**ImprintSeq, a novel tool to interrogate DNA methylation at human imprinted regions and diagnose multilocus imprinting disturbance**

Eguzkine Ochoa1, Sunwoo Lee1, Benoit Lan-Leung1, Renuka Dias2,3,4, Ken K Ong5,6, Jessica A Radley7,8, Gustavo Pérez de Nanclares9, Rosa Martinez9, Graeme Clark1,10, Ezequiel Martin10, Luis Castaño9, Leonardo Bottolo1,11,12, Eamonn R Maher1,\*

1 Department of Medical Genetics, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK.

2 Institutes of Metabolism and Systems Research, Vincent Drive, University of Birmingham, Birmingham B15 2TT, UK.

3 Centre for Endocrinology, Diabetes and Metabolism, Vincent Drive, Birmingham Health Partners, Birmingham B15 2TH, UK.

4 Department of Paediatric Endocrinology and Diabetes, Birmingham Children's Hospital NHS Foundation Trust, Steelhouse Lane, Birmingham B4 6NH, UK.

5 MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus Box 285, Cambridge, CB2 0QQ, UK.

6 Department of Paediatrics, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus Box 181, Cambridge, CB2 0QQ, UK.

7 West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women’s and Children’s Hospitals NHS Foundation Trust, Birmingham, UK.

8 London North West Regional Genetics Service, St. Mark’s and Northwick Park hospitals, Harrow, Middlesex, U.K.

9 Biocruces Bizkaia Health Research Institute, Hospital Universitario Cruces, CIBERDEM, CIBERER, Endo-ERN, University of the Basque Country (UPV-EHU), Bizkaia, Spain
10 Stratified Medicine Core Laboratory NGS Hub, Cambridge Biomedical Campus, Cambridge, UK.

11 The Alan Turing Institute, London, UK.

12 MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.

\* To whom correspondence should be addressed. Tel: 01223 746714; Fax: 01223 746777 ; Email: erm1000@medschl.cam.ac.uk

**ABSTRACT**

*Purpose*: Disruptions of genomic imprinting are associated with congenital imprinting disorders (CIDs) and other disease states including cancer. CIDs are most often associated with altered methylation at imprinted differential methylated regions (iDMRs). In some cases, multiple iDMRs are affected causing multi-locus imprinting disturbances (MLIDs). The availability of accurate, quantitative and scalable high-throughput methods to interrogate multiple iDMRs simultaneously would enhance clinical diagnostics and research.

*Methods*: We report the development of a custom targeted methylation sequencing panel that covered most relevant 63 iDMRs for CIDs and the detection of MLIDs. We tested in 70 healthy controls and 147 individuals with CIDs. We distinguished loss and gain of methylation per DMR and classified high and moderate methylation alterations.

*Results*: Across a range of CIDs with a variety of molecular mechanisms, ImprintSeq performed at 98.4% sensitivity, 99.9% specificity and 99.9% accuracy (when compared to previous diagnostic testing). ImprintSeq was highly sensitive for detecting MLIDs and enabled diagnostic criteria for MLID to be proposed. In a child with extreme MLID profile a probable genetic cause was identified.

*Conclusion*: ImprintSeq provides a novel assay for clinical diagnostic and research studies of CIDs, MLIDs and the role of disordered imprinting in human disease states.

**INTRODUCTION**

Genomic imprinting is an epigenetic process that regulates parent-of-origin specific expression of >100 human genes with critical roles in growth, metabolism and development1,2. This process is mediated by differential epigenetic marks that establish allele-specific expression patterns during gametogenesis in male and female germ cells. After fertilization, the methylation status of imprinting marks (imprinted differentially methylated regions;iDMRs) are protected and maintained during the second wave of global epigenetic reprogramming1,2. iDMRs are associated with parent-of-origin allele-specific patterns of DNA methylation, chromatin structure and expression of non-coding RNAs1,2. Disruption of a single (or multiple) germline imprinting mark(s) causing altered expression of imprinted genes is associated with congenital imprinting disorders(CIDs)1. Disordered genomic imprinting has also been implicated in complex disorders such as autism and embryonal tumours(e.g., Wilms tumour)3.

Disturbances of the establishment or maintenance of imprinting can be caused by multiple mechanisms including primary genetic alterations (e.g., pathogenic sequence variants, deletions/duplications, cytogenetic rearrangement, uniparental disomies) with epigenetic consequences or by primary epigenetic alterations(epimutations) without underlying genetic alterations4. Each CID has been linked with locus-specific defects and disease-specific molecular alterations (*Supplementary Table 1*)1. For example, depending on the parental-origin, deletions at 15q11-q13 may cause Angelman(AS) or Prader-Willi(PWS) syndromes and epimutations at 11p15.5 may cause Beckwith-Wiedemann spectrum(BWSp) or Silver-Russell syndromes(SRS) (*Supplementary Table 1*). Interestingly, a subset of CID cases exhibit multiple imprinting alterations across the genome, i.e. multi-locus imprinting disturbances(MLIDs)5. MLIDs are more common in CIDs mainly caused by epimutations (*Supplementary Table 1*). Children with MLIDs may have atypical phenotypes6.

For clinical molecular diagnosis of CIDs, methylation analysis at iDMRs is the standard methodology worldwide. Whilst whole-genome bisulfite sequencing(WGBS) or array-based technologies can provide genome-wide methylation assessments, WGBS is generally too expensive for routine diagnostics, array-based technologies cover only partially some iDMRs and targeted methylation assays such as MS-MLPA7 are most widely-used. To provide an alternative methodology suitable for clinical diagnostics and research, we designed a quantitative scalable high-throughput method to interrogate individual CpG methylation status at iDMRs using a custom targeted methylation sequencing panel, ImprintSeq.

**MATERIAL AND METHODS**

**Samples**

Genomic DNA from a total of 217 samples was extracted from peripheral blood using Gentra Puregene Blood Kit(Qiagen) or other standard methods and quantified by Qubit™ dsDNA BR Assay Kit(Invitrogen, ThermoFisher). The cohort included 70 healthy controls and 147 with a CID (see *Supplementary Table 2*). For 133 individuals with a CID a molecular diagnostic report was available. For 14 individuals with a CID a molecular diagnostic report was not available and for these individuals the molecular classification was based on the ImprintSeq results (which were consistent with the clinical phenotype). Appropriate human subjects’ approvals and written informed consent were obtained from all participants. The study was approved by South Birmingham Research Ethics Committee.

