were several potential limitations to our study. First, whole genome methylation analysis by array is restrictive to the sequences captured on the array: many more candidate imprinted regions may have potentially been obtained from whole genome bisulfite sequencing; second, additional HIL cohorts with other IDs may have provided further candidates; third, the grouping of disease cases was necessary for statistical purposes, but may have masked the hypomethylation of less strongly-affected loci. For the candidate regions that have been identified there are further limitations to expression analysis in the form of low frequency SNPs and potentially imprinted transcript identification. DNA methylation is only one component of the cellular machinery of imprinting, and the methylation signature does not necessarily colocate with the gene(s) under its control, or as has been observed in the case of the candidate region *PPIEL*, not even residing within a CpG island.

Further work is required to exploit the findings of this study. The candidate imprinted loci identified here must be characterised to determine whether their epimutation has any bearing on clinical features in the context of HIL or in as-yet undescribed ID. These or similar patients may be more comprehensively analysed by whole genome bisulfite sequencing to increase capture of candidate genes. Greater resolution may also be obtained if a bioinformatic pipeline can be developed for statistically robust analysis of individuals, rather than groups of patients; indeed, such analysis might be the basis for a comprehensive clinical genetic diagnosis of HIL. Analysis of further patients may support accurate stratification of patient groups with common epigenetic signatures—with or without common phenotype. This in turn would support the search for candidate trans-acting gene mutations by exome analysis. Identification of common DNA motifs in hypomethylated loci may also indicate

association with common trans-acting factors (by analogy with *ZFP57*), and such motifs would be the focus for cis-acting mutations in IDs. Overall, the potential benefits are disproportionate to the rarity of the patients being analysed, and may include novel insight into the basic mechanisms of human epigenetics, as well as novel loci that may be implicated in many other disorders including Down Syndrome and bipolar disorder.

REFERENCES

1 Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* 2012;335:1499–502.
2 Seisenberger S, Peat JR, Hore TA, Santos F, Dean W, Reik W. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Phil Trans Roy Soc Lond B* 2013;368:20110330.
3 Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* 2011;43:811–14.
4 Peters J, Beechey C. Identification and characterisation of imprinted genes in the mouse. *Brief Funct Genomic Proteomic* 2004;2:320–33.
5 Yamazawa K, Ogata T, Ferguson-Smith AC. Uniparental disomy and human disease: an overview. *Am J Med Genet* 2010;154C:329–34.
6 Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 2010;329:643–8.
7 DeVeale B, van der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet* 2012;8:e1002600.
8 Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, Plomin R, Mill J. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 2010;86:196–212.
9 Kelsey G, Bartolomei MS. Imprinted genes … and the number is? *PLoS Genet* 2012;8:e1002601.
10 Horsthemke B. Mechanisms of imprint dysregulation. *Am J Med Genet* 2010;154C:321–8.
11 Eggermann T, Leisten I, Binder G, Begemann M, Spengler S. Disturbed methylation at multiple imprinted loci: an increasing observation in imprinting disorders. *Epigenomics* 2011;3:625–37.
12 Poole RL, Docherty LE, Al Sayegh A, Caliebe A, Turner C, Baple E, Wakeling E, Harrison L, Lehmann A, Temple IK, Mackay DJG. Targeted methylation testing of a patient cohort broadens the epigenetic and clinical description of imprinting disorders. *Am J Med Genet* 2013;161A:2174–82.
13 Docherty LE, Kabwama S, Lehmann A, Hawke E, Harrison L, Flanagan SE, Ellard S, Hattersley AT, Shield JP, Ennis S, Mackay DJG, Temple IK. Clinical presentation of

6q24 transient neonatal diabetes mellitus (6q24 TNDM) and genotype-phenotype correlation in an international cohort of patients. *Diabetologia* 2013;56:758–62.

