Imprinted loci may be more widespread in humans than previously appreciated and enable limited assignment of parental allelic transmissions in unrelated individuals

Gabriel Cuellar Partida[1], Charles Laurin[2], Susan M. Ring[2], Tom R. Gaunt[2], Caroline L. Relton[2], George Davey Smith[2], David M. Evans[1,2]

1 University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia

2 Medical Research Council (MRC) Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom

Key Words: Methylation, Imprinting, Parent of origin effects, GWAS, ALSPAC

Corresponding Author:
Dr Gabriel Cuellar Partida
The University of Queensland Diamantina Institute
Brisbane, Queensland
Australia
Tel: +61 7 3443 7051
Fax: +61 7 3443 6966
Email: g.cuellarpartida@uq.edu.au

Prof David M Evans
The University of Queensland Diamantina Institute
Brisbane, Queensland
Australia
Tel: +61 7 3443 7051
Fax: +61 7 3443 6966
Email: d.evans1@uq.edu.au

Genome-wide survey of parent-of-origin effects on DNA methylation identifies candidate imprinted loci in humans.

Abstract

Genomic imprinting is an epigenetic mechanism leading to parent-of-origin dependent gene expression. So far, the precise number of imprinted genes in humans is uncertain. In this study, we leveraged genome-wide DNA methylation in whole blood measured longitudinally at 3 time points (birth, childhood and adolescence) and GWAS data in 740 Mother-Child duos from the Avon Longitudinal Study of Parents and Children (ALSPAC) to systematically identify imprinted loci. We reasoned that *cis*-meQTLs at genomic regions that were imprinted would show strong evidence of parent-of-origin associations with DNA methylation, enabling the detection of imprinted regions. Using this approach, we identified genome-wide significant *cis*-meQTLs that exhibited parent-of-origin effects (POEs) at 35 novel and 50 known imprinted regions ($10^{-10} < P < 10^{-300}$). Among the novel loci, we observed signals near genes implicated in cardiovascular disease (*PCSK9*), and Alzheimer's disease (*CR1*), amongst others. Most of the significant regions exhibited imprinting patterns consistent with uniparental expression, with the exception of twelve loci (including the *IGF2, IGF1R,* and *IGF2R* genes), where we observed a bipolar-dominance pattern. POEs were remarkably consistent across time points and were so strong at some loci that methylation levels enabled good discrimination of parental transmissions at these and surrounding genomic regions. The implication is that parental allelic transmissions could be modelled at many imprinted (and linked) loci and hence POEs detected in GWAS of unrelated individuals given a combination of genetic and methylation data. Our results indicate that modelling POEs on DNA methylation is effective to identify loci that may be affected by imprinting.

Introduction

Genomic imprinting is an epigenetic mechanism in which genes are silenced in a parent-of-origin specific manner. The first experimental evidence for genomic imprinting was provided by investigations during the 1980s when researchers failed to produce viable mice embryos using only the paternal or maternal genome[1]. The precise evolutionary mechanisms that give rise to genomic imprinting are unknown. One hypothesis postulates that imprinting provides a mechanism through which maternal and paternal genomes exert counteracting growth effects during development with paternal genes encouraging growth and solicitation of maternal care, even at the expense of the mother's health, while maternal alleles are orientated toward success of all offspring, who do not necessarily share the same father[2]. There is some empirical evidence to support of this hypothesis. For example, in contrast to expression of the paternally derived insulin-like growth factor 2 (*IGF2*) gene which promotes cell proliferation, expression of the maternally derived *CDKN1C* and *PHLDA2* genes act as negative regulators of this process[3].

It is widely accepted that imprinted genes are regulated by *cis*-acting regulatory elements, called imprinting control elements (ICEs), which carry parental-specific epigenetic modifications such as DNA methylation[4]. DNA methylation mainly occurs at the C5 position of CpG dinucleotides and is known to influence transcription[4]. Promoter regions of imprinted genes are usually rich in CpG sites and within differentially methylated regions (DMRs); where the repressed allele is methylated and the active allele is unmethylated. Although typical imprinting of a region results in monoallelic expression of the paternal or maternal allele, studies have shown that loci can deviate from this canonical pattern and show differential expression in a parent-of-origin-dependent manner[5,6].