**Targeted methylation sequencing: TrueMethyl® On Target**

ImprintSeq was developed as a collaboration between Cambridge Epigenetix(CEGX, Cambridge, UK) company, Tecan group(NuGen Technologies, Redwood, CA) and the University of Cambridge. This hybridization-based custom panel uses 3,989 probes to cover 203,279 base pairs across 63 genomic regions involved in genomic imprinting (*Table 1*). Probes were designed for both strands of bisulfite converted DNA. The design interrogates over 9,257 CpG sites (see *Supplementary Table 3*) in low input DNA(50-200 ng). Methylation levels were calibrated at 0%, 50% and 100% by mixing different ratios of unmethylated/methylated human control DNA(Zymo, CA, USA). Samples were bisulphite converted with EZ DNA Methylation-Gold kit(Zymo, CA, USA) following the manufacturer's instructions. Details of the experimental protocol are provided in *Supplementary Data*.

**Methylation analysis and interpretation**

After filtering by coverage and quality (see *Supplementary Data*), correction of the methylation levels for 145 cases and 70 healthy controls was performed using MethylCal8. After correcting the observed methylation levels of 215 individuals, we calculated the median of the methylation levels across CpGs (median methylation level (MML)) for each individual and each iDMR. For each iDMR, MMLs below the healthy controls’ 3 standard deviations (SDs) confidence interval were considered to undergo loss of methylation (LOM) and those above 3SDs confidence interval were designated gain of methylation (GOM). To avoid false-positive diagnoses we selected a stringent clinically-motivated 3SD confidence interval to guarantee low Type-I error (α = 0.0027).

To increase the specificity of MLID diagnosis, we divided significant LOM/GOM events according to the magnitude of the alteration in the differential median methylation level (DMML) between controls and patients: high values of DMML (high methylation alterations(HMA)) and moderate values of DMML (moderate methylation alterations(MMA)). The detection of the two categories was performed by using a mixture model of beta densities9, an unsupervised clustering algorithm that classifies the absolute value of DMMLs into HMA and MMA groups. The in-house implementation was based on the expectation-maximization algorithm (EM algorithm)10. The optimal classification cut-off point that separates the HMA and MMA groups corresponds to the point where the densities of the two beta components of the mixture model intersect. In our data set based on 943 significant DMMLs (722 LOM and 221 GOM), the cut-off point was detected at 0.185. (see *Supplementary Figure 3*).

ImprintSeq methylation values were compared to those from Illumina Infinium HumanMethylation450 BeadChips (Illumina, San Diego, CA) analysis (see *Supplementary Methods*). Details of whole exome sequencing (performed in two individuals) are in *Supplementary Methods*).

**RESULTS**

**Methylation profiling in healthy controls**

Methylation profiling for 63 iDMRs was determined in 70 healthy control samples (*Figure 1*). Mean coverage per iDMR varied from 154.5 to 1,128.5 reads (see *Supplementary Table 4*) and MML ranged from 0.0 to 1.0 with a median of 0.50 (see *Supplementary Table 5*). For 8 iDMRs associated with CIDs (i.e., CIDs-associated iDMRs: PLAGL1:alt-TSS-DMR, MEST:alt-TSS-DMR, GRB10:alt-TSS-DMR, H19/IGF2:IG-DMR, KCNQ1OT1:TSS-DMR, MEG3:TSS-DMR, SNURF:TSS-DMR and GNASA/B:TSS-DMR)11–15, mean coverage per iDMR varied from 165.2 to 974.1 reads (see *Supplementary Table 4*) and MML ranged from 0.21 to 0.62 (median 0.43) (see *Supplementary Table 5*). None of 70 controls had an MML outside 3SDs confidence interval at the 8 CIDs-associated iDMRs.

**Diagnosis of Congenital Imprinting disorders with ImprintSeq**

131 individuals (data from two individuals were excluded for technical reasons) with a CID and molecular diagnostic report (AS(n=1), BWS(n=50), PHP1b(n=42), SRS(n=24) and TNDM(n=14)) were analysed at 8 CID-associated iDMRs (*Figure 1*). The overall classification by ImprintSeq along with the previous diagnostic testing result is summarised in *Figure 2*.

*Beckwith-Wiedemann Spectrum(BWSp) testing*

Fifty individuals with BWSp were studied. Thirty-nine had epimutation at IC1 (GOM at H19–IGF2:IG-DMR; n=1) or IC2 (LOM at KCNQ1OT1:TSS-DMR; n=38). ImprintSeq identified 38 LOM at IC2 and one GOM at IC1. The IC2 MML for controls was 0.46 and 84% (32/38) BWS cases with IC2 LOM showed an MML below 0.1 (*Supplementary Table 6*).

Six individuals with BWSp and previously diagnosed paternal UPD(11p15) were analysed. In four cases, LOM at KCNQ1OT1:TSS-DMR and GOM at H19–IGF2:IG-DMR and IGF2:alt-TSS-DMR, consistent with patUPD11 was detected. The MML at KCNQ1OT1:TSS-DMR in these cases was 0.15 to 0.3 which is consistent with mosaic patUPD11 (*Supplementary Table 6*). Two cases previously classified as BWSp with patUPD(11p15) did not show methylation alterations characteristic of UPD, one showed GOM at H19–IGF2:IG-DMR, IGF2:Ex9-DMR and IGF2:alt-TSS-DMR (MML-1.0, 0.81 and 0.78 respectively) but no alteration at KCNQ1OT1:TSS-DMR (MML=0.44) and the other one, no alteration was detected at 11p15.5 (MMLs at H19–IGF2:IG:DMR, IGF2:Ex9-DMR and IGF2:alt-TSS-DMR and KCNQ1OT1:TSS-DMR were 0.56, 0.52, 0.48 and 0.46 respectively).

All five individuals with suspected BWSp who previously tested negative in a clinical diagnostic laboratory had given a normal ImprintSeq testing result.

