AJHG The American Journal of Human Genetics

# DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | AJHG-D-16-00165R3 |
| **Full Title:** | DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome |
| **Article Type:** | Article |
| **Keywords:** | Genomic imprinting; uniparental disomy; 15q11.2; Prader-Willi syndrome; Angelman syndrome |
| **Corresponding Author:** | Andrew James Sharp, Ph.D<br>Icahn School of Medicine at Mount Sinai<br>New York, UNITED STATES |
| **First Author:** | Ricky Shantilal Joshi, Ph.D |
| **Order of Authors:** | Ricky Shantilal Joshi, Ph.D |
| | Paras Garg, MSc |
| | Noah Zaitlen, Ph.D |
| | Tuuli Lappalainen, Ph.D |
| | Corey Watson, Ph.D |
| | Nidha Azam, MSc |
| | Daniel Ho Ho, BA |
| | Xin Li, Ph.D |
| | Stylianos Antonarakis, Ph.D |
| | Han Brunner, Ph.D |
| | Karin Buiting, Ph.D |
| | Sau Wei Cheung, Ph.D |
| | Bradford Coffee, Ph.D |
| | Thomas Eggermann, Ph.D |
| | David Francis, Ph.D |
| | Joep Geraedts, Ph.D |
| | Giorgio Gimelli, Ph.D |
| | Samuel Jacobson, Ph.D |
| | Cedric Le Caignec, Ph.D |
| | Nicole de Leeuw, Ph.D |
| | Thomas Liehr, Ph.D |
| | Deborah Mackay, Ph.D |
| | Stephen Montgomery, Ph.D |
| | Alistair Pagnamenta, Ph.D |
| | Peter Papenhausen, Ph.D |
| | David Robinson, Ph.D |
| | Claudia Ruivenkamp, Ph.D |

| | |
|---|---|
| | Charles Schwartz, Ph.D |
| | Bernhard Steiner, Ph.D |
| | David Stevenson, Ph.D |
| | Urvashi Surti, Ph.D |
| | Thomas Wassink, Ph.D |
| | Andrew James Sharp, Ph.D |
| **Abstract:** | Genomic imprinting is a mechanism in which gene expression varies depending on parental origin. Imprinting occurs through differential epigenetic marks on the two parental alleles, with most imprinted loci marked by the presence of differentially methylated regions (DMRs). To identify sites of parental epigenetic bias, here we have profiled DNA methylation patterns in a cohort of 57 individuals with uniparental disomy (UPD) for 19 different chromosomes, defining imprinted DMRs as sites where the maternal and paternal methylation levels diverge significantly from the biparental mean. Using this approach we identified 77 DMRs, including nearly all those described in previous studies, in addition to 34 DMRs not previously reported. These include a DMR at TUBGCP5 within the recurrent 15q11.2 microdeletion region, suggesting potential parent-of-origin effects associated with this genomic disorder. We also observed a modest parental bias in DNA methylation levels at every CpG analyzed across ~1.9Mb of the 15q11-q13 Prader-Willi/Angelman syndrome region, demonstrating that the influence of imprinting is not limited to individual regulatory elements such as CpG islands, but can extend across entire chromosomal domains. Using RNAseq data we detected signatures consistent with imprinted expression associated with nine novel DMRs. Finally, using a population sample of 4,004 blood methylomes we define patterns of epigenetic variation at DMRs, identifying rare individuals with global gain or loss of methylation across multiple imprinted loci. Our data provide a detailed map of parental epigenetic bias in the human genome, providing insights into potential parent-of-origin effects. |

Icahn School of Medicine at Mount Sinai

Dr Andrew Sharp, PhD, Associate Professor
The Mindich Child Health & Development Institute
Department of Genetics and Genomic Sciences
Icahn School of Medicine at Mount Sinai
Hess Center for Science and Medicine,
1470 Madison Avenue, Room S8-116, Box 1498
New York, NY 10029 USA
Office: +1-212-824-8942
Fax: +1-646-537-8527
Email: andrew.sharp@mssm.edu

## The American Journal of Human Genetics Decision AJHG-D-16-00165R2

**"DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome"**

June 28th 2016

Dear Editor,

I have uploaded a revised version of the manuscript addressing all points raised. Please do not hesitate to contact me if you require any further details.

Sincerely,

**AJHG-D-16-00165R2: “DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome” by Joshi et al.**

**RESPONSE TO EDITORS COMMENTS**

*Please address the following formatting requirements:*

*1. Accept all changes, remove line numbers, and turn off the 'Track changes while editing' feature of MS Word.*

This has been done.

*2. Please provide a single PDF that contains all supplemental figures and legends, supplemental tables, and Supplemental References (in this order). If a supplemental table cannot fit onto two 8.5" x 11" pages, please instead supply the table as an Excel file (its title & legend should appear within the supplemental PDF). Please do NOT include the title or author list in the PDF; we will add a coversheet with this information. In addition, please do not include page numbers in your final PDF.*

This has been done.

# DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome

Ricky S. Joshi[1]*, Paras Garg[1]*, Noah Zaitlen[2], Tuuli Lappalainen[3], Corey T. Watson[1], Nidha Azam[1], Daniel Ho[1,#], Xin Li[4], Stylianos E. Antonarakis[5], Han G. Brunner[6], Karin Buiting[7], Sau Wei Cheung[8], Bradford Coffee[9], Thomas Eggermann[10], David Francis[11], Joep P. Geraedts[12], Giorgio Gimelli[13], Samuel G. Jacobson[14], Cedric Le Caignec[15], Nicole de Leeuw[6], Thomas Liehr[16], Deborah J. Mackay[17], Stephen B. Montgomery[4], Alistair T. Pagnamenta[18], Peter Papenhausen[19], David O. Robinson[17], Claudia Ruivenkamp[20], Charles Schwartz[21], Bernhard Steiner[22], David A. Stevenson[23], Urvashi Surti[24], Thomas Wassink[25], Andrew J. Sharp[1]

[1] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA

[2] Department of Medicine, UCSF MC2552, 1700 4th Street, Byers Hall Suite 503C, SF, CA 94158, USA

[3] New York Genome Center, 101 Avenue of the Americas, 7th floor, New York, NY 10013, USA; Department of Systems Biology, Columbia University, New York, New York 10032, USA

[4] Departments of Pathology, Genetics and Computer Science, Stanford University School of Medicine, Stanford, CA 94305, USA

[5] Department of Genetic Medicine and Development, University of Geneva Medical School, 9th floor, 1, rue Michel-Servet, CH – 1211 Geneva, Switzerland

[6] Department of Human Genetics, Radboud University Medical Center, P.O. Box 9101, 6500 HB, Nijmegen, the Netherlands

[7] Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Hufelandstrasse 55, D-45122 Essen, Germany

[8] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

[9] Emory Genetics Laboratory, Emory University, Atlanta, Georgia, USA

[10] Institute of Human Genetics, University Hospital, RWTH, Aachen, Germany

[11] Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville 3052, Melbourne, Australia.

[12] Department of Genetics and Cell Biology, Research Institute GROW, Faculty of Health, Medicine and Life Sciences, Maastricht University, PO Box 5800, Maastricht, AZ 6202, The Netherlands

[13] Laboratorio di Citogenetica, Istituto G. Gaslini, Genova, Italy

[14] Scheie Eye Institute, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, 51 N. 39th Street, 19104, Philadelphia, PA, USA

[15] CHU Nantes, Service de Génétique Médicale, Institut de Biologie, 9 quai Moncousu, 44093 Nantes, France; INSERM, UMR 957, Nantes 44035, France; Université de Nantes, Nantes atlantique universités, Pathophysiology of Bone Resorption and Therapy of Primary Bone Tumours, Nantes, France

[16] Jena University Hospital, Friedrich Schiller University, Institute of Human Genetics, Kollegiengasse 10, D-07743 Jena, Germany

[17] Wessex Regional Genetics Laboratory Salisbury District Hospital, Salisbury, Wilts. SO2 8BJ

[18] National Institute for Health Research Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

[19] Division of Cytogenetics, LabCorp, Center for Molecular Biology and Pathology, Research Triangle Park, NC, USA.

[20] Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

[21] J.C. Self Research Institute, Greenwood Genetic Center, Greenwood, South Carolina, USA

[22] Institute of Medical Genetics, University of Zurich, Schwerzenbach, Switzerland

[23] Division of Medical Genetics, Lucile Salter Packard Children's Hospital, 300 Pasteur Drive, Boswell Building A097, Stanford, CA 94304

[24] Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA

[25] Department of Psychiatry, University of Iowa Iowa City, Iowa

[#] Present address: Novartis Institutes for Biomedical Research, 100 Technology Square, 8503C, Cambridge, MA 02139

* These authors contributed equally

Address for correspondence: Andrew J. Sharp, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, Hess Center for Science and Medicine, 1470 Madison Avenue, Room S8-116, Box 1498, New York, NY 10029 USA. Telephone: +1-212-824-8942, Fax: +1-646-537-8527, Email: andrew.sharp@mssm.edu

Short title: Imprinting in UPD cases

Keywords: Genomic imprinting, uniparental disomy, 15q11.2, Prader-Willi syndrome, Angelman syndrome

## Abstract

Genomic imprinting is a mechanism in which gene expression varies depending on parental origin. Imprinting occurs through differential epigenetic marks on the two parental alleles, with most imprinted loci marked by the presence of differentially methylated regions (DMRs). To identify sites of parental epigenetic bias, here we have profiled DNA methylation patterns in a cohort of 57 individuals with uniparental disomy (UPD) for 19 different chromosomes, defining imprinted DMRs as sites where the maternal and paternal methylation levels diverge significantly from the biparental mean. Using this approach we identified 77 DMRs, including nearly all those described in previous studies, in addition to 34 DMRs not previously reported. These include a DMR at *TUBGCP5* within the recurrent 15q11.2 microdeletion region, suggesting potential parent-of-origin effects associated with this genomic disorder. We also observed a modest parental bias in DNA methylation levels at every CpG analyzed across ~1.9Mb of the 15q11-q13 Prader-Willi/Angelman syndrome region, demonstrating that the influence of imprinting is not limited to individual regulatory elements such as CpG islands, but can extend across entire chromosomal domains. Using RNAseq data we detected signatures consistent with imprinted expression associated with nine novel DMRs. Finally, using a population sample of 4,004 blood methylomes we define patterns of epigenetic variation at DMRs, identifying rare individuals with global gain or loss of methylation across multiple imprinted loci. Our data provide a detailed map of parental epigenetic bias in the human genome, providing insights into potential parent-of-origin effects.