Multiple studies have shown that imprinted genes affect prenatal growth control, normal brain development and postnatal metabolism[7-10]. The monoallelic expression of imprinted loci produces genetic vulnerabilities that can lead to monogenic syndromes. In humans, abnormal imprinting patterns at specific loci can result in genetic disorders such as Beckwith-Wiedemann and Silver-Russell syndromes which primarily affect growth, and Angelman and Prader Willi syndromes which have marked effects on growth and behaviour[11]. Evidence is also growing that imprinted genes play a significant role in complex human traits. Early linkage studies found evidence that genomic imprinting was important in the genetic aetiology of mental disorders such as Alzheimer's and schizophrenia as well as type 2 diabetes and BMI[12-14]. More recently, large-scale genome-wide association studies (GWAS) have found a handful of SNPs in imprinted genes that exhibit parent of origin effects and are associated with traits including age at menarche, breast cancer, basal cell carcinoma or type 2 diabetes [15-18].

Given that genomic imprinting appears to play a role in the genetic aetiology of multiple complex phenotypes, identifying novel imprinted genes is of considerable interest. However, the extent to which genes exhibit imprinted expression throughout the human genome is unknown. The number of validated imprinted genes in humans lies somewhere between 40 and 100 according to reviews[19-21], while some databases such as geneimprint (http://www.geneimprint.com/) and the Otago imprinting database[22] list many more that have yet to be validated. Several methods have been used to identify imprinted loci, including analysis of differential expression between parthenogenotes and androgenotes in mice[23], bioinformatic approaches that look for novel imprinted loci based on genomic features found in known imprinted regions[24], and creating gene knockouts of paternal/maternal alleles in mice[25]. More recently, whole genome scans of imprinted regions have been performed using next-generation sequencing technologies to measure differential gene expression between maternally and paternally derived genes using RNA-seq[26-28] or to measure differential methylation with MethylC-Seq[29]. Although some of these more recent approaches have been applied to human genomes, the number of studies have been limited and constrained to small sample sizes[27,30,31], thus limiting the ability to reliably detect imprinted genes.

Imprinted regions in the human genome can also be detected using statistical approaches that model parent-of-origin effects (POEs) of genetic variants on DNA methylation and gene expression. In the presence of imprinting, SNPs affecting DNA methylation (mQTLs) or gene expression (eQTLs) have a different effect depending on their parental origin. In this work, we leverage genome-wide DNA methylation and genotypic data of up to 740 Mother-Child duos from the Avon Longitudinal Study of Parents and Children (ALSPAC) to systematically identify imprinted loci. We subsequently investigated the association between SNPs in these imprinted regions and a diverse range of phenotypes.

Methods

Data

Study sample

ALSPAC is a geographically based UK cohort that recruited pregnant women residing in Avon (South West England) with an expected date of delivery between 1 April 1991 and 31 December 1992. A total of 15 247 pregnancies were enrolled, with 14 775 children born[32,33]. Of these births, 14 701 children were alive at 12 months. Ethical approval was obtained from the ALSPAC Law and Ethics committee, and the local research ethics committees. Appropriate consent was obtained from the participants for genetic analysis. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary [http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/].

The data used in this study corresponds to the mother-child pairs from the ALSPAC cohort who took part in the Accessible Resource for Integrative Epigenomic Studies (ARIES, http://www.ariesepigenomics.org.uk/)[34,35]. We used genotypic data from 740 mother-child duos, and DNA methylation data from the 740 children. Each child had DNA methylation measured at three time points – i.e. cord blood, peripheral blood (whole blood, buffy coats, white blood cells or blood spots) during childhood (~7 years) and during adolescence (15, 17 years).

DNA Methylation

Description of the DNA methylation assays can be found elsewhere[7]. In brief, genome-wide methylation was measured using the Illumina Infinium HumanMethylation450 (HM450) arrays. These arrays were scanned using Illumina iScan and the initial quality review was done in GenomeStudio. A wide range of batch variables were measured for each sample during the data generation, including quality control (QC) metrics from the standard control probes on the array. Samples failing QC were not included in the analysis. Data points with a low signal:noise ratio (detection $p > 0.01$) or with methylated or unmethylated read counts of zero were also excluded from analysis. Genotype probes in the HM450 array of the same individual at different time points were used to identify and remove sample mismatches. DNA methylation at each CpG probe was normalised using the Touleimat and Tost algorithm implemented in the R package wateRmelon[36] to reduce the non-biological differences between probes. From the CpG probes passing QC, we selected those that have been shown to provide genuine measurements of DNA methylation as described in Naeem *et al.* [37], leaving 294,841 CpG probes for the analysis. Beta values (i.e. the proportion of methylation) were used for all the analyses.