*Silver-Russell syndrome(SRS) testing*

In the SRS group (n=24), 17 cases had been diagnosed with an IC1 epimutation (LOM at H19–IGF2:IG-DMR) and in two cases no alteration had been detected by MS-MLPA. All previously positive cases showed H19–IGF2:IG-DMR LOM below 3SDs confidence interval, although the degree of LOM was generally less pronounced than that seen in BWSp with IC2 epimutations (see above) or PHP1b with LOM at GNAS A/B:TSS-DMR (see below) (*Supplementary Table 6*). Only 6 out of 17 (35%) SRS cases with H19–IGF2:IG-DMR LOM had a MML below 0.1 (the other 10 had values of 0.1-0.27 (*Supplementary Table 6*). One of the two patients with a clinical diagnosis of SRS but normal prior diagnostic testing showed LOM at H19–IGF2:IG-DMR with ImprintSeq (MML of 0.49 compared to MML in control cohort of 0.56 (lower boundary of 3SDs confidence interval was 0.50) (*Supplementary Table 6*).

In five individuals diagnosed with SRS and maternal UPD7, ImprintSeq demonstrated almost complete GOM at GRB10:alt-TSS-DMR, PEG10:TSS-DMR, MEST:alt-TSS-DMR and HTR5A:TSS-DMR consistent with matUPD7.

*Pseudohyoparathyroidism 1B(PHP1b) testing*

Forty-one individuals with a clinical diagnosis of PHP1b were studied. Previous MS-MLPA testing had classified 14 of them as GNAS A/B:TSS-DMR LOM and 27 with GNAS A/B:TSS-DMR LOM, GNAS-XL:TSS-DMR LOM, GNAS-AS1:TSS-DMR LOM, GNASNESP:TSS-DMR GOM. Testing by ImprintSeq showed complete concordance with the previous diagnosis from service molecular genetic laboratories. All PHP1b cases with GNAS A/B:TSS-DMR LOM showed a complete LOM (all MLL <0.1) (*Supplementary Table 6*). MML of those PHP1b cases with GNAS A/B:TSS-DMR LOM, GNAS-XL:TSS-DMR LOM, GNAS-AS1:TSS-DMR LOM, GNASNESP:TSS-DMR GOM was consistent with complete GOM (>0.89 MML) at GNASNESP:TSS-DMR and complete LOM (<0.05 MML) at GNAS A/B:TSS-DMR, GNAS-XL:TSS-DMR and GNAS-AS1:TSS-DMR (*Supplementary Table 6*) in most cases (20/27). Two cases showed partial changes in all GNAS DMRs and 5 of 27 (19%) showed partial LOM at GNAS-XL:TSS-DMR and complete LOM at GNAS A/B:TSS-DMR, GNAS-AS1:TSS-DMR and complete GOM at GNASNESP:TSS-DMR.

One PHP1B case was known to have a maternal duplication at GNAS locus covering GNAS-XL:TSS DMR, GNAS-AS1:TSS-DMR and GNASNESP:TSS-DMR16. ImprintSeq profiling demonstrated LOM at GNAS A/B:TSS-DMR (MML = 0.05) and GNASNESP:TSS-DMR, and GOM at GNAS-XL:TSS-DMR and GNAS-AS1:TSS-DMR ((MMLs 0.21, 0.65 and 0.56 respectively) consistent with the maternal duplication observed by SNP array.

*Testing for other CIDs: Transient Neonatal Diabetes Mellitus (TNDM) and Angelman syndrome (AS)*

Fourteen individuals with TNDM (2 with PLAGL1:(alt)-TSS-DMR-LOM, 6 with paternal UPD6 and 6 with a PLAGL1:(alt)-TSS-DMR duplication) were tested. The 2 TNDM with PLAGL1:(alt)-TSS-DMR LOM showed MLL <0.05 (Supplementary Table 6). The 6 TNDM cases with paternal UPD6, showed complete LOM at PLAGL1:(alt)-TSS-DMR (all MMLs ~0.02, see *Supplementary Table 6*) and almost complete LOM at one or more DMRs at chromosome 6 (FAM50B:TSS DMR, IGF2R:Int2-DMR and WDR27:Int13-DMR). Analysis of 6 TNDM cases with duplication at PLAGL1:(alt)-TSS-DMR demonstrated partial LOM in this DMR (MML 0.193 to 0.274), these methylation changes were undetectable previously by MS-MLPA. In an individual previously diagnosed with mosaic AS, there was significant LOM (MML < 0.15) at SNRPN:Int1-DMR2, SNRPN:Int1-DMR1, SNRPN alt-TSS-DMR and SNURF:TSS-DMR, validating previous diagnostic.

To determine sensitivity, specificity and accuracy of the assay, we included all CID cases with diagnostic reports excluding those classified as NA (not available) or NAD (no alteration detected) (*Supplementary Table 2*). Overall, across a range of CIDs with a variety of molecular mechanisms (compared to previous diagnostic results), ImprintSeq performed at 98.4% of sensitivity, 99.9% specificity and 99.9% accuracy assuming 0.01% prevalence (fraction of people found to have the condition with the total number of people studied). From this cohort only 2 cases were not validated, the two patUPD11 BWS cases.

**ImprintSeq testing for multi-locus imprinting disturbances (MLIDs)**

To investigate the utility of ImprintSeq to detect MLID we interrogated 63 imprinted regions in 145 individuals with known or suspected CIDs (the 131 individuals with a CID studied in a diagnostic laboratory studied to assess the diagnostic potential of ImprintSeq plus 14 further individuals with a CID for whom a diagnostic report was not available. For these 14 individuals we considered that the ImprintSeq results (which were consistent with the clinical phenotype) could be used for diagnostic classification). Applying our method, 22/70 controls had significant DNA methylation alterations (GOM/LOM) at one or more iDMRs (median 0.37, range 0-2) (*Figure 3*). The most frequent iDMR with LOM was RTL1 and the most frequent iDMR with GOM was IGF2R:Int2-DMR (none of the controls showed LOM/GOM at a CIDs-associated iDMRs).