## Introduction

Genomic imprinting is a mechanism in which the expression of genes varies depending on the gender of the parent from which they are inherited. Imprinting is mediated via an epigenetic mechanism involving differential DNA methylation and histone modifications on the two parental alleles, with most imprinted genes marked by CpG-rich differentially methylated regions (DMRs).[1] Although some reports suggest >100 genes show evidence of imprinting in human[2], recent genome-wide surveys of imprinted DNA methylation and gene expression have only been able to confirm about half this number[3,4].

Uniparental disomy (UPD) is a condition where both homologs of a chromosome pair are inherited from the same parent, either from the mother (maternal, or UPDmat) or from the father (paternal, or UPDpat). UPD can lead to functional nullisomy for imprinted genes, resulting in clinically recognizable syndromes such as Prader-Willi/Angelman syndrome (PWS [MIM: 176270], AS [MIM: 105830]), UPD(15)mat/pat), Beckwith-Wiedemann (BWS [MIM: 130650], UPD(11)pat) or transient neonatal diabetes mellitus (TNDM, [MIM: 601410], UPD(6)pat).[5,6] In contrast, UPD for other chromosomes has not been associated with phenotypic consequences, leading to the presumption that these chromosomes do not harbor imprinted genes.[7]

Several studies have utilized allele-specific gene expression as a method of detecting imprinted expression in humans,[4,8-11] although these approaches can be hindered by the requirement for informative polymorphisms. However, in model organisms the use of reciprocal inter-strain or inter-species hybrids can generate offspring with extremely high rates of heterozygosity, allowing the efficient detection of parentally-biased expression.[12,13]

Alternative methods to detect imprinting have used the fact that the maternal and paternal genomes have differential epigenetic marks at most imprinted loci. This approach has the advantage over expression-based methods, in that these differential methylation marks are often preserved even in tissues that lack imprinted expression.[14] Because one of the key features of imprinted genes is the presence of parent-of-origin-specific methylation,

the systematic comparison of DNA methylation patterns in maternal versus paternal chromosomes provides a powerful approach for the identification of imprinted loci that does not rely on the availability of polymorphisms. We previously applied this method to study parental epigenetic bias on chromosome 15,[15] and it has since been utilized in several other studies.[3,16,17]

We have assembled a collection of 53 individuals with maternal and paternal UPD and four Turner syndrome samples for 19 different chromosomes, allowing the efficient detection of DMRs associated with imprinted genes for most of the human genome. This cohort is further supplemented by additional tissue samples with uniparental inheritance for every chromosome. Here we report the identification of 77 DMRs in the human genome that show consistent parental epigenetic bias. Using RNAseq data from 33 tissues, we further show that many of these loci show evidence of imprinted expression. Finally, we perform a population-scale methylome analysis to define patterns of normal epigenetic variation at imprinted DMRs, and identify rare individuals who exhibit global epigenetic shifts across multiple imprinted loci.

## Materials and Methods

### *Sample cohort*

DNA extracted from peripheral blood was obtained from 53 cases with uniparental disomy (UPD) for one or more autosomal chromosomes, and from four individuals with Turner syndrome (45,X karyotype, with a single X chromosome of either maternal or paternal origin). In each case uniparental disomy for the entire chromosome was confirmed by either microsatellite and/or SNP array analysis by comparison with parental DNA samples. In addition, we also tested tissue samples from four Ovarian Teratomas (OT) and four samples of Complete Hydatidiform Mole (CHM). OTs possess a 46,XX karyotype that is completely maternally derived, resulting from a meiotic error during oocyte maturation that results in the formation of a cyst containing tissues from each of the three embryonic germ cell layers. In contrast, sporadic CHMs possess a 46,XX karyotype that is entirely paternally derived, arising from an anucleate oocyte that is fertilized by a sperm that subsequently duplicates its haploid genome. A list of samples used for genome-wide methylation profiling is shown in Supplemental Figure 1. The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation, and this study was approved by the IRB of the Icahn School of Medicine at Mount Sinai (Study ID 12-00567).

### *Illumina Infinium HumanMethylation450 BeadChip array analysis*

DNA methylation profiling was performed using the Illumina Infinium HumanMethylation450 BeadChip platform (450k array) according to manufacturer's instructions (Illumina, San Diego, CA, USA). Data were processed using the methylation module of GenomeStudio v1.9 software using default parameters. 470,681 autosomal and 11,122 X-linked probe sequences (50-mer oligonucleotides) out of total 482,421 were uniquely mapped to the reference human genome hg19 (GRCh37) using BSMAP,[18] allowing alignments with up to 2bp of mismatch and 3bp of gaps. To reduce hybridization artefacts resulting from

mismatches between the probe and target sequence, we removed probes with a 1000 Genomes SNP (≥5% MAF) overlapping the last 5bp of the probe closest to and including the CpG site being assayed.[19] In each individual, probes with a detection p>0.01 were removed. Intensity data from the remaining probes were normalized using the *lumi* package,[20] and methylation values per probe (β-values) generated, defined as the ratio of the methylated signal intensity divided by the sum of the methylated and unmethylated signal intensities. Finally, we utilized BMIQ (β mixture quantile method) normalization to correct for the differences related to the Infinium type I and type II assay types on the 450k array.[21] To annotate probes in relation to genes, their positions were overlapped with Refseq gene annotations, including ±2kb of transcription start sites (TSS).

To identify regions that exhibited a parent of origin (PofO) bias in methylation levels, we performed a test in which we compared data from each uniparentally inherited chromosome against all other individuals with biparental inheritance of that chromosome. For an imprinted DMR, we expect the β-values of biparental chromosomes to be intermediate between individuals with maternal and paternal UPD for that chromosome, whose β-values should diverge in opposite directions from the control mean. We employed an outlier test using the Z-score method, with β-values converted to a Z-score using the following equation:

$$z_{ij} = \frac{\beta_{ij} - \mu_j}{\sigma_j}$$

Where $\beta_{ij}$ is the β-value from probe j for a UPD individual i on uniparental chromosome, $\mu_j$ and $\sigma_j$ are the mean and standard deviations of β-values from probe j on biparentally inherited chromosomes. Where we had multiple UPD samples for a single chromosome, we used Stouffer's Z-score method to combine individual Z-scores, as follows:

$$Z_j = \sum_i \frac{w_i z_{ij}}{\sqrt{\sum_i w_i^2}}$$

Where $Z_i$ is a combined Z-score, $z_{ij}$ represents individual probe-level Z-scores as derived from equation 1, with $w_i$ as individual weights. To allow paternal and maternal UPDs

to have opposite directionality, we assigned +1 weight to β-values for paternal UPDs and -1 to β-values from maternal UPDs. The p-value for the combined Z-score was calculated using a normal distribution table. For analysis of the X chromosome, due to the confounder of X chromosome inactivation the control mean was calculated using only male samples and compared with 45,X individuals in which the single X chromosome was of maternal or paternal origin.

In order to robustly identify DMRs influenced by parental origin and avoid false positives caused by single probes, we applied a sliding window analysis to identify regions containing multiple CpGs that were individually significant and with consistent directionality under the outlier test. We estimated optimal thresholds for this analysis using data for chromosomes 7 and 15, which are both well characterized in terms of imprinting. We selected pairs of subjects with UPD for chromosomes 7 or 15 and determined the number of probes contained in putative imprinted DMRs at different combinations of window size, p-value threshold, and minimum number of significant probes per window (Supplemental Figure 2). Based on this analysis, we chose to classify imprinted DMRs as 3kb windows containing ≥2 probes, each with $p<0.001$ showing consistent directionality. To further minimize false positives, we required a β-value difference ≥0.05 between the mean of UPD cases and biparental controls. We determined the fraction of probes passing these thresholds in a comparison of biparentally inherited chromosomes (i.e. control versus control analysis), yielding an estimated false positive rate of ~1% compared to the number of true positives detected in the comparison of uniparentally inherited chromosomes versus controls performed in the same individuals.

***Validation of uniparental methylation and expression***

To unambiguously determine methylation levels of the maternal and paternal alleles in normal control samples with biparentally inherited chromosomes, we identified HapMap individuals who carried heterozygous SNPs within putative DMRs identified in the 1000 Genomes Project. Lymphoblastoid cell lines (Coriell Cell Repositories, Camden, NJ) were

cultured according to supplier's recommendations and cells harvested. 2µg of genomic DNA was extracted (Puregene Core kit B, Qiagen) and bisulfite converted (Epitect Bisulfite kit, Qiagen). Primers were designed to amplify bisulfite converted DNA using Methyl Primer Express v1.0. Bisulfite-treated DNA was amplified using HotStar Taq DNA polymerase (Qiagen). PCR products were gel purified (QIAquick gel extraction kit) and ligated into the TOPO cloning vector using the TOPO TA cloning kit (Life Technologies). Subsequent ligations were transformed into TOP10 chemically competent cells (Life Technologies) and grown on agar plates supplemented with 100mg/ml ampicillin. Individual colonies containing an insert were grown overnight in 5ml LB medium (100mg/ml ampicillin) and DNA extracted with the QIAprep spin kit. Cloned inserts were sequenced using the M13R primer and parental alleles were separated based on the presence of an informative heterozygous SNP.

Quantitative methylation analysis was performed using MassARRAY EpiTYPER assays (Sequenom Inc., CA, USA). Primers incorporating a T7 promoter tag were designed using Epidesigner and used to amplify bisulfite-converted DNA from 20 biparental controls and samples with UPDmat and/or UPDpat. Following *in vitro* transcription, the synthesized RNAs were cleaved with RNase A and the products quantified as per the manufacturer's instructions.