14 Azzi S, Rossignol S, Steunou V, Sas T, Thibaud N, Danton F, Le Jule M, Heinrichs C, Cabrol S, Gicquel C, Le Bouc Y, Netchine I. Multilocus methylation analysis in a large cohort of 11p15-related foetal growth disorders (Russell Silver and Beckwith Wiedemann syndromes) reveals simultaneous loss of methylation at paternal and maternal imprinted loci. *Hum Mol Genet* 2009;18:4724–33.

15 Mackay DJG, Callaway JLA, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JMD, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in *ZFP57*. *Nat Genet* 2008;40:949–51.

16 Parry DA, Logan CV, Hayward BE, Shires M, Landolsi H, Diggle C, Carr I, Rittore C, Touitou I, Philibert L, Fisher RA, Fallahian M, Huntriss JD, Picton HM, Malik S, Taylor GR, Johnson CA, Bonthron DT, Sheridan EG. Mutations causing familial biparental hydatidiform mole implicate c6orf221 as a possible regulator of genomic imprinting in the human oocyte. *Am J Hum Genet* 2011;89:451–8.

17 Van den Veyver IB, Al-Hussaini TK. Biparental hydatidiform moles: a maternal effect mutation affecting imprinting in the offspring. *Hum Reprod Update* 2006;12:233–42.

18 Judson H, Hayward BE, Sheridan E, Bonthron DT. A global disorder of imprinting in the human female germ line. *Nature* 2002;416:539–42.

19 Choufani S, Shapiro JS, Susiarjo M, Butcher DT, Grafodatskaya D, Lou Y, Ferreira JC, Pinto D, Scherer SW, Shaffer LG, Coullin P, Caniggia I, Beyene J, Slim R, Bartolomei MS, Weksberg R. A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. *Genome Res* 2011;21:465–76.

20 Court F, Martin-Trujillo A, Romanelli V, Garin I, Iglesias-Platas I, Salafsky I, Guitart M, Perez de Nanclares G, Lapunzina P, Monk D. Genome-wide allelic methylation analysis reveals disease-specific susceptibility to multiple methylation defects in imprinting syndromes. *Hum Mut* 2013;34:595–602.

21 Nakabayashi K, Trujillo AM, Tayama C, Camprubi C, Yoshida W, Lapunzina P, Sanchez A, Soejima H, Aburatani H, Nagae G, Ogata T, Hata K, Monk D. Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. *Hum Mol Genet* 2011;20:3188–97.

22 Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S. IMA: an R package for high-throughput analysis of Illumina's 450 K Infinium methylation data. *Bioinformatics* 2012;28:729–30.

23 Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PloS One* 2009;4:e6767.

24 El-Maarri O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, Wienker T, Oldenburg J. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet* 2007;122:505–14.

25 Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;11:587.

26 Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M, Kravariti E, Toulopoulou T, Murray RM, Mill J. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet* 2011;20:4786–96.

27 Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450 K technology. *Epigenomics* 2011;3:771–84.

28 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;57:289–300.

29 Smyth GK. Limma: linear models for microarray data. In: Gentleman VR, Carey SD, Irizarry R, Huber W. eds. *Bioinformatics and computational biology solutions using R and bioconductor*. New York: Springer, 2005:397–420.

30 Crawford JR, Garthwaite PH. Single-case research in neuropsychology: a comparison of five forms of t-test for comparing a case to controls. *Cortex* 2012;48:1009–16.

31 Poole RL, Leith DJ, Docherty LE, Shmela ME, Gicquel C, Temple IK, Mackay DJG. Beckwith-Wiedemann syndrome caused by maternally-inherited mutation of an OCT-binding motif in the IGF2/H19 imprinting control region, ICR1. *Eur J Hum Genet* 2012;20:240–3.

32 Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. *PNAS USA* 2012;109:7332–7.

33 Antonarakis SE. Parental origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. Down Syndrome Collaborative Group. *NEJM* 1991;324:872–6.