The 63 iDMRs were then analysed in the 145 individuals with a clinical and/or molecular diagnosis of CID. Overall, 141 of 145 individuals demonstrated LOM/GOM at one or more iDMRs (median 5.19, range 1 to 35). 20 iDMRs demonstrated LOM, 7 iDMRs demonstrated GOM and 32 both LOM and GOM. The iDMRs located outside of CID-associated imprinted gene clusters (nonCID-iDMRs) with the most frequent LOM were FANCC:In61-DMR, SNU13:alt-TSS-DMR and IGF1R:Int2-DMR (14.3%, 10.2% and 9.5% of cases respectively). The nonCID-iDMRs with the most frequent GOM were ZDBF2/GPR1-AS:IG-DMR and PEG13:TSS-DMR (20.5% and 14.6% of cases). To replicate the methylation profiling results of ImprintSeq at non-CID-iDMRs, we compared the ImprintSeq methylation results in a subset of CID cases (BWS (n=43) and SRS (n=22)) and healthy controls (n=1,472) for whom methylation profiling with the Illumina Infinium Human Methylation 450 BeadChip array (450K) had been performed (*Supplementary Figure 2*). Within this cohort ImprintSeq identified 247 significant methylation alterations (157 LOM, 90 GOM) compared to 129 significant methylation alterations (96 LOM, 33 GOM) detected by 450K. The comparison revealed 91.5% (119) of LOM/GOM detected by the 450K analysis were also detected by ImprintSeq and 8.5% (10) of LOM/GOM findings detected by 450K analysis were not detected by ImprintSeq. 140 significant LOM/GOM alterations (in 51 patients) detected by ImprintSeq were not identified by 450k array. iDMR alterations that were more likely to be identified by ImprintSeq analysis were ZDBF2/GPR1-AS:IG-DMR GOM, IGF2:Ex9-DMR LOM, PEG10:TSS-DMR LOM, MEST:alt-TSS-DMR LOM/GOM or IGF1R:Int2-DMR-LOM and 450K analysis were GNASNESP:TSS-DMR-GOM, and L3MBTL1:alt-TSS-DMR-LOM. The number of CpGs per DMR interrogated by 450K and ImprintSeq differs, being significantly higher in ImprintSeq (*Table 1*). This might explain the limited capacity of the array to detect methylation alteration at some specific DMRs (PPIEL:Ex1-DMR, IGF1R:Int2-DMR, ERLIN2:Int6-DMR and L3MBTL1 alt-TSS-DMR).

ImprintSeq detected a higher frequency of MLID in CID subgroups than previously reported but the number of iDMRs analysed was much larger than in most previous reports and the number of CpGs at each iDMR analysed is higher than for array-based technologies or MS-MLPA. To develop diagnostic criteria for the identification of MLID by ImprintSeq, we divided the LOM/GOM events detected by ImprintSeq into two groups: high (HMA) and moderate methylation alterations (MMA) (see Methods) (*Figure 4*). Most of LOM/GOM findings (85%) detected by 450K analysis corresponded to ImprintSeq HMAs (*Supplementary Figure 2*) and 450K analysis displayed limited capacity of to detect MMAs (only 6.1% LOM/GOM MMA signals were detected with 450K). Applying the HMA/MMA criteria, all LOM/GOM alterations at diagnostic iDMRs regions were classified as HMAs but the majority (6/7) of methylation changes associated with duplications in diagnostic regions were classified as MMAs (*Figure 4b*). We then compared the effect of classifying alterations as HMA/MMAs at iDMRs that were located outside of CID-associated imprinted gene clusters (nonCID-iDMRs). For some nonCID-iDMRs, increasing the diagnostic threshold did not impact the ImprintSeq LOM/GOM findings (PPIEL:Ex1-DMR, DIRAS3:Ex2-DMR, IGF2R:Int2-DMR, ERLIN2:Int6-DMR, FANCC:In61-DMR, INPP5F:Int2-DMR, IGF1R:Int2-DMR, NNAT:TSS-DMR, WRB:alt-TSS-DMR and SNU13:alt-TSS-DMR) (113 of 115 detected LOM/GOM findings were classified as HMAs) (*Supplementary Figure 1*). However for other non-CID-iDMRs (ZDBF2/GPR1-AS:IG-DMR, PEG13:TSS-DMR, MAGEL2:TSS-DMR, *TCEB3C*, *NLRP2*, *RTL1* and *GABRA5*) (*Supplementary Figure 1*) most alterations (91/97) were classified as MMAs (*Figure 4*). An interesting observation was that LOM/GOM at primary iDMRs was more likely to be classified as a HMA than at LOM/GOM at secondary iDMRs (secondary DMRs acquire differential DNA methylation after fertilization whereas primary/germline iDMRs (giDMRs) become differentially methylated in the germline17. 82% (338/410) of LOM alterations detected by ImprintSeq in primary iDMRs (82% (305/370) oocyte and 82% (33/40) sperm primary iDMRs) were classified as HMA whereas only 44% (16/36) in secondary iDMR were HMAs (*P* < 10-5) (See *Supplementary Table 8*). However, for GOM alterations there was no apparent difference between the proportion of HMAs at primary and secondary iDMRs (63% (63/99) and 76% (41/54) respectively, *P*= 0.1478) (See *Supplementary Table 8*).

Considering only HMAs, 3 of 70 controls and 56 of 145 with CIDs had LOM or GOM at one or more nonCID-iDMRs and 0 of 70 controls and 31 of 145 with CIDs had LOM or GOM at two or more nonCID-iDMRs. To avoid overdiagnosis of MLID in a clinical diagnostic testing setting we proposed that, after excluding UPD, the diagnostic criteria for MLID should consist of (in addition to the primary CID-associated epimutation) either a LOM/GOM HMA at a CID-associated iDMR or two LOM/GOM HMAs at non-CID iDMRs. Applying these criteria the frequencies of MLID in specific CID-molecular subgroups were: BWS with IC1 GOM (0/3, 0%), BWS with IC2 LOM (17/40, 42%), PHP1b with GNAS A/B:TSS DMR LOM (0/14, 0%), SRS with IC1 LOM (2/17, 12%) and TNDM with PLAGL1:alt-TSS-DMR LOM (2/2, 100%) (See *Supplementary Table 7*). Application of the proposed MLID criteria had a marked effect on two CID cases: a patient with BWS with 22 LOM/GOM iDMR alterations but only 3 HMA (CID0047) and a patient with PHP1b with 15 LOM/GOM iDMR alterations but only 5 HMAs (CID0081). CID0047 showed 3 HMAs but no primary CID-associated epimutation and CID0081 showed 5 HMAs, 4 of them in GNAS-locus (probable UPD20) and only one HMA at non-CID iDMRs. After applying the proposed diagnostic criteria for MLIDs in these cases, none of them was considered to be a MLID positive case.