Allele-specific expression studies were performed using HapMap cell lines carrying transcribed polymorphisms in the gene of interest. RNA was extracted from cell pellets using TRIzol (Life technologies, Carlsbad, USA). Following DNAse treatment, cDNA was generated by reverse transcription using Superscript III (Life Technologies, Carlsbad, USA), PCR performed using HotStar Taq DNA polymerase (Qiagen), and amplified products were Sanger sequenced. A complete list of all PCR primers used in these validation assays is shown in Supplemental Table 1.

***Study of allelic expression bias using population-scale RNAseq data***

We utilized results from a recently developed methodology that searches for consistent allelic bias in RNAseq data produced by the GTEx and GEUvadis projects, representing data

from 33 different tissues and cell types. The complete description of the how the test statistic for imprinting in generated is given in Baran et al. [4]. Briefly, the method takes as input genotypes and counts of RNAseq alleles at each SNP in each individual. A series of filtering steps are applied to remove technical and functional confounders such as genotyping error and genes with random monoallelic expression. Finally each gene is evaluated for consistency with monoallelic expression in a likelihood framework learned from the data that considers three categories of gene expression: biallelic, biased, and monoallelic. The biased category contains genes with eQTLs that induce non-biallelic expression. Because the imprinted allele is not random with respect to parent of origin, the likelihood also uses a "flip test" to ensure that the expressed allele is randomly drawn as opposed to the nonsense mediated decay case in which the same allele will induce monoallelic expression.

***Population survey of methylation levels at imprinted loci***

We downloaded data from six studies that used the 450k array to profile DNA methylation patterns in peripheral blood of individuals sampled from the general population (GSE40279, GSE42861, GSE53045, GSE55763, GSE56105 and GSE36064). These studies included both pediatric and elderly populations, cigarette smokers, individuals of European and African American ancestry, and 354 individuals with a diagnosis of rheumatoid arthritis. After removal of outlier samples identified by PCA, we retained a total of 4,004 individuals. Probes with detection p>0.01 were removed, the remaining autosomal probes normalized using *lumi* and BMIQ as detailed above to minimize study-specific biases, and data for 1,022 probes overlapping sites with parentally-biased methylation marks defined in this study and by Court et al.[3] were extracted. To identify individuals with global changes at imprinted loci, we calculated the mean deviation of these 1,022 parentally-biased probes in each individual from the population average. In order to focus on loci with methylation specifically on one parental allele and control for the β-value distribution, we repeated this analysis using a subset of 579 imprinted probes with population mean β-values between 0.4-0.6, calculating

the mean deviation of each individual against all autosomal probes with mean β-values between 0.4-0.6.

## Results

### *Identification of differentially methylated regions in individuals with uniparental disomy*

To identify DMRs where the epigenetic state varied dependent on parental origin, we utilized a sliding window approach to search for sites containing multiple probes showing reciprocal differences in DNA methylation levels in UPD cases compared to biparental controls. A total of 844 autosomal probes mapping to 77 non-redundant regions passed our stringency criteria (2 probes with $p<0.001$ in the same direction within a 3kb window) (Figure 1, Supplemental Table 2). While 43 of these loci were already known to have PofO-specific methylation, 34 loci had not been previously reported. The magnitude of the parental methylation bias we detected at some loci was subtle, likely indicating why these were not detected in previous studies. Consistent with the known excess of maternally methylated loci, 60 loci showed a maternal methylation bias and 17 a paternal methylation bias. As expected, many of these loci showing PofO-specific methylation marks were located at or adjacent to genes with imprinted expression.[4] Comparison with a recent study of DNA methylation in four cases of genome-wide UPD[3] showed that, when considering the 18 chromosomes assayed in our UPD cases, we detected 37 of 42 (88%) loci reported in this study. Inspection of the five loci that we did not classify as imprinted revealed that although each contained probes with significant p-values, these were either single probes or occurred >3kb from the nearest neighboring significant probe, and thus did not meet our discovery thresholds. For example, we identified single probes located intragenic within *MEG3* [MIM: 605636] and *MEG8* [MIM: 613648] at 14q32.2 that yielded p-values for parental methylation bias of $3.7x10^{-93}$ and $3.9x10^{-244}$, but in both cases there were no other significant probes located within close proximity.

In addition to searching for autosomal regions showing parental epigenetic bias, we also compared DNA methylation profiles in four individuals with Turner syndrome (45,X) where the single X is of either maternal or paternal origin. However, even when using

relaxed statistical thresholds, consistent with a previous report[22] we were unable to detect any evidence of imprinting on the human X chromosome.

***Consistent parental bias in DNA methylation extending over 1.9Mb of 15q11-q13***

Previous studies of imprinting in the PWS/AS region have concluded that parental-specific methylation marks occur specifically at CpG-rich regions associated with the *MKRN3* [MIM: 603856]*, SNURF*/*SNRPN* [MIM: 182279], *NDN* [MIM: 602117] and *MAGEL2* [MIM: 605283] genes.[3,15] However, in addition to those DMRs detected using our statistical pipeline, visual analysis of data from four individuals with UPD(15)mat and four individuals with UPD(15)pat uncovered subtle global biases of parental methylation levels extending across the entire 15q11-q13 region. Visualization of the difference in methylation levels between individuals with UPD(15) against the mean of biparental controls uncovered a clear reciprocal pattern in which methylation level consistently varied in relation to parental origin at every CpG assayed over ~1.9Mb of the PWS/AS region, starting ~180kb proximal to *MKRN3* and extending distally to include the *SNORD115* [MIM: 609837] and *SNORD116* [MIM: 605436] clusters of imprinted small nucleolar RNAs (Figure 2). This extended region of parental epigenetic bias included CpGs with preferential methylation of both the maternal and paternal alleles, with the direction of the parental bias switching multiple times over the contiguous region. In addition to the previously known extent of imprinting in the PWS/AS region, our statistical analysis identified a maternally-methylated locus in the CpG island shore at the promoter of *TUBGCP5* (Tubulin-gamma complex-associated protein 5, [MIM: 608147]), a gene which lies at the centromeric end of the recurrent PWS/AS deletion region between Break Point 1 (BP1) and BP2. In this position, copy number of *TUBGCP5* is altered by both larger Type 1 (BP1-BP3) PWS/AS deletions, and also by recurrent deletions and duplications of 15q11.2. These latter rearrangements are risk factors for a variety of neuropsychiatric and neurodevelopmental problems with incomplete penetrance,[23,24] although associated PofO effects have not been reported to date. Visual inspection also suggested the presence of a possible region of increased maternal methylation associated

with *MIR4509* located close to BP1 in 15q13.1, although this locus did not reach statistical significance in our formal analysis.

### *Identification of differentially methylated regions in tissues with maternally- or paternally-derived genomes*

To complement our analysis of UPD cases, we also performed methylation profiling in four samples of OT and four samples of CHM, which have a full chromosome complement exclusively of maternal or paternal origin, respectively. As a result of their divergent developmental origins, there are multiple epigenetic differences between these two tissue types that are unrelated to parental origin, complicating their use for the identification of imprinted loci. However, as the majority of imprinted DMRs are characterized by intermediate levels of methylation in biparental tissues, to increase specificity for identifying putative imprinted regions, we focused on loci that showed both highly divergent methylation in OT versus CHM (mean β-values <0.25 and >0.75, or vice versa) and intermediate methylation levels in peripheral blood of biparental controls (mean β-value between 0.4 and 0.6). This analysis identified a total of 580 probes (Supplemental Table 3), including many previously described sites of PofO epigenetic bias (Supplemental Figure 3). 179 (31%) of these probes were identified as showing significant PofO methylation bias in our analysis of UPD samples, providing immediate cross-validation of parentally-biased methylation at these sites, including four loci not reported in previous studies (*ZBTB8B*, *BRDT* [MIM: 602144], *CHD7* [MIM: 608892] and *MIR8069-1/MIR8069-2*) (Supplemental Figure 4). Utilizing a more relaxed stringency (OT/CHM showing mean reciprocal β-values <0.3 and >0.7), 236 of the 844 (28%) probes identified showing significant PofO-biased methylation in UPD cases were cross-validated, including two further putative imprinted DMRs: *LOC151121* and *ANKRD20A11P*. In every case, the directionality of the PofO methylation bias was consistent between UPD cases and OT/CHM.

### *Validation of parental methylation and expression bias*

To validate putative PofO methylation biases detected by the 450k array, we performed quantitative methylation profiling of DMRs associated with *RPS2P32* and *SVOPL* [MIM: 611700] using Sequenom EpiTYPER assays, including additional control and UPD samples not tested in our discovery phase. Consistent with published studies of *SVOPL*,[3] these assays confirmed preferential methylation of the maternal allele at both loci (Figure 3, Supplemental Figure 5). Utilizing bisulfite cloning of informative HapMap trios, we also confirmed preferential maternal methylation at the putative DMRs associated with *ERLIN2* [MIM: 611605], *HTR2A* [MIM: 182135], *WDR27, WRB* [MIM: 602915], *JAKMIP1* [MIM: 611195] and *LOC101927815* detected in our initial genome-wide screen (Supplemental Figure 6). Sanger sequencing of transcribed SNPs using informative HapMap trios showed consistent paternal expression bias for *PPIEL* and *LOC101927815*, and maternal expression bias for *ERLIN2* (Supplemental Figure 7). In addition, analysis of RNAseq data from a three-generation pedigree[25] further confirmed a strong paternal expression bias of *PPIEL* at 1p34.3, and weaker but consistent paternal expression biases for both *DESI1* [MIM: 614637] and *SNU13* [MIM: 601304] at 22q13.2 (Supplemental Figure 8). We also observed data suggesting a consistent ~2-fold over-expression of the maternal allele of *HM13* [MIM: 607106], although we were unable to exclude that this may be due to preferential mapping bias to the reference allele (Supplemental Table 4).