Genotypes

Mother-child duos participating in ARIES were previously genotyped as part of a former ALSPAC study, the details of which can be found elsewhere[32,33,38]. Briefly, children were genotyped on Illumina HumanHap550 quad-chip platforms by the Wellcome Trust Sanger Institute (Cambridge, UK) and by the Laboratory Corporation of America (Burlington, USA) using support from 23andMe. Mothers were genotyped on Illumina HumanHap660W quad-chip platform by Centre National de Génotypage (Évry, FR). Standard QC was applied to SNPs and individuals. Individuals were excluded based on genotype rate (< 5%), sex mismatch, high heterozygosity, and cryptic relatedness (defined as identity by descent > 0.125). In order to remove individuals of non-European descent, principal components (PCs) were derived from LD-pruned SNPs with MAF>0.05 using plink[39]. Individuals laying 5 s.d. beyond the 1000 Genomes European population PCs 1 and 2 centroid were excluded. SNPs with a minor allele frequency (MAF) < 1%, genotyping rate <5%, or with a deviation from Hardy-Weinberg disequilibrium ($p < 1 \times 10^{-6}$) were removed from the analysis.

Genotype Imputation was performed by first phasing the genotypes using SHAPEIT V2[40], and then imputing to the 1000 Genomes European reference panel (phase 1, version 3) using Impute (v2.2.2)[41]. Genotypes were removed if they deviated from Hardy–Weinberg equilibrium $p < 5 \times 10^{-6}$, MAF <5% (the high threshold was to minimize the possibility of low frequency variants producing chance parent of origin effects through statistical fluctuation) or imputation info score <0.8. Best guess genotypes were used for subsequent analyses. The final imputed dataset used for the analyses presented here contained 2,158,724 SNPs.

Statistical analysis

Identifying transmission of the alleles

The crucial first step in identifying POEs is assigning alleles to their parental origin. In order to achieve this, we applied the duoHMM algorithm implemented in the software SHAPEIT V2[42] to the most likely imputed genotypes from the ALSPAC mothers and children. This algorithm leverages linkage disequilibrium (LD) and identity-by-descent (IBD) sharing in order to phase genotypes and resolve the parental origin of alleles at each SNP. Using a custom written *Perl* script, the phased genotypes were formatted in a way such that heterozygotes where the minor allele was inherited from the mother were coded as 1, homozygotes were coded as 0 and heterozygotes where the minor allele was transmitted by the father were coded as -1. In order to confirm the accuracy of our approach to resolve the transmission of the alleles, we compared the haplotypes of the mothers and children. We observed that for each of the children, the alleles of the haplotype inferred to be the one inherited from the mother, matched to those from the mother 99.9% of the time. We attribute the 0.1% of mismatches to genotyping or imputation errors in mothers or children.

Regression model

In order to identify SNPs in the genome displaying POEs on DNA methylation from the three time points (birth, childhood and adolescence), we employed a regression model[6,43] to estimate: the additive effect β_A, defined as the equal contribution of each minor allele to the phenotype; (ii) the dominance effect β_D which measures the deviation of the heterozygote from the mean phenotypic value of the two homozygotes; and the parent-of-origin effect β_P, which is the mean difference between heterozygotes (i.e. the heterozygote where allele "A" is paternally transmitted, and the heterozygote where allele "A" is maternally transmitted). In matrix annotation, with intercept term β_0, the mean phenotypic value for each possible genotype can be modelled as:

$$\boxed{\begin{bmatrix} AA \\ Aa \\ aA \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_A \\ \beta_D \\ \beta_P \end{bmatrix}}$$

This coding of genotypes enables testing for effects that are strictly due to parent-of-origin effects, as under Hardy-Weinberg equilibrium the parent-of-origin vectors are orthogonal to the additive and dominant effects.