In total 21/145 (14%) individuals with CID satisfied the MLID diagnostic criteria with a median of 4 HMAs (range 2-26) per individual.

**Characteristics of a case with an extreme MLID profile detected by ImprintSeq**

The most extreme MLID profile was detected in a child with atypical SRS (case # CID0120) and 26 HMAs. There was no relevant family history and no underlying cause for MLID had been diagnosed. We undertook exome sequencing in the child and interrogated a panel of MLID-associated genes (*ZFP57/NLRP2*/*NLRP5*/*NLRP7/PADI6*)1finding a previously unreported heterozygous *PADI6* frameshift insertion (NM\_207421.4:c.1874dup NP\_997304.3:p.(Asn626GlufsTer38). Exome sequencing in both parents revealed the presence of the same frameshift insertion in *PADI6* in the mother but not the father. Also, a heterozygous *PADI6* missense substitution (NM\_207421.4:c.1456T>C, NP\_997304.3:p.(Cys486Arg)) and a heterozygous *NLRP5* missense substitution (NM\_153447.4:c.1111C>T NP\_703148.4:p.(Leu371Phe)) were both detected in the mother. Both missense variants were classified as VUSs (PM1, PM2, PP3 and PM1, BP4 respectively) by ACMG/AMP criteria.

**DISCUSSION**

Disruption of establishment or maintenance of genomic imprinting has been described in a variety of disorders including congenital imprinting disorders and neoplasia1,2. Disturbances in genomic imprinting are commonly associated with alterations in the methylation status of iDMRs and so methylation profiling of iDMRs is usually the first-line investigation for the diagnosis of CIDs. Different methodologies have been used for iDMR profiling in clinical diagnostic and research settings including methylation-specific PCR, pyrosequencing, MS-MLPA, methylation arrays and whole-genome bisulphite sequencing(WGBS)18. For diagnostic testing MS-MLPA is most widely used whereas methylation arrays are widely employed in research studies. An advantage of methylation arrays such as the Illumina Infinium and EPIC arrays is the large number of CpGs that are tested (485,577 and 850,000 CpGs) respectively (though only a small fraction of CpGs profiled are at iDMRs). WGBS offers comprehensive analysis but is expensive, and the depth of coverage may be inadequate to detect partial methylation alterations. MS-MLPA can detect copy number changes but can be insensitive to mosaic methylation defects that are common in some CIDs. Accordingly, we and others have sought to develop cost-effective targeted methods to accurately measure methylation at a base-pair resolution in multiple iDMRs. Recently, Klobucar *et al* described IMPLICON19, a sequencing-based assay developed to analyse iDMRs in mice which was then adapted to investigate 14 human iDMRs. IMPLICON successfully detected aberrant methylation at iDMRs in human-induced pluripotential stem cells and in 2 patients with a CID. Although both IMPLICON and ImprintSeq were designed for similar purposes there are significant differences, ImprintSeq targets a larger number of iDMRs than IMPLICON (63 and 14 respectively). Furthermore, IMPLICON is an amplicon-based method (while ImprintSeq is a capture-based approach)20 that has not yet been validated against other diagnostic assays for a range of CIDs as described here.

We examined ImprintSeq for the diagnosis of CIDs and detection of MLIDs. ImprintSeq showed excellent correlation with the previous diagnostic results from genetics service laboratories. ImprintSeq was particularly capable of identifying CID patients with epimutations (97/97) and duplications (7/7) and most UPD cases (15/17), but two patients reported to have BWSp with paternal 11p15.5 UPD were not detected. patUPD in BWSp is invariably mosaic and its diagnosis may require specialist investigations such as SNP arrays or testing of multiple tissues when the level of mosaicism is low15. Based on our results, the theoretical threshold for mosaicism for ImprintSeq is ~20-30% (unpublished data). Lower levels of mosaic UPD (e.g. 5-20%) can be detected by SNP arrays (37) and borderline ImprintSeq results could be followed up by further analyses. In the longer-term, with further development, we would expect that given the high depth of coverage of ImprintSeq combining data from different loci and adding SNP information generated by ImprintSeq should make it possible to detect lower levels of UPD.

In the last decade, there has been increasing interest in the occurrence of MLIDs in CIDs5. However, the clinical implications of finding MLIDs is still debated. Though there are examples of in which the presence of MLID has been linked to an atypical CID phenotype21,22 , in many cases the effect of MLID is unclear and MLID has been reported with an apparently normal phenotype6. A difficulty in studying MLID in CID cohorts is the variability in the iDMRs investigated between studies and the inconsistency of analytical methods and criteria for diagnosing MLID. We directly compared iDMR profiling by ImprintSeq and methylation array. Most LOM/GOM events detected by methylation array were detected by ImprintSeq but there were a large number of LOM/GOM alterations detected by ImprintSeq but not array methylation profiling. We had expected that ImprintSeq should be more sensitive because of the larger number of CpGs covered (see *Table 1*) and the higher depth of coverage. Nevertheless, whilst ImprintSeq is intended to be a sensitive diagnostic test for CID, we wished to avoid overdiagnosing MLID and used stringent criteria to diagnose HMAs. Interestingly, we found that HMA events accounted for most alterations at primary iDMRs and MMAs at secondary iDMRs. Primary iDMRs (oocyte and sperm gDMR) are established during gametogenesis in primordial germ cells and secondary iDMRs are established after fertilization17 and consequently, secondary iDMRs might be more likely to show mosaic and/or polymorphic variation. Whereas CID-associated iDMRs have been extensively profiled in diagnostic laboratories, for non-CID-associated iDMRs there is generally much less information about polymorphic epigenetic variation23. Hence, we proposed that diagnostic criteria for MLID should concentrate on HMAs and differentially consider LOM/GOM at CID-associated iDMRs and non-CID-associated iDMRs. The proposed MLID diagnostic criteria (exclusion of UPD and, in addition to the primary CID-associated epimutation, either a LOM/GOM HMA at another CID-associated iDMR or LOM/GOM HMAs at two non-CID iDMRs) gave frequencies of MLID of 0% in control subjects, 0% in BWS with IC1 GOM, 42% in BWS with IC2 LOM, 0% in PHP1b with GNAS A/B:TSS DMR LOM and 12% in SRS with IC1 LOM. ImprintSeq profiling of larger cohorts of CID individuals will enable the validation of these proposed diagnostic criteria and the significance of HMAs at specific iDMRs for patient phenotype to be assessed.