***Allelic expression bias of DMR-associated genes using population-scale RNAseq data***

We searched for evidence that would support putative imprinted gene expression associated with sites of PofO methylation bias identified using UPD samples. We utilized results from a recently developed methodology that searches for consistent allelic bias in RNAseq data (independent of genotype) from large populations as a potential signature of imprinted expression, applied to RNAseq data produced by the GTEx and GEUvadis projects, representing data from 33 different tissues and cell types.[4] Considering all transcripts located within ±10kb of each of our PofO DMRs, we first excluded all genes with prior evidence of imprinted expression, leaving a total of 30 independent transcripts that

contained informative SNPs with sufficient expression level in at least one tissue to be effectively assayed for evidence of allelic expression bias. Despite the fact that none of these transcripts showed strong evidence of imprinting in previous studies, in comparison to all other genes located >10kb from our PofO DMRs, we observed a significant excess of allelic bias around our previously undefined DMRs in both the GTEx ($p=1.2x10^{-5}$, Mann-Whitney Rank Sum test) and GEUvadis (p=0.013) data sets. Two genes in the GEUvadis population (*HM13* and *RPS2P32*) and seven genes in one or more GTEx tissues (*PSCA* [MIM: 602470], *SORD* [MIM: 182500], *LIPI* [MIM: 609252], *IGF2R* [MIM: 147280], *ANKRD20A11P*, *RPS2P32* and *SMOC2* [MIM: 607223]) each yielded test statistics for imprinting >20, corresponding to the top 1.02% of all genes tested (Supplemental Tables 5 and 6). Example plots showing consistent allelic bias for *LIPI*, *RPS2P32P* and *ANKRD20A11* in the GTEx data are shown in Supplemental Figure 9.

***Population variation of methylation levels at imprinted loci***

We investigated patterns of epigenetic variability at imprinted DMRs in 4,004 blood DNA methylomes profiled using the 450k array, representing a deep sample of methylation levels from the general population. This revealed considerable epigenetic variation at some imprinted loci. For example, at the *VTRNA2-1* [MIM: 614938] DMR, 22% of individuals showed hypomethylation of both parental alleles (mean β-value of DMR probes <0.1), indicating that the imprinting mark at this locus is polymorphic, as has been observed previously[26,27] (Supplemental Table 7, Supplemental Figure 10). Similarly, five individuals (~0.1% of the population) showed hypomethylation of the *FAM50B* [MIM: 614686] DMR at 6p25.2, and six individuals showed hypermethylation of both alleles (mean β-value of DMR probes >0.9) at the *HM13*/*MCTS2P* [MIM: 607106] DMR at 20q11.21, indicating a low but appreciable frequency of altered parental-specific methylation at these loci. However, most imprinted DMRs exhibited intermediate methylation levels in all individuals, indicating that parental methylation at these loci is tightly constrained in the population.

We also identified rare individuals who exhibited consistently raised or lowered methylation levels at dozens of imprinted genes across the genome, suggesting a generalized disturbance of imprinting (Figure 4a, Supplemental Figure 10). Considering all probes on the 450k array that show a PofO bias in UPD samples, we identified seven individuals who showed global hypomethylation at PofO DMRs (mean β-value of imprinted probes <0.05 below the population mean), and one individual who showed a mean β-value at all PofO DMRs that was >0.05 above the population mean. Focusing on the subset of PofO biased CpGs with intermediate β-values (population mean 0.4-0.6) as an approximation for loci where methylation occurs specifically on one allele, the most extreme individual identified showed a mean β-value at PofO biased DMRs that was 0.12 below the population average ($p=1.8x10^{-20}$), and could be clearly visualized as showing consistent hypomethylation across dozens of DMRs (Figure 4, Supplemental Figure 10). Altered methylation levels occurred at both maternally and paternally methylated regions (Pearson correlation of mean differences for probes in maternally and paternally methylated loci, r=0.83), and were specific to regions of PofO methylation bias, as matched probes outside of imprinted regions in these same individuals showed mean β-value differences <0.01 versus the population mean (Figure 4c).

## Discussion

Using quantitative methylation profiling in a cohort of 57 individuals carrying chromosomes inherited from a single parent, we identify 77 autosomal regions showing a PofO methylation bias. Of these, 34 have not been previously reported, significantly expanding the number of loci showing PofO epigenetic bias identified to date in somatic tissues. However, despite phenotypic evidence from the analysis of Turner syndrome cases,[28] in contrast to the mouse we could find no evidence of imprinting on the human X chromosome.[29,30]

Of particular note we detected subtle but consistent PofO biases at all CpGs tested within the PWS/AS imprinted domain, extending apparently continuously over 1.9Mb of 15q11-q13. Previous studies have suggested imprinted epigenetic marks within this region are confined to specific regulatory elements, such as CpG islands at the imprinted transcripts *SNRPN* [MIM: 182279] and *MAGEL2* [MIM: 605283].[3,15] However, the quantitative nature of our assay reveals consistent epigenetic bias dependent on parental origin across all CpGs in this region. We examined other imprinted loci for this same phenomenon, but did not observe any similar patterns of parental bias in methylation over 100s of kb at any other genomic regions, suggesting this may be unique to the PWS/AS region. However, for many other chromosomes both the sparse 450k array coverage and smaller number of UPD cases available means that our power to detect such phenomena was limited.

Our observations make two important points that challenge current paradigms of imprinted loci. First, parentally biased epigenetic marks are not always confined to discrete regulatory elements such as CpG islands or gene promoters, but as illustrated by our observations in the PWS/AS region can extend over entire chromosomal domains. Second, not all differentially methylated sites associated with imprinted genes show a hemi-methylated pattern, with one parental allele that is fully methylated and the other allele largely unmethylated. Indeed, several of the loci we identify as showing PofO bias in methylation levels exhibit quite subtle differences between the maternal and paternal

genomes, with the two parental alleles differing from the biparental mean by only ~10%. For example, the 86 probes contained within DMRs in the *SNORD115/116* [MIM: 609837 and 605436] cluster have a mean β-value of 0.83 in biparental controls, indicating they are predominantly methylated on both parental alleles, but show a slight tendency for preferential paternal methylation, with mean β-values of 0.75 on the maternal allele and 0.91 on the paternal allele. These subtle PofO biases likely explain why many signals of imprinting we detect have not been identified in prior studies, which have tended to focus on sites showing highly divergent PofO methylation. Thus, our studies indicate that the contribution of the two parental epigenomes is much more subtle and complex than previously thought.

There are several potential mechanisms that might contribute to these subtle parental epigenetic biases that we observed. While subtle methylation differences may be an inherent and previously unrecognized feature of the two parental alleles, one possible alternative explanation is that rather than being a primary mark of imprinting, it may be a secondary consequence of imprinted gene expression that results in differences in methylation levels between the parental alleles within transcribed regions. Alternatively it may be that the PofO epigenetic mark occurs strongly only in certain cell types or developmental time points, and by studying DNA derived from post-natal blood we are only able to detect either a weak residual epigenetic mark, or one that is diluted by mosaicism.

We suggest that future transcription studies may reveal that similar subtle PofO biases also occur in gene expression patterns, although by nature these are difficult to reliably detect. Indeed, the study of F1 hybrids in model organisms, a powerful system for the study of PofO effects, has identified several examples of genes showing partial imprinting in the placenta (e.g. *Phactr2* [MIM: 608724] in mouse and *IGF2R* [MIM: 147280] in equids).[12,13]

We also identified PofO biased DMRs on both chromosome 21 and within the recurrent 15q11.2 deletion/duplication region [MIM: 615656], suggesting the potential for PofO effects associated with aneuploidy of these regions. For example, there is some

evidence suggesting that the parental origin of the additional chromosome 21 in Down syndrome [MIM: 190685] influences the penetrance of heart defects in this disorder.[31] We suggest that detailed phenotype studies of trisomy 21 and carriers of the 15q11.2 deletion/duplication may reveal differences between individuals with maternally and paternally derived aneuploidy of these regions.

We note that a number of other genes associated with DMRs we identify have previously been linked with complex diseases. For example, polymorphisms in the serotonin receptor *HTR5A* have been linked to risk of schizophrenia,[32,33] while haploinsufficiency for *JRK* [MIM: 603210] is associated with risk of epilepsy,[34] a condition in which a strong maternal transmission bias has been observed in numerous studies.[35-38]

Multiple different lines of evidence suggest that many of the DMRs we detect represent genuine regions of parental bias: (i) Several occur either within or adjacent to the known 15q11-q13 imprinted region, including a cluster of DMRs associated with the imprinted *SNORD* cluster distal to *SNRPN*, and a DMR associated with the promoter of *TUBGCP5* in 15q11.2. (ii) Six of the novel DMRs we identified (*ZBTB8B, BRDT, CHD7, MIR8069-1/MIR8069-2, LOC151121 and ANKRD20A11P*) also showed the same PofO methylation bias in other uniparental tissues. (iii) DMRs associated with *SORD* [MIM: 182500] and *JAKMIP1* were highlighted as possible imprinted regions during initial analysis in previous studies of imprinting, but were later excluded as they failed subsequent filtering thresholds.[3,39] Furthermore *SORD* was also identified as showing a maternal methylation bias in placenta.[40] (iv) In common with several well-established imprinted loci, the DMR we identify at *ZBTB8B* overlaps one of 63 syntenic ZFP57 binding sites identified in mouse ES cells, a protein involved in the maintenance of imprinting.[41] (v) Validation experiments using targeted methylation assays confirmed significant parentally biased methylation levels for six loci. (vi) Analyses of two populations with RNAseq data identified significant enrichments for signals of imprinted expression, with eight genes associated with novel DMRs yielding scores within the top 1% of all transcripts tested. While the relaxed threshold and limited number of informative individuals contributing to some of these expression signals mean that

it would be premature to conclude that all these genes are definitively imprinted, these data provide evidence that supports imprinted expression associated with many of the novel DMRs we identified. PCR and RNAseq analysis of family groups confirmed paternally biased expression of five other transcripts associated with PofO DMRs we describe. However, linking the presence of parental methylation marks to imprinted gene expression can be challenging. For example, while *IGF1R* [MIM: 147370] contains a maternally methylated DMR located within intron 2, there is currently no evidence that *IGF1R* itself shows imprinted expression, or any evidence for PofO effects associated with *IGF1R* mutations. Instead, recent studies have revealed imprinted expression of a lncRNA, *IRAIN*, antisense within *IGF1R*.[42] Therefore considerable caution should be applied when inferring the imprinting status of a specific gene from the presence of a *cis*-linked parental methylation bias.