Given that DNA methylation is affected by sex and age, these factors were incorporated into the model as covariates, along with the first 3 ancestry informative principal components (PCs) derived from genome-wide SNP genotypes in order to control for population stratification, as well as the first 10 PCs derived from the control matrix of the Illumina HumanMethylation450 assays to control for batch effects. The following model:

$$\boldsymbol{CpG} = \beta_0 + \beta_A \boldsymbol{A} + \beta_D \boldsymbol{D} + \beta_P \boldsymbol{P} + \sum_{i=1}^{\#cov} \beta_i cov_i + \epsilon$$

was fitted to the 294,841 DNA methylation CpG probes against SNPs within 500 Kb from the CpG probe (i.e. SNPs in *cis*). SNPs beyond 500kb from the CpG site were not assessed as it would have increased the multiple testing burden by 3 orders of magnitude and the number of individuals in this study may not yield enough power to detect reliable associations of SNPs in *trans*[34]. In this model, CpG is the column vector of DNA methylation values of a CpG probe; β_0 is the intercept; β_A the regression coefficient of the SNP additive effect; $\boldsymbol{A}$ is the vector of genotypes in additive coding; β_D the regression coefficient of the SNP dominance effect; $\boldsymbol{D}$ is the vector of genotypes in dominance coding; $\beta_P$is the regression coefficient of the SNP parent-of-origin effect; ***P*** is the vector of genotypes in parent-of-origin coding; β_i the regression coefficient of the covariates; and cov_i are the covariates specified above.

Given that DNA methylation values suffer from heteroscedasticity, White-Huber standard errors[44] were computed to estimate the significance of the POE term β_P using the *sandwich* package in R.

Significance Threshold

In total, ~400M statistical tests were performed. Given that neighbouring SNPs usually display a high degree of correlation between each other due to LD, the number of independent tests was empirically estimated using a matrix spectral decomposition algorithm of the correlation matrix[45,46]. We applied this algorithm in 100 randomly selected autosomal genomic regions of 1Mb each and observed that the number of independent SNPs was 0.33 times (95% CI: 0.28, 0.38) the number of SNPs tested. Hence the effective number of tests was set to 132M and the Bonferroni significance threshold was set at P-value < 3.7e-10. We note however, that this threshold may still be conservative as the correlation between CpG probes has not been taken into account.

Functional analyses

We performed a gene-set enrichment analysis using the WEB-based Gene SeT AnaLysis Toolkit to investigate if the identified loci implicated particular biological pathways, phenotypes and diseases. We created the gene list by mapping each CpG probe displaying a statistically significant POE to its physically closest gene. We opted for this approach over mapping genes based on whether the SNP exerting the POE on the CpG was reported to be an eQTL of a certain gene in the GTEx database, as these eQTLs are based on additive effects. Moreover, we observed that by mapping genes based on shortest distance, many of them corresponded to known imprinted genes. Gene enrichment analysis was assessed against the Gene Ontology[47], DisGeNET[48] and GLAD4U[49] databases.

With a similar intent, we used Phenoscanner[50] to identify if the SNPs exerting statistically significant POEs on CpG have been implicated in other phenotypes.

Predicting parental transmission in heterozygote individuals using methylation status

During this project we observed that DNA methylation at some CpG sites could potentially be used to infer the parental transmission in heterozygote individuals of samples without parental genotypes. Under a uniparental expression pattern of imprinting, one of the parental alleles remains inactive leading to the phenotypic mean of one of the heterozygote groups (e.g. minor allele inherited by the mother) being equal to the mean of the minor allele homozygote, while the phenotypic mean of the other heterozygote group (e.g. minor allele inherited by the father) is equal to the mean of the major allele homozygote. With this premise, we fitted a logistic model to the homozygous individuals for each of the statistically significant SNPs found in this study:

$$logit(\boldsymbol{H}) = \beta_0 + \beta \boldsymbol{CpG} + \varepsilon$$

Where ***H*** is a vector with labels 0 for minor allele homozygotes and 1 for major allele homozygotes and ***CpG*** is the DNA methylation at the relevant CpG site.