In most children with a CID and MLID no cause will be identified. However, a history of assisted reproductive technology(ART) fertilization has been linked to an increased risk of MLIDs in some studies and in a small subset of patients with MLIDs a genetic cause may be identified24–27. We tested the family of the child with the most extreme ImprintSeq MLID profile (26 HMAs at iDMRs) and identified that her mother was a compound heterozygote for a loss of function *PADI6* variant and a missense substitution (classed as VUS). A missense VUS in *NLRP5* was also detected in the mother. These findings would be consistent with maternal biallelic pathogenic variants in *PADI6* presenting a CID and MLID in the child. Previously maternal biallelic *PADI6* pathogenic variants have been associated with infertility, embryonic arrest and BWS with MLIDs28–31. Such an interpretation would argue for the clinical utility of MLIDs to identify families with a high recurrence risk for CIDs or other MLID-related disorders.

Healthy controls studied showed none significant alterations at 8 CID-associated iDMRs. However, we detected significant alterations at non-diagnostic iDMRs in 22 of 70 and HMAs in 10 of 70. The most frequently affected iDMR was IGF2R:Int2-DMR (GOM in 3/70 controls), which is known to be polymorphically imprinted in humans23,32. This suggests other iDMRs could show polymorphic imprinted methylation in general population and this might be the cause of alterations detected in healthy controls. Beside this, some studies suggested the induction of iDMRs DNA methylation alterations by the exposure to some environmental factors33–35, however these studies are limited. To determine which iDMRs are more frequently affected in the general population, in which percentage and the consequences of them for human health, will require a quantitative high-throughput method, like ImprintSeq, to interrogate large cohort of healthy individuals.

The main limitations of this work is the sample size used taking into account the high variability of alterations and mosaicism levels described in these patients. However, despite the low frequency of these rare disorders, we presented the largest CID cohort study to date with high-throughput technologies. Larger cohort of samples representing this variability will be necessary to improve the analysis and diagnostic of CIDs and MLIDs.

In conclusion, ImprintSeq is a novel method that can diagnose the iDMR methylation alterations that account for most CIDs. Investigation of larger patient cohorts is required to refine and optimise the diagnostic promise of ImprintSeq but the assay is a highly promising tool for clinical diagnostics and research into the role of imprinting in human disease.

**ACKNOWLEDGEMENT AND FUNDING**

## This research was co-funded by the NIHR Cambridge Biomedical Research Centre and Rosetrees Trust (to EO, SL and ERM), Marie Curie Initial Training Network Award (INGENIUM) (BLL, ERM), from the European Union, Alan Turing Institute under the Engineering and Physical Sciences Research Council [EP*/*N510129*/*1 to LB]. Medical Research Council [MR/L00002/1, MC\_UU\_12015/1, and MC\_UU\_12015/2 to KO. The University of Cambridge has received salary support (ERM.) from the NHS in the East of England through the Clinical Academic Reserve. The views expressed are those of the authors and not necessarily those of the NHS or Department of Health. Funding for open access charge: Cambridge Open Access: https://www.openaccess.cam.ac.uk/. We acknowledge support from the NIHR UK Rare Genetic Disease Research Consortium.

**ETHICS STATEMENT**

Appropriate human subjects’ approvals and written informed consent were obtained from all participants. The study was approved by South Birmingham Research Ethics Committee.

**DATA AVAILABILITY**

The datasets supporting the conclusions of this article is(are) included within the article (and in supplementary material. Beta-values from Illumina Infinium HumanMethylation450 BeadChips analysis in 65 CID cases (BWS (n = 43) and SRS cases (n = 22)) have been deposited in GEO (accession number GSE166531 [www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166531](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166531)). The bioinformatics’ pipelines used for the analysis of ImprintSeq data have been deposited in GitHub (https://github.com/eguz8A/ImprintSeq).

**AUTHOR INFORMATION**

Conceptualization: E.O.; L.B.; E.R.M.; Data curation: E.O.; S.L; B.L.L; K.K.O; E.M.; L.B.; E.R.M.; Formal analysis: E.O.; L.B.; Funding acquisition: K.K.O.; L.B.; E.R.M.; Investigation: E.O.; S.L; B.L.L; G.C; E.M; L.B.; E.R.M.; Resources: R.D,; K.K.O; J.A.R.; G.P.N.; R.M.; L.C.; E.R.M.; Supervision: E.O.; L.B.; E.R.M.; Writing – original draft: E.O.; L.B.; E.R.M.; Writing – review & editing: E.O.; S.L; B.L.L; R.D.; K.K.O.; J.A.R.; G.P.N.; R.M.; G.C; E.M.; L.C. L.B.; E.R.M.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**REFERENCES**

1. Monk D, Mackay DJG, Eggermann T, Maher ER. Genomic imprinting disorders : lessons on how genome , epigenome and environment interact. *Nat Rev Genet*. 2019;20:235-248. doi:10.1038/s41576-018-0092-0

2. Peters J. The role of genomic imprinting in biology and disease: An expanding view. *Nat Rev Genet*. 2014;15(8):517-530. doi:10.1038/nrg3766

3. Coorens THH, Treger TD, Al-Saadi R, et al. Embryonal precursors of Wilms tumor. *Science (80- )*. 2019;366(6470):1247-1251. doi:10.1126/science.aax1323