34 Choufani S, Shuman C, Weksberg R. Beckwith-Wiedemann syndrome. *Am J Med Genet* 2010;154C:343–54.

35 Eggermann T. Russell-Silver syndrome. *Am J Med Genet* 2010;154C:355–64.

36 Boonen SE, Mackay DJ, Hahnemann JM, Docherty L, Grønskov K, Lehmann A, Larsen LG, Haemers AP, Kockaerts Y, Dooms L, Vu DC, Ngoc CT, Nguyen PB, Kordonouri O, Sundberg F, Dayanikli P, Puthi V, Acerini C, Massoud AF, Tümer Z, Temple IK. Transient neonatal diabetes, *ZFP57*, and hypomethylation of multiple imprinted loci: a detailed follow-up. *Diabetes Care* 2013;36:505–12.

37 Scott RH, Douglas J, Baskcomb L, Huxter N, Barker K, Hanks S, Craft A, Gerrard M, Kohler JA, Levitt GA, Picton S, Pizer B, Ronghe MD, Williams D; Factors Associated with Childhood Tumours (FACT) Collaboration, Cook JA, Pujol P, Maher ER, Birch JM, Stiller CA, Pritchard-Jones K, Rahman N. Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. *Nat Genet* 2008;40:1329–34.

38 Beygo J, Ammerpohl O, Gritzan D, Heitmann M, Rademacher K, Richter J, Caliebe A, Siebert R, Horsthemke B, Buiting K. Deep bisulfite sequencing of aberrantly methylated loci in a patient with multiple methylation defects. *PLOS ONE* 2013;8:e76953.

39 Liu S, Rauhut R, Vornlocher HP, Luhrmann R. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* 2006; 12:1418–30.

40 Kuratomi G, Iwamoto K, Bundo M, Kusumi I, Kato N, Iwata N, Ozaki N, Kato T. Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol Psych* 2008;13:429–41.

41 Egeo A, Mazzocco M, Sotgia F, Arrigo P, Oliva R, Bergonon S, Nizetic D, Rasore-Quartino A, Scartezzini P. Identification and characterization of a new human cDNA from chromosome 21q22.3 encoding a basic nuclear protein. *Hum Genet* 2998;102:289–93.

42 Murata K, Degmetich S, Kinoshita M, Shimada E. Expression of the congenital heart disease 5/tryptophan rich basic protein homologue gene during heart development in medaka fish, Oryzias latipes. *Dev Growth Differ* 2009;51:95–107.

43 Yu Y, Xu F, Peng H, Fang X, Zhao S, Li Y, Cuevas B, Kuo WL, Gray JW, Siciliano M, Mills GB, Bast RC Jr. NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *PNAS USA* 1999;96:214–9.

44 Zhang A, Skaar DA, Li Y, Huang D, Price TM, Murphy SK, Jirtle RL. Novel retrotransposed imprinted locus identified at human 6p25. *NAR* 2011;39:5388–400.

45 Sharp AJ, Migliavacca E, Dupre Y, Stathaki E, Sailani MR, Baumer A, Schinzel A, Mackay DJ, Robinson DO, Cobellis G, Cobellis L, Brunner HG, Steiner B, Antonarakis SE. Methylation profiling in individuals with uniparental disomy identifies novel differentially methylated regions on chromosome 15. *Genome Res* 2010;20:1271–8.

46 Noguer-Dance M, Abu-Amero S, Al-Khtib M, Lefevre A, Coullin P, Moore GE, Cavaille J. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum Mol Genet* 2010;19:3566–82.

47 Pollard KS, Serre D, Wang X, Tao H, Grundberg E, Hudson TJ, Clark AG, Frazer K. A genome-wide approach to identifying novel-imprinted genes. *Hum Genet* 2008;122:625–34.

48 Li J, Bench AJ, Vassiliou GS, Fourouclas N, Ferguson-Smith AC, Green AR. Imprinting of the human *L3MBTL* gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. *PNAS USA* 2004;101:7341–6.