We then used this fitted logistic model to predict the pattern of allelic transmission for each heterozygote individual at the putatively imprinted SNPs. Note that this approach can also predict the allelic transmission at other patterns of imprinting (e.g. bipolar or polar dominance) as it splits heterozygote individuals into those that are above the phenotypic mean of the (e.g. minor allele) homozygous individuals and those that are below the phenotypic mean of the (e.g. major allele) homozygous individuals. To measure how well this method performed, we computed the Area Under the receiver operating Characteristic curve (AUC) for each SNP.

We estimated the sample size that would be required to achieve 80% statistical power to detect POEs using this approach to infer parental transmission compared to having actual parental genotypes and being able to identify each heterozygote group correctly (as was the case in this study). We simulated 500 runs for each SNP where POEs explained: 0.5%, 1%, 2%, 4% and 9% of the variance (R^2) using known parent-of-origin coded genotypes (i.e. 0 for homozygotes, and -1 or 1 for each of the heterozygote groups, AUC=1). We then estimated how the variance explained degraded when using the inferred genotypes coded as 0 for homozygotes and as an expected dosage for heterozygotes: p - (1-p), where p is the probability of being in heterozygous group 1 and 1-p the probability of being in heterozygous group 2. For example, when we simulated a POE using the known parent-of-origin coded genotypes (i.e. AUC=1) that explained R^2=1%, the variance explained would drop to R^2=0.09% when using the inferred (AUC=0.75) parent-of-origin coded genotypes (as expected, R^2 would normally degrade relative to AUC and MAF). We then used the function pwr.r.test from the "pwr" package in R that implements a Z' transformation of the correlation[51] to derive the sample size required to achieve 80% power with α=0.0005.

Results

Identification of imprinted DMRs

We identified 365 CpG sites with at least one SNP exerting POEs with a p-value less than our Bonferroni significance threshold of $3.7\text{x}10^{-10}$ [Supplementary table 1]. These CpG sites were distributed among 85 loci, each of which was defined to be at least 2Mb apart from one another [Figure 1 and Table 1]. By inspecting RefSeq[52], geneimprint (http://www.geneimprint.com/) and Otago imprinting[22] (http://igc.otago.ac.nz) databases and the literature[21,53-61], we identified 150 known imprinted loci (each defined to be at least 2Mb in each direction from one another) [Supplementary Table 2]. Of the 85 loci we identified at genome-wide significant levels, 50 mapped to these known imprinted regions, while 35 appeared to be novel. Distance between each identified locus and the closest known imprinted locus is included in Supplementary Table 3. The POEs identified were remarkably consistent across the different time points (i.e., birth, childhood and adolescence), with 70 loci identified as statistically significant at at least two time points. All the remaining loci with the exception of the *FAM30A* locus (which showed a uniparental expression pattern) showed at least a nominally significant parent of origin p-value (<0.05) between the SNP and methylation at the relevant CpG site at all 3 time points [Table 1].

The strongest POEs were observed within known imprinted loci. For instance, we observed partial correlations (R) as high as 0.90 between parent-of-origin coded SNPs and CpG sites near the *NAP1L5* and *GNAS* genes. For the novel candidate imprinted loci we observed partial correlations as high as $R=0.73$ for a CpG near *MAP2*. In Supplementary Tables 4-6, we have included the summary statistics of each CpG site with at least one significant SNP at each of the different time points along with additive and dominance effect statistics.

We found diverse patterns of imprinting where the effect depended on the combination of the alleles [Figure 2]. For example the distribution of DNA methylation at the CpG probe cg06982169 near the known imprinted genes *H19* and *IGF2* displayed a bipolar dominance[6] pattern where the phenotypic value of the two homozygotes did not differ, and one of the heterozygotes had a larger phenotypic value than the two homozygotes and the other heterozygote had a smaller value [Figure 2a]. This type of pattern was also observed for loci containing the *FAR2P1, NAP1L5, FAM50B, HYMAI, IGF2R, AIRN, KCNQ1OT1, IGF1R, DDX11* and *GNAS* genes. Most of the loci identified displayed a DNA methylation distribution consistent with uniparental effects, where one of the alleles led to a larger average phenotypic value than the other and one of the chromosomes was putatively silenced. Figure 2b shows an example of this methylation pattern, where the mean DNA methylation of the CpG probe cg09336323 near *MAP2* increases only if the minor allele "T" is inherited from the father. We observed an instance of polar underdominance at a CpG site near *LOC100506082* where one of the heterozygous groups displayed a lower phenotypic mean than the other three genotype groups (*P-value* = 1.05E-11) [Figure 2c]. Finally, the locus containing *ZNF331* showed an uncharacterized distribution of phenotypes amongst the four possible genotypes (*P-value* = 1.46E-105) [Figure 2d].