Although we excluded probes on the 450k array with multiple alignments (see Methods), we utilized relatively permissive thresholds that retained probes with weak homology in the genome, meaning that some of the loci with putative parental epigenetic bias that we report have multiple paralogs. For example, BLAST analysis indicates the *RPS2P32P* DMR has 38 significant matches in the human genome with a mean of 85% sequence identity. While it is possible that cross hybridization artifacts and underlying copy number variations of the target regions might influence methylation profiles reported for this locus, the overall weight of evidence strongly indicates a genuine parental epigenetic bias at this locus. The *RPS2P32P* DMR was initially discovered by array profiling of four UPD samples, with significant and consistent signal seen for six independent contiguous probes. We then replicated this using a PCR assay in enlarged set of eight UPD samples, all of which showed the same maternal methylation bias as observed initially. In addition, the transcript for *RPS2P32P* yielded nominally significant evidence of imprinted expression by RNAseq in two independent cohorts comprising >600 individuals. The same DMR also showed methylation shifts by array in two individuals identified from the general population who presented with consistent hypo- or hypermethylation at dozens of imprinted loci. Finally, analysis of 4,004 controls showed that the methylation profiles for the *RPS2P32P* DMR

show very little variance in the normal population, strongly arguing that despite having multiple paralogs, probes targeting this locus report stable data. Therefore, the convergence of all these data supporting the conclusion of imprinting of *RPS2P32P* is robust, despite its paralogous nature.

Our study has several limitations. The 450k array platform we used to perform DNA methylation profiling only provides targeted coverage of the genome, assaying a small fraction of total CpGs, and as a result will miss some sites with parental epigenetic bias. For example, one previously reported DMR within 16q24.1[43] is not represented on the 450k array, and our use of thresholds requiring multiple CpGs showing consistent parental bias meant we failed to identify a small number of known DMRs where probe coverage was poor. Second, our sample collection did not include samples with constitutive UPD(10), UPD(11), UPD(18) or UPD(19), which occur infrequently.[7] Finally, as all the DNA samples from UPD cases were derived from peripheral blood, this limited our ability to detect parental methylation bias that might be confined to specific tissues.

While loss of imprinting (LOI) at several different imprinted regions is a known cause of syndromic congenital disorders,[6] the functional consequences of disruptions at other imprinted loci remain unknown. Our population analysis of imprinted DMRs indicates that many show tightly constrained methylation levels in the general population. In contrast, at several other DMRs including *HM13*/*MCTS2P*, *FAM50B* and *SNU13*, we observed a small fraction of individuals displaying consistent hypo- or hypermethylation. While we do not have detailed phenotype data for these individuals, these observations suggest that LOI of these loci in blood cells may be compatible with normal development. Surprisingly, within our population analysis we observed methylation profiles indicating occasional LOI at DMRs classically associated with congenital disorders, including hypomethylation of the *KCNQ1* [MIM: 607542] DMR in one individual, a defect that is normally associated with BWS.[44] While we do not know the clinical history of these individuals, it is unlikely that individuals suffering from rare congenital imprinting disorders would occur in a population sample of 4,004 individuals. This may therefore represent somatic mosaicism for LOI, the presence of a

deletion including the imprinted DMR, or alternatively suggest that LOI at these loci can occur with incomplete penetrance or variable expressivity. We also identified rare individuals who exhibited consistently raised or lowered methylation levels across dozens of imprinted loci. While we are unable to perform further studies of these cases, mutations in genes such as *NLRP2* [MIM: 609364], *NLRP5* [MIM: 609658], *NLRP7* [MIM: 609661], *ZFP57* [MIM: 612192] and *TRIM28* [MIM: 601742] have all been associated with disturbed methylation at multiple imprinted loci.[45-48] We suggest that detailed investigation of individuals with unusual methylation patterns at multiple imprinted loci may provide further insights into the establishment and maintenance of imprinting.

## Data Access

The 450k array data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE64244.


## Acknowledgements

This work was supported by NIH grants DA033660, HG006696, HD073731 and MH097018, and research grant 6-FY13-92 from the March of Dimes to AJS, a Beatriu de Pinos Postdoctoral Fellowship to RS (2011BP-A00515), and the Bundesministerium für für Bildung und Forschung to KB (BMBF; Imprinting diseases, grant No. 01GM1513A).


## Disclosures

The authors declare no conflicts of interest.

## Web Resources

The URLs for data presented herein are as follows:

http://igc.otago.ac.nz/

http://www.1000genomes.org

http://www.epidesigner.com/start3.html

http://www.ncbi.nlm.nih.gov/geo/

http://omim.org/

## References


1. Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. 2, 21-32.

2. Morison, I.M., Paton, C.J., Cleverley, S.D. (2001). The imprinted gene and parent-of-origin effect database. Nucleic Acids Res. 29, 275-6.

3. Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J.V., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. 24, 554-69.

4. Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E.K., Rivas, M.A., Pirinen, M., Gutierrez-Arcelus, M., Smith, K.S., Kukurba, K.R., et al. (2015). The landscape of genomic imprinting across diverse adult human tissues. Genome Res. 25, 927-36.

5. Robinson, W.P. (2000). Mechanisms leading to uniparental disomy and their clinical consequences. Bioessays 22, 452-459.

6. Eggermann, T., Perez de Nanclares, G., Maher, E.R., Temple, I.K., Tümer, Z., Monk, D., Mackay, D.J., Grønskov, K., Riccio, A., Linglart, A., Netchine, I. (2015). Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. Clin. Epigenetics 14, 123.

7. Kotzot, D., Utermann, G. (2005). Uniparental disomy (UPD) other than 15: phenotypes and bibliography updated. Am. J. Med. Genet. A. 136, 287-305.

8. Pollard, K.S., Serre, D., Wang, X., Tao, H., Grundberg, E., Hudson, T.J., Clark, A.G., and Frazer, K. (2008). A genome-wide approach to identifying novel-imprinted genes. Hum. Genet. 122, 625-634.

9. Morcos, L., Ge, B., Koka, V., Lam, K.C., Pokholok, D.K., Gunderson, K.L., Montpetit, A., Verlaan, D.J., Pastinen, T. (2011). Genome-wide assessment of imprinted expression in human cells. Genome Biol. 12, R25.

10. Barbaux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.B., Pinard, A., Auer, J., Bessieres, B., Barlier, A., et al. (2012). A genome-wide approach reveals novel imprinted genes expressed in the human placenta. Epigenetics 7, 1079-1090.

11. Metsalu, T., Viltrop, T., Tiirats, A., Rajashekar, B., Reimann, E., Kõks, S., Rull, K., Milani, L., Acharya, G., Basnet, P., et al. (2014). Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. Epigenetics 9, 1397-409.

12. Wang, X., Soloway, P.D., Clark, A.G. (2011). A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics 189, 109-22.

13. Wang, X., Miller, D.C., Harman, R., Antczak, D.F., Clark, A.G. (2013). Paternally expressed genes predominate in the placenta. Proc Natl Acad Sci U S A 110, 10705-10710.

14. Dockery, L., Gerfen, J., Harview, C., Rahn-Lee, C., Horton, R., Park, Y., Davis, T.L. (2009). Differential methylation persists at the mouse *Rasgrf1* DMR in tissues displaying monoallelic and biallelic expression. Epigenetics 4, 241-247.

15. Sharp, A.J., Migliavacca, E., Dupre, Y., Stathaki, E., Sailani, M.R., Baumer, A., Schinzel, A., Mackay, D.J., Robinson, D.O., Cobellis, G., et al. (2010). Methylation profiling in individuals with uniparental disomy identifies novel differentially methylated regions on chromosome 15. Genome Res. 20, 1271-1278.

16. Nakabayashi, K., Trujillo, A.M., Tayama, C., Camprubi, C., Yoshida, W., Lapunzina, P., Sanchez, A., Soejima, H., Aburatani, H., Nagae, G., et al. (2011). Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. Hum. Mol. Genet. 20, 3188-97.

17. Choufani, S., Shapiro, J.S., Susiarjo, M., Butcher, D.T., Grafodatskaya, D., Lou, Y., Ferreira, J.C., Pinto, D., Scherer, S.W., Shaffer, L.G., et al. (2011). A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. Genome Res. 21, 465-76.

18. Xi, Y., Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10, 232.

19. Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H., Hvilsom, C., Navarro, A., Esteller, M., Sharp, A.J., Marques-Bonet, T. (2013). Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet. 9, e1003763.

20. Du, P., Kibbe, W.A., Lin, S.M. (2008). *lumi*: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547-8.

21. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. Bioinformatics 29, 189-96.

22. Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., Antonarakis, S.E. (2011) DNA methylation profiles of human active and inactive X chromosomes. Genome Res. 21, 1592-600.

23. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. 43, 838-46.

24. Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G.B., Jonsdottir, G.A., Doyle, O.M., et al. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 505, 361-6.

25. Li, X., Battle, A., Karczewski, K.J., Zappala, Z., Knowles, D.A., Smith, K.S., Kukurba, K.R., Wu, E., Simon, N., Montgomery, S.B. (2014). Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. Am. J. Hum. Genet. 95, 245-56.

26. Romanelli, V., Nakabayashi, K., Vizoso, M., Moran, S., Iglesias-Platas, I., Sugahara, N., Simón, C., Hata, K., Esteller, M., Court, F., and Monk, D. (2014). Variable maternal methylation overlapping the *nc886/vtRNA2-1* locus is locked between hypermethylated repeats and is frequently altered in cancer. Epigenetics. 9, 783-90.

27. Silver, M.J., Kessler, N.J., Hennig, B.J., Dominguez-Salas, P., Laritsky, E., Baker, M.S., Coarfa, C., Hernandez-Vargas, H., Castelino, J.M., Routledge, M.N., Gong, Y.Y., Herceg, Z., Lee, Y.S., Lee, K., Moore, S.E., Fulford, A.J., Prentice, A.M., and Waterland, R.A. (2015). Independent genomewide screens identify the tumor suppressor *VTRNA2-1* as a human epiallele responsive to periconceptional environment. Genome Biol. 16, 118.