4. Eggermann T, Perez de Nanclares G, Maher ER, et al. Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin Epigenetics*. 2015;7(1). doi:10.1186/s13148-015-0143-8

5. Sanchez-Delgado M, Riccio A, Eggermann T, et al. Causes and Consequences of Multi-Locus Imprinting Disturbances in Humans. *Trends Genet*. 2016;32(7). doi:10.1016/j.tig.2016.05.001

6. Sparago A, Verma A, Patricelli MG, et al. The phenotypic variations of multi-locus imprinting isturbances associated with maternal-effect variants of NLRP5 range from overt imprinting disorder to apparently healthy phenotype. *Clin Epigenetics*. 2019;11(1):1-10. doi:10.1186/s13148-019-0760-8

7. Grafodatskaya D, Choufani S, Basran R, Weksberg R. An Update on Molecular Diagnostic Testing of Human Imprinting Disorders. *J Pediatr Genet*. 2016;06(01):003-017. doi:10.1055/s-0036-1593840

8. Ochoa E, Zuber V, Fernandez-jimenez N, et al. MethylCal : Bayesian calibration of methylation levels. 2019:1-14. doi:10.1093/nar/gkz325

9. McLachlan GJ, Peel D. *Finite Mixture Models*. Wiley; 2000.

10. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J R Stat Soc*. 1977;39(1):1-38.

11. Monteagudo-Sánchez A., Garin I., de Nanclares G.P. MD. The Use of Methylation-Sensitive Multiplex Ligation-Dependent Probe Amplification for Quantification of Imprinted Methylation. In: *CpG Islands. Methods in Molecular Biology Book Series (MIMB, Volume 1766)*. Humana Press, New York, NY; 2018:109-121. doi:https://doi.org/10.1007/978-1-4939-7768-0\_6

12. Garin I, Mantovani G, Aguirre U, et al. European guidance for the molecular diagnosis of pseudohypoparathyroidism not caused by point genetic variants at GNAS : an EQA study. *Eur J Hum Genet*. 2015;(May 2014):438-444. doi:10.1038/ejhg.2014.127

13. Temple IK, Shield JPH. 6q24 transient neonatal diabetes. *Rev Endocr Metab Disord*. 2010;(October):199-204. doi:10.1007/s11154-010-9150-4

14. Wakeling EL, Brioude F, Lokulo-sodipe O, et al. Diagnosis and management of Silver–Russell syndrome: first international consensus statement. *Nat Rev Endocrinol*. 2017;13:105-124. doi:10.1038/nrendo.2016.138

15. Brioude F, Kalish JM, Mussa A, et al. Clinical and molecular diagnosis, screening and management of Beckwith–Wiedemann syndrome: an international consensus statement. *Nat Rev Endocrinol*. 2018;14(4):229-249. doi:10.1038/nrendo.2017.166.Clinical

16. Perez-Nanclares G, Velayos T, Vela A, Muñoz-Torres M, Castaño L. Pseudohypoparathyroidism type Ib associated with novel duplications in the GNAS locus. *PLoS One*. 2015;10(2). doi:10.1371/journal.pone.0117691

17. Ferguson-Smith AC. Genomic imprinting: The emergence of an epigenetic paradigm. *Nat Rev Genet*. 2011;12(8):565-575. doi:10.1038/nrg3032

18. Tost J. Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns. In: Jeltsch A, Jurkowska R, eds. *DNA Methyltransferases - Role and Function. Advances in Experimental Medicine and Biology*. Vol 945. ; 2016:343-430. doi:https://doi.org/10.1007/978-3-319-43624-1\_15

19. Klobucar T, Kreibich E, Krueger F, et al. IMPLICON : an ultra-deep sequencing method to uncover DNA methylation at imprinted regions. 2020;48(16). doi:10.1093/nar/gkaa567

20. Barros-Silva D, Marques CJ, Henrique R, Jerónimo C. Profiling DNA methylation based on next-generation sequencing approaches: New insights and clinical applications. *Genes (Basel)*. 2018;9(9). doi:10.3390/genes9090429

21. Tee L, Lim DHK, Dias RP, et al. Epimutation profiling in Beckwith-Wiedemann syndrome: Relationship with assisted reproductive technology. *Clin Epigenetics*. 2013;5(1):1-10. doi:10.1186/1868-7083-5-23

22. Mackay DJG, Bliek J, Lombardi MP, et al. Discrepant molecular and clinical diagnoses in Beckwith-Wiedemann and Silver-Russell syndromes. *Genet Res (Camb)*. 2019:10-14. doi:10.1017/S001667231900003X

23. Monk D, Arnaud P, Apostolidou S, et al. Limited evolutionary conservation of imprinting in the human placenta. *Proc Natl Acad Sci U S A*. 2006;103(17):6623-6628. doi:10.1073/pnas.0511031103

24. Hattori H, Hiura H, Kitamura A, et al. Association of four imprinting disorders and ART. *Clin Epigenetics*. 2019;11(1):1-12. doi:10.1186/s13148-019-0623-3

25. Hiura H, Okae H, Chiba H, et al. Imprinting methylation errors in ART. *Reprod Med Biol*. 2014;13(4):193-202. doi:10.1007/s12522-014-0183-3

26. Rossignol S, Steunou V, Chalas C, et al. The epigenetic imprinting defect of patients with Beckwitn-Wiedemann syndrome born after assisted reproductive technology is not restricted to the 11p15 region. *J Med Genet*. 2006;43(12):902-907. doi:10.1136/jmg.2006.042135

27. Chen Z, Hagen DE, Elsik CG, et al. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *Proc Natl Acad Sci U S A*. 2015;112(15):4618-4623. doi:10.1073/pnas.1422088112

28. Mackay DJG, Callaway JLA, Marks SM, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet*. 2008;40(8):949-951. doi:10.1038/ng.187

29. Zheng W, Chen L, Dai J, et al. New biallelic mutations in PADI6 cause recurrent preimplantation embryonic arrest characterized by direct cleavage. *J Assist Reprod Genet*. 2020;37(1):205-212. doi:10.1007/s10815-019-01606-7

30. Cubellis MV, Pignata L, Verma A, et al. Loss-of-function maternal-effect mutations of PADI6 are associated with familial and sporadic Beckwith-Wiedemann syndrome with multi-locus imprinting disturbance. *Clin Epigenetics*. 2020;12(1):1-13. doi:10.1186/s13148-020-00925-2

31. Xu Y, Shi Y, Fu J, et al. Mutations in PADI6 Cause Female Infertility Characterized by Early Embryonic Arrest. *Am J Hum Genet*. 2016;99(3):744-752. doi:10.1016/j.ajhg.2016.06.024

32. Xu Y, Goodyer C, Deal C, Polychronakos C. Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochem Biophys Res Commun*. 1993;197(2):747-754.