28. Skuse, D.H., James, R.S., Bishop, D.V., Coppin, B., Dalton, P., Aamodt-Leeper, G., Bacarese-Hamilton, M., Creswell, C., McGurk, R., Jacobs, P.A. (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. Nature 387, 705-8.

29. Davies, W., Isles, A., Smith, R., Karunadasa, D., Burrmann, D., Humby, T., Ojarikre, O., Biggin, C., Skuse, D., Burgoyne, P., Wilkinson, L. (2005). *Xlr3b* is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nat. Genet. 37, 625-9.

30. Raefski, A.S., O'Neill, M.J. (2005). Identification of a cluster of X-linked imprinted genes in mice. Nat. Genet. 37, 620-4.

31. Muranjan, M., Chaudhari, T., Vundinti, B.R. (2010). Phenotypic heterogeneity and parental origin of extra chromosome 21 in Down syndrome. Indian Pediatr. 47, 429-32.

32. Iwata, N., Ozaki, N., Inada, T., and Goldman, D. (2001). Association of a *5-HT(5A)* receptor polymorphism, Pro15Ser, to schizophrenia. Mol. Psychiatry 6, 217-219.

33. Dubertret, C., Hanoun, N., Adès, J., Hamon, M., Gorwood, P. (2004). Family-based association studies between *5-HT5A* receptor gene and schizophrenia. J. Psychiatr. Res. 38, 371-6.

34. Donovan, G.P., Harden, C., Gal, J., Ho, L., Sibille, E., Trifiletti, R., Gudas, L.J., and Toth, M. (1997). Sensitivity to *jerky* gene dosage underlies epileptic seizures in mice. J. Neurosci.17, 4562-4569.

35. Annegers, J.F., Hauser, W.A., Elveback, L.R., Anderson, V.E., Kurland, L.T. (1976). Seizure disorders in offspring of parents with a history of seizures - a maternal-paternal difference? Epilepsia 17, 1-9.

36. Ottman, R., Annegers, J.F., Hauser, W.A., Kurland, L.T. (1988). Higher risk of seizures in offspring of mothers than of fathers with epilepsy. Am. J. Hum. Genet. 43, 257-64.

37. Doose, H., Neubauer, B.A. (2001). Preponderance of female sex in the transmission of seizure liability in idiopathic generalized epilepsy. Epilepsy Res. 43, 103-14.

38. Pal, D.K., Durner, M., Klotz, I., Dicker, E., Shinnar, S., Resor, S., Cohen, J., Harden, C., Moshé, S.L., Ballaban-Gill, K., et al. (2006). Complex inheritance and parent-of-origin effect in juvenile myoclonic epilepsy. Brain Dev. 28, 92-8.

39. Docherty, L.E., Rezwan, F.I., Poole, R.L., Jagoe, H., Lake, H., Lockett, G.A., Arshad, H., Wilson, D.I., Holloway, J.W., et al. (2014). Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J. Med. Genet. 51, 229-38. Erratum in: J Med Genet 2014 51, 478.

40. Yuen, R.K., Jiang, R., Peñaherrera, M.S., McFadden, D.E., Robinson, W.P. (2011). Genome-wide mapping of imprinted differentially methylated regions by DNA methylation profiling of human placentas from triploidies. Epigenetics Chromatin 4, 10

41. Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 44, 361-72.

42. Sun, J., Li, W., Sun, Y., Yu, D., Wen, X., Wang, H., Cui, J., Wang, G., Hoffman, A.R., Hu, J.F. (2014). A novel antisense long noncoding RNA within the *IGF1R* gene locus is imprinted in hematopoietic malignancies. Nucleic Acids Res. 42, 9588-601.

43. Szafranski, P., Dharmadhikari, A.V., Brosens, E., Gurha, P., Kolodziejska, K.E., Zhishuo, O., Dittwald, P., Majewski, T., Mohan, K.N., Chen, B., et al. (2013). Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder. Genome Res. 23,23-33.

44. Smilinich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., et al. (1999). A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc. Natl. Acad. Sci. U.S.A. 96, 8064-9.

45. Mackay, D.J., Callaway, J.L., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H.V., Goodship, J.A., Haemers, A.P., et al. (2008). Hypo-methylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in *ZFP57*. Nat. Genet. 40, 949-51.

46. Meyer, E., Lim, D., Pasha, S., Tee, L.J., Rahman, F., Yates, J.R., Woods, C.G., Reik, W., Maher, E.R. (2009). Germline mutation in *NLRP2* (*NALP2*) in a familial imprinting disorder (Beckwith-Wiedemann Syndrome). PLoS Genet. 5, e1000423.

47. Alexander, K.A., Wang, X., Shibata, M., Clark, A.G., García-García, M.J. (2015). TRIM28 controls genomic imprinting through distinct mechanisms during and after early genome-wide reprogramming. Cell Rep. 13, 1194-205.

48. Docherty, L.E., Rezwan, F.I., Poole, R.L., Turner, C.L., Kivuva, E., Maher, E.R., Smithson, S.F., Hamilton-Shield, J.P., Patalan, M., Gizewska, M., et al. (2015). Mutations in *NLRP5* are associated with reproductive wastage and multilocus imprinting disorders in humans. Nat Commun. 6, 8086.

49. Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S., and Nicholls, R.D. (1999). Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. Am. J. Hum. Genet. 65, 370-386.

## Legends to Figures and Tables

**Figure 1. Karyogram showing the distribution of imprinted differentially methylated regions identified in individuals with uniparental disomy.** Each imprinted DMR is represented according to RefSeq gene annotations: DMRs overlapping a gene or within ±2kb of a TSS are shown as single gene names, while intergenic DMRs are indicated by black arrows joining names of the two flanking genes. Paternally methylated loci are shown in blue on the right, while maternally methylated loci are shown in red on the left of each chromosome. Novel imprinted DMRs (those not reported in previous studies) are boxed. In total 77 DMRs were identified, 60 of which showed a maternal methylation bias and 17 a paternal methylation bias. Chromosomes 10, 11, 18 and 19 are shown in grey as these were not tested for imprinted methylation marks in UPD samples, and thus we do not show known imprinted DMRs associated with *INPP5Fv2* (chr10), *IGF2* [MIM: 147470] and *KCNQ1* (chr11), and *ZNF331* [MIM: 606043] and *PEG3* [MIM: 601483] (chr19). Although not shown, analysis of the chromosome X in four cases of Turner syndrome failed to detect any evidence of loci showing PofO methylation bias.

**Figure 2. Reciprocal methylation differences based on parental origin extend over >2Mb of 15q11-q13, including regions that undergo recurrent rearrangement in human disease.** Comparison of relative methylation levels in eight individuals with UPD(15)mat or UPD(15)pat versus biparental controls reveals a clear reciprocal pattern in which methylation varies in relation to parental origin across hundreds of CpGs. Although in many cases these differences are quite subtle, presenting with differences between UPD cases and biparental controls of <10%, this analysis shows a consistent PofO epigenetic bias at every CpG tested over an ~1.85Mb interval within the minimal PWS/AS recurrent deletion region (chr15:23,650,000-25,500,000). In addition two DMRs located more proximally associated with *MIR4509* and *TUBGCP5* are clearly visible. Both are located within the 15q11.2 region that undergoes frequent recurrent deletion/duplication in individuals with intellectual disability

and schizophrenia, and which has also been associated with reduced cognition in controls,[23,24] suggesting the existence of possible PofO effects associated with this genomic disorder. In the ideogram of chromosome 15 at the top of the figure, the zoomed ~3Mb region (chr15:22,500,000-25,650,000) shown in the plot is highlighted as a red square. Below the plot, black bars indicate the positions of the recurrent breakpoints (BP1, BP2 and BP3) and deletion regions observed in the 15q11.2 and Prader-Willi/Angelman syndromes.[49] Also shown are the positions of RefSeq genes, segmental duplications and DMRs tested with the 450k array that showed a significant PofO association ($p<0.001$).

**Figure 3. Maternally biased methylation at the *RPS2P32* locus.** (A) Discovery of a maternally methylated DMR at chromosome 7p15.3 overlapping *RPS2P32* by methylation profiling using the 450k array. Six CpGs assayed by the 450k array in two UPD(7)mat and two UPD(7)pat cases show significant reciprocal methylation differences when compared to biparental controls (p-values between $8.4x10^{-7}$ and $3.6x10^{-13}$). (B) Results of targeted methylation analysis of the *RPS2P32* locus using a Sequenom EpiTYPER assay. Four additional individuals with UPD(7)mat who were not tested by the 450k array were profiled using this PCR-based assay, with all four showing much higher methylation levels than controls at most CpGs assayed. Methylation levels for cases of UPD(7)mat, UPD(7)pat and biparental controls are shown as red, blue and black filled circles, respectively, with the dotted black line connecting the mean methylation in controls for consecutive CpGs assayed. In (B), methylation values in one CHM (paternally-derived) sample are shown as red crosses, and blue crosses show data from one OT (maternally derived) sample. The differing shades of grey represent 1, 1.5 and 2 standard deviations from the mean of control methylation values. The location of *RPS2P32* (black bar), individual CpGs (black tick marks) and the location of a CpG island (green bar) are also shown.

**Figure 4. Rare individuals with global hyper- or hypomethylation of multiple imprinted loci.** Analysis of 4,004 blood-DNA methylomes from the general population identified

individuals showing consistently increased or decreased methylation levels at the majority of imprinted DMRs. (A) Plots show data for four imprinted loci located on different chromosomes. In each plot, the solid red line shows the profile of individual 1056 that showed the greatest mean reduction in methylation across all imprinted DMRs, while the solid blue line is individual 753 who showed the greatest mean increase in methylation at all imprinted DMRs. Methylation profiles for the other 4,002 individuals are shown by thin dashed lines, with the population mean shown as a black solid line. Similar plots for 35 other loci that show a PofO epigenetic bias are shown in Supplemental Figure 10. (B) Considering those probes within imprinted DMRs with intermediate methylation (population mean β-values between 0.4-0.6), individual 1056 showed 12% lower methylation than the population average ($p=1.8x10^{-20}$, Mann Whitney Rank Sum test), while individual 753 showed 6% higher methylation than the population average ($p=3.9x10^{-12}$). (C) These shifts occurred specifically at imprinted DMRs: considering all autosomal probes on the 450k array with population mean β-value between 0.4-0.6, these two individuals showed mean β-value differences of +0.005 and +0.008 compared to the population mean (non-significant).

## Supplemental Data

Supplemental Data include eight figures and seven tables

**Supplemental Table 1.** PCR primers used in validation assays of imprinted methylation and expression.

**Supplemental Table 2.** Probes within DMRs showing significant PofO methylation bias in UPD samples.

**Supplemental Table 3.** Probes showing large reciprocal differences in methylation levels between Ovarian Teratoma and Complete Hydatidiform Mole samples, and with intermediate methylation levels in blood of biparental controls (mean β-values between 0.4-0.6).

**Supplemental Table 4.** Analysis of allelic read counts using RNAseq data within a three-generation HapMap pedigree[25] shows consistent preferential maternal expression of a transcribed SNP within *HM13/ HM13-AS1* in five individuals from this pedigree.

**Supplemental Table 5.** Evidence for imprinted expression from allelic bias in RNAseq data of genes located within 10kb of DMRs in the GEUvadis cohort (462 lymphoblastoid cell lines).[4] Transcripts with a test statistic >20 (corresponding to the top 1.02% of all tests performed) are highlighted in green, while genes not previously reported to be imprinted are highlighted in red.

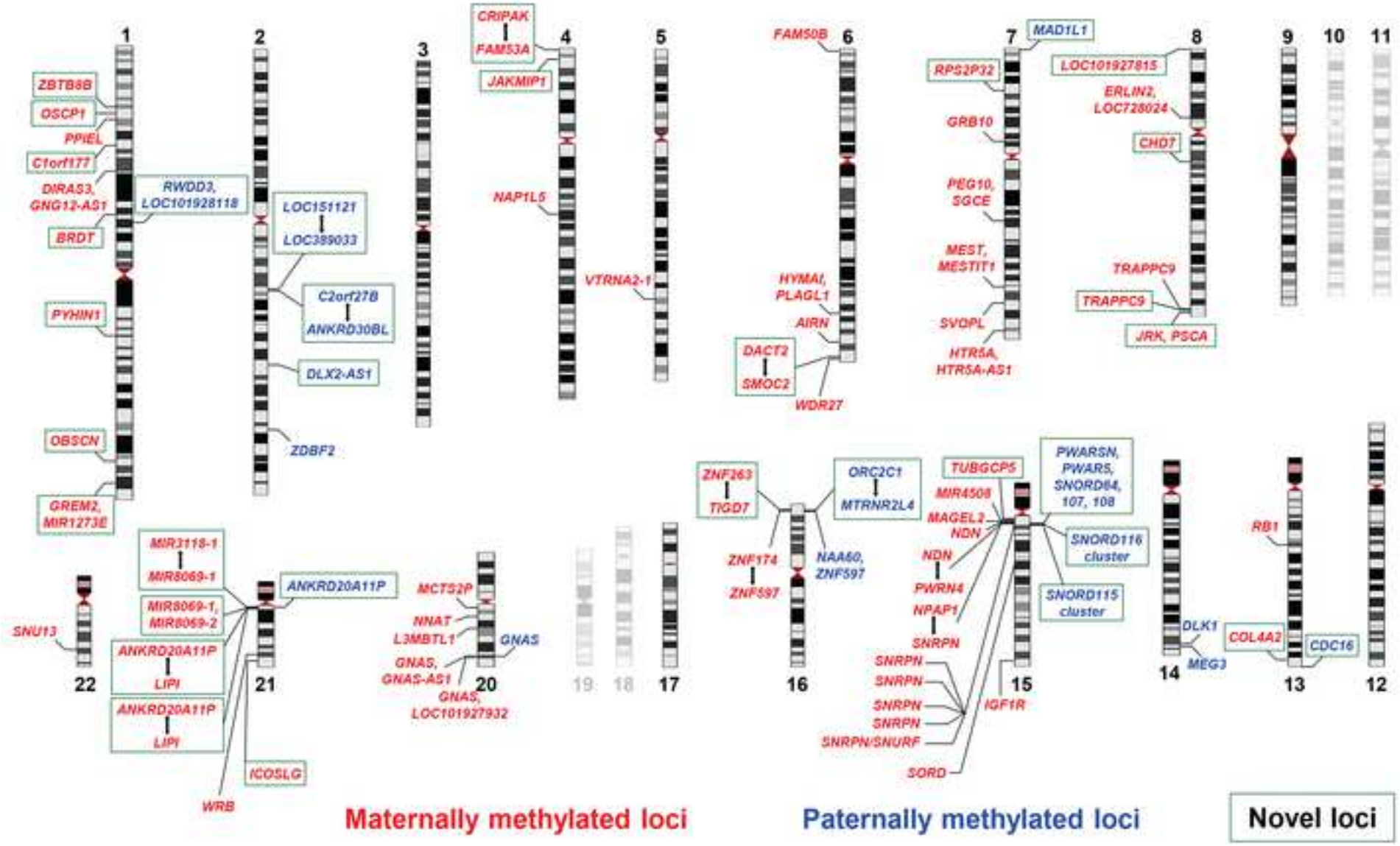
**Supplemental Table 6.** Evidence for imprinted expression from allelic bias in RNAseq data of genes located within 10kb of DMRs in the GTEx cohort (33 different tissues).[4] Transcripts with a test statistic >20 (corresponding to the top 1.02% of all tests performed) are

highlighted in green, while genes not previously reported to be imprinted are highlighted in red.