33. Martin EM, Fry RC. Environmental Influences on the Epigenome : Exposure- Associated DNA Methylation in Human Populations. Annu Rev Public Health. 2018; 39:309-333

34. Nye MD, King KE, Darrah TH, et al. Maternal blood lead concentrations, DNA methylation of MEG3 DMR regulating the DLK1/MEG3 imprinted domain and early growth in a multiethnic cohort. *Environ Epigenetics*. 2016;2(1):1-8. doi:10.1093/eep/dvv009

35. Hoyo C, Daltveit AK, Iversen E, et al. Erythrocyte folate concentrations, CpG methylation at genomically imprinted domains, and birth weight in a multiethnic newborn cohort. *Epigenetics*. 2014;9(8):1120-1130. doi:10.4161/epi.29332

36. Waterland RA, Garza C. Potential mechanisms of metabolic imprinting that lead to chronic disease. *Am J Clin Nutr*. 1999;69(2):179-197. doi:10.1093/ajcn/69.2.179

37. Russo S, Calzari L, Mussa A et al. [A multi-method approach to the molecular diagnosis of overt and borderline 11p15.5 defects underlying Silver-Russell and Beckwith-Wiedemann syndromes.](https://pubmed.ncbi.nlm.nih.gov/26933465/) Clin Epigenetics. 2016 Mar 1;8:23

**TABLE AND FIGURES LEGENDS**

**Table 1:** **ImprintSeq design.** Imprinting DMRs covered, location, number of CpGs in Illumina Infinium Human Methylation 450 BeadChip array (450K), number of CpGs in ImprintSeq and mean coverage detected per DMR. P, paternal; M, maternal; ID, isoform dependent; gDMR, germline DMR. \* Individual CpGs covered by probes.

**Figure 1:** **Flow chart of the study.** A total of 223 samples were studied with our hybridization-based custom-sequencing panel to interrogate 63 imprinted DMRs (iDMRs) including 147 CID samples, 70 controls and 3 calibration points in replicates (0%, 50% and 100% methylation). After the extraction of raw DNA methylation per CpG, data was filtered by quality and coverage and corrected by MethylCal using standard controls. Two CID samples were excluded from the analysis during filtering. In order to compare the level of methylation per DMR between samples, we calculated median methylation level per DMR (MML) per sample. Afterward we determined the methylation profiling and 3SDs confidence interval of 63 iDMRs in the control cohort. Then we compared it with each case sample. Loss of methylation (LOM) was defined as MML below healthy controls’ MML 3SDs confidence interval and gain of methylation (GOM) as MML above healthy controls’ 3SDs confidence interval. To estimate the value of ImprintSeq as diagnostic assay, we compared results obtained with ImprintSeq in 8 CIDs-associated iDMRs and those from 131 CID cases with diagnostic report (by MS-MLPA and/or SNP array) . To determine the ability of ImprintSeq to detect MLIDs, first LOM/GOM detected with ImprintSeq in 63 iDMRs were divided in two groups based on the magnitude of the differential median methylation level, high/moderate methylation alterations (HMA/MMA). The results obtained were compared with Illumina Infinium HumanMethylation450 BeadChips (450K) results. Then, we applied diagnostic criteria for MLID proposed in this work.

**Figure 2: Diagnostic of congenital imprinting disorders with ImprintSeq.** Significant DNA methylation alterations across 8 imprinting DMRs (iDMRs) related with diagnostic of congenital imprinting disorders (CIDs-associated) (bold) and 14 additional iDMRs to detect uniparental disomies (UPD) in 131 samples with diagnostic report. In orange are represented gain-of-methylation detected by using a 3SDs confidence interval based on healthy controls’ median methylation level (MML) and in purple are represented loss-of-methylation identified with the same confidence interval. Samples are organized by disease/status and iDMRs are organized by chromosome position. Beckwith-Wiedemann syndrome (BWS), Silver-Russell syndrome (SRS), Pseudohypoparathyroidism 1B (PHP1b), Transient neonatal diabetes mellitus (TNDM), Angelman syndrome (AS), not alteration detected (NAD).

**Figure 3: Identification of LOM/GOM with ImprintSeq.** Significant DNA methylation alterations across 63 imprinting DMRs in 145 samples. In orange are represented gain-of-methylation detected by using a 3SDs confidence interval based on healthy controls’ median methylation level and in purple are represented loss-of-methylation identified with the same confidence interval. Samples are organized by disease and DMRs are organized by chromosome position. Beckwith-Wiedemann syndrome (BWS), Silver-Russell syndrome (SRS), Pseudohypoparathyroidism 1B (PHP1b), Transient neonatal diabetes mellitus (TNDM), Angelman syndrome (AS).

**Figure 4: Classification of LOM/GOM based on the magnitude of the differential median methylation level and detection of MLIDs.** Significant DNA methylation alterations across 63 imprinting DMRs in 145 samples. (a) High methylation alterations (HMA) was declared when the difference in absolute value between the median methylation level in the healthy controls and CIDs is greater than the estimated cut-off point (see Methods section). (b) Moderate methylation alterations (MMA) was declared when the difference in absolute value between median methylation level in control and the one detected in the case is lower than the estimated cut-off point. In orange are represented gain-of-methylation by using 3SDs confidence interval and in purple are represented loss-of-methylation identified with the same confidence interval. Samples are organized by disease and DMRs are organized by chromosome position. Beckwith-Wiedemann syndrome (BWS), Silver-Russell syndrome (SRS), Pseudohypoparathyroidism 1B (PHP1b), Transient neonatal diabetes mellitus (TNDM), Angelman syndrome (AS).