**Supplemental Table 7.** Beta values for 1,022 imprinted probes in 4,004 population controls.

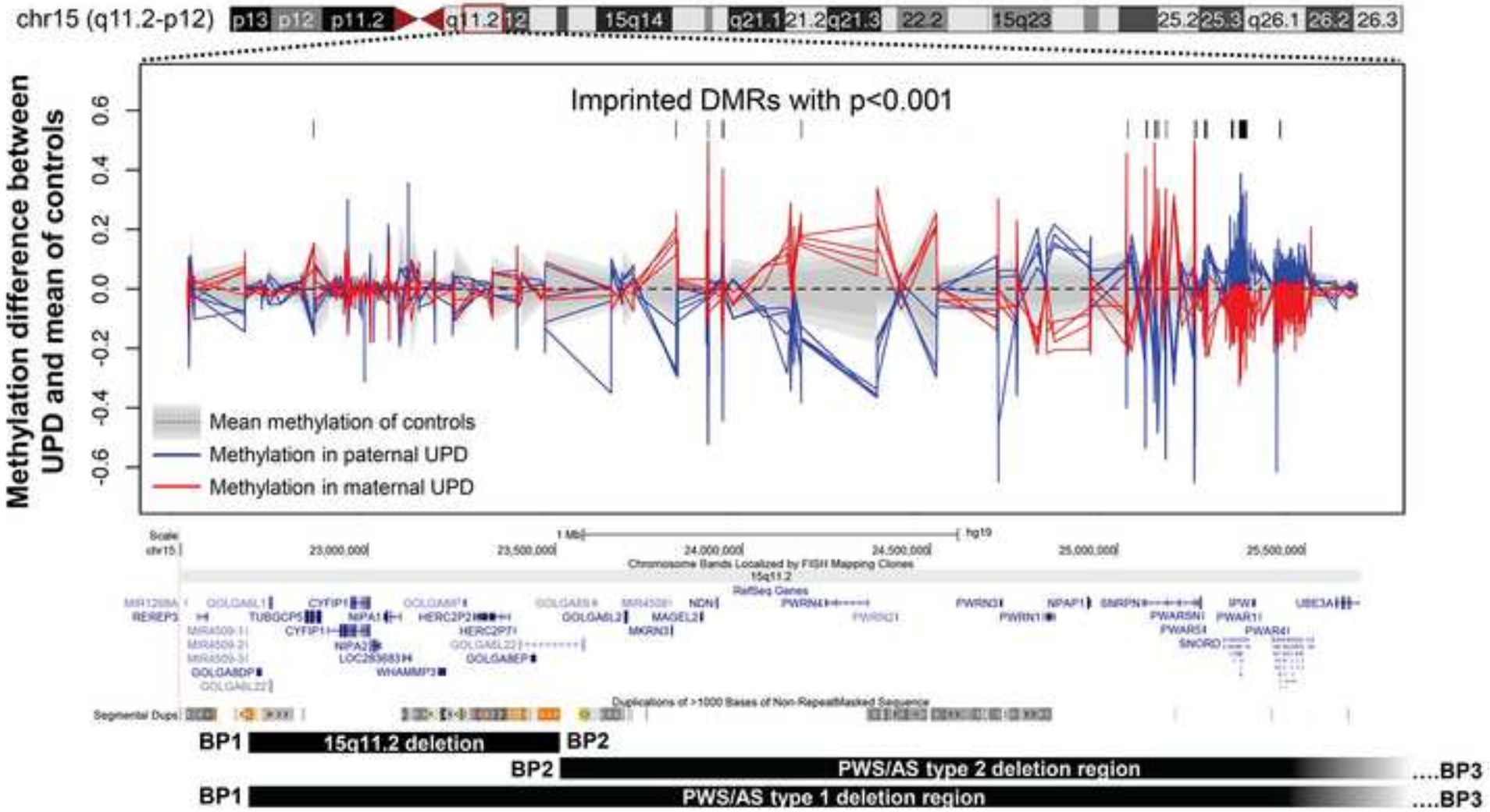
Figure

Click here to download Figure Figure1_hg19_revision 2 version_300dpi.tif



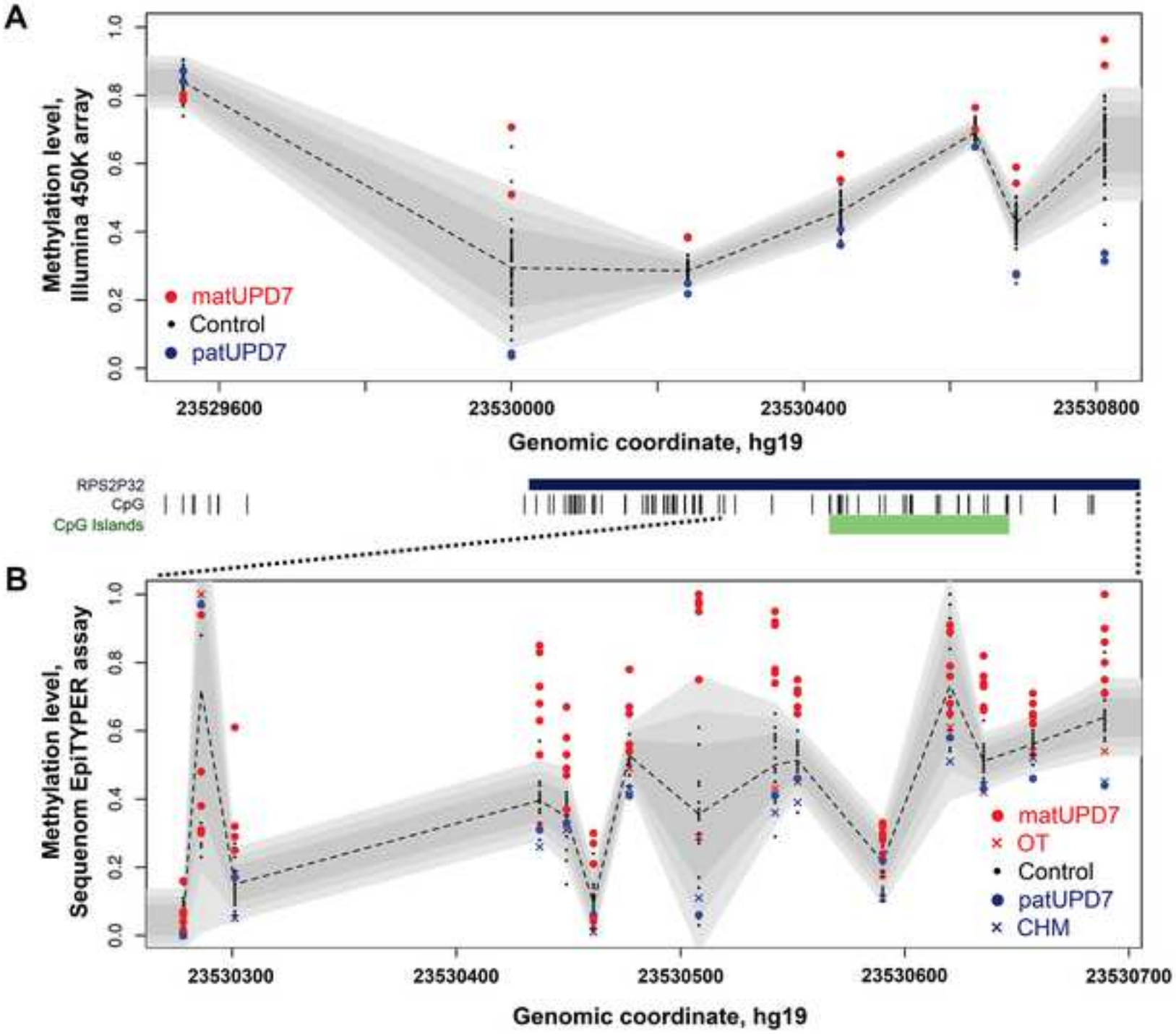
Figure

Click here to download Figure Figure2_PWS_AS region_hg19_v2_300dpi.tif



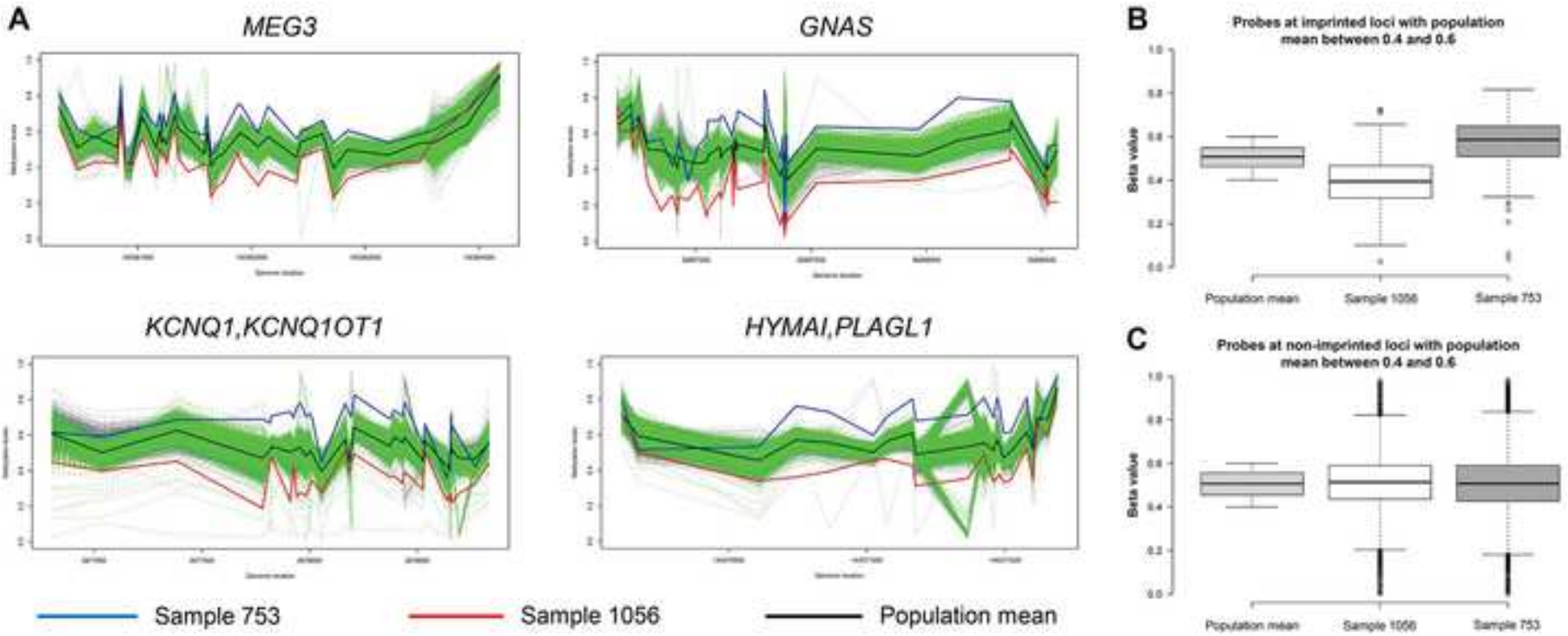
Figure

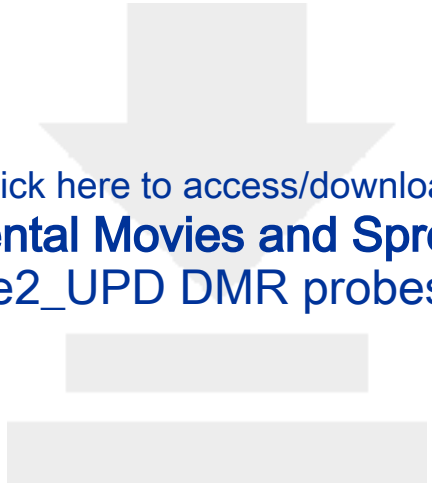
Click here to download Figure Figure3_RPS2P32 updated to hg19_300dpi.tif



Figure

Click here to download Figure Figure4_Controls showing global shifts at 4 imprinted loci_with boxplots_300dpi.tif





Click here to access/download

**Supplemental Movies and Spreadsheets**

Suppl Table2_UPD DMR probes_hg19.xlsx

Click here to access/download

**Supplemental Movies and Spreadsheets**

Suppl Table3_Hits in OT vs CHM_2 stringencies.xlsx

Click here to access/download

**Supplemental Movies and Spreadsheets**

Suppl Table6_GTEx RNAseq allelic bias.xlsx

# *AJHG* Conflict of Interest Form

Cell Press, 50 Hampshire Street, 5th floor, Cambridge, MA 02139, USA

**Please complete this form electronically and upload the file with your final submission.**

*AJHG* requires all authors to disclose any financial interest that might be construed to influence the results or interpretation of their manuscript.

As a guideline, any affiliation associated with a payment or financial benefit exceeding $10,000 p.a. or 5% ownership of a company or research funding by a company with related interests would constitute a financial interest that must be declared. This policy applies to all submitted research manuscripts and review material.

Examples of statement language include: AUTHOR is an employee and shareholder of COMPANY; AUTHOR is a founder of COMPANY and a member of its scientific advisory board. This work was supported in part by a grant from COMPANY.

Please disclose any such interest below on behalf of all authors of this manuscript.

---

**Please check one of the following:**

☒ None of the authors of this manuscript have a financial interest related to this work.

☐ Please print the following disclosure statement in the Acknowledgments section:

**Please provide the following information:**

☒ Please check this box to indicate that you have asked every author of this work to declare any conflicts of interest. Your answers on this form are on behalf of every author of this work.

Manuscript # (research articles only): AJHG-D-16-00165R2

Title: DNA methylation profiling of uniparental disomy cases provides a map of parental epigenetic bias in the human genome

Author list: Ricky Joshi, Paras Garg, Noah Zaitlen, Tuuli Lappalainen, Corey T. Watson, Nidha Azam, Daniel Ho, Xin Li, Stylianos E. Antonarakis, Han G. Brunner, Karin Buiting, Sau Wei Cheung, Bradford Coffee, Thomas Eggermann, David Francis, Joep P. Geraedts, Giorgio Gimelli, Samuel G. Jacobson, Cedric Le Caignec, Nicole de Leeuw, Thomas Liehr, Deborah J. Mackay, Stephen B. Montgomery, Alistair T. Pagnamenta, Peter Papenhausen, David O. Robinson, Claudia Ruivenkamp, Charles Schwartz, Bernhard Steiner, David A. Stevenson, Urvashi Surti, Thomas Wassink, Andrew J. Sharp

Your name: Andrew J. Sharp

Date: June 28 2016