Functional analyses

To shed light on the potential function of genes within the identified loci, we performed gene enrichment analyses based on the gene ontology database[47]. Not surprisingly, we found a statistically significant overrepresentation of genes involved in insulin receptor binding (*P* = 3.63E-06; *FDR* = 6.64E-03) and imprinting (*P* = 2.76E-06; *FDR* = 2.36E-02). Beside these associations, we did not observe any other pathway with an *FDR* < 0.05; however, the top results included pathways involved in mitotic division and hormone secretion [Table 2]. Similarly, we assessed whether the genes tended to be involved in certain diseases. We found several diseases with an FDR < 0.05, most of which related to early growth or the nervous system, including neurodegenerative diseases such as Alzheimer's and Parkinson's disease, and syndromic disorders known to be due to imprinting aberrations, such as Beckwith–Wiedemann and Prader-Willi syndromes. These findings are summarized in Table 3.

We also investigated whether SNPs that showed parent of origin dependent associations with methylation had also been previously associated with other phenotypes (i.e. not necessarily in a parent of origin fashion). Using Phenoscanner[50], we identified 19 SNPs that reached genome-wide significance for POEs in our study and were also associated with one or more phenotypes ($p < 5 \times 10^{-6}$) including height, BMI, schizophrenia, fasting glucose, type 1 and type 2 diabetes **[Table 4]**.

Using methylation to determine allelic transmissions

Given that many of the loci showing parent of origin effects were associated very strongly with patterns of methylation, we were interested in the extent to which patterns of methylation might be used to determine parental transmissions in heterozygous individuals. We examined the performance of a simple statistical approach to determining transmissions at loci showing evidence of imprinting through first modelling the methylation levels of homozygous individuals, and then using this information to estimate the transmission status of each heterozygous individual (see methods). Supplementary Table 7 displays the accuracy by which the heterozygous genotypes groups could be inferred using methylation levels at the single most strongly associated CpG site at each locus. The most accurate discrimination between heterozygote groups using DNA methylation was observed for the SNP rs11699704 near *GNAS-AS1* using DNA methylation at the CpG site cg03606258, achieving an AUC = 0.92. For the 85 loci identified in this study, the median accuracy was AUC = 0.73 (interquartile range: 0.68 – 0.79).

Although for the majority of loci, the parental origin of alleles is difficult to determine with appreciable accuracy using DNA methylation alone, it may be the case that given very large numbers of individuals, it may still be possible to detect POEs in a large GWAS study of unrelated individuals when EWAS data is also present. In Supplementary Table 7 we show the sample size that would be required to achieve 80% power to detect POEs at candidate loci ($\alpha = 0.0005$). The sample size required increased with lower AUC and lower MAF. For example, on average, a SNP inferred with an AUC ~ 0.75 and a MAF ~ 0.25 required a sample size 12x larger than if the SNP was inferred with perfect discrimination (AUC = 1). For more common SNPs (MAF > 0.4) and AUC ~ 0.75 the required sample would be 5x larger.

Discussion

Summary of imprinted loci

In this work, we presented a genome-wide scan of SNPs' POEs on DNA methylation from peripheral blood at multiple time points. We found that most of the POEs of SNPs on DNA methylation are constant throughout birth, childhood and adolescence. This observation is consistent with previous studies, which showed that although patterns of DNA methylation at many CpG sites in peripheral blood cells are not stable over time, the additive genetic effects of SNPs on methylation appear to be remarkably consistent longitudinally[34]. We also showed that investigating POEs on DNA methylation is a powerful method of identifying regions of the genome likely to be affected by genomic imprinting. This assertion is supported by the fact that most statistically significant associations in our study corresponded to known imprinted loci and that the associations were with genetic variants in *cis* – i.e. it is unlikely that *cis* effects at genes are a product of maternal or paternal effects on children's DNA methylation, as we expect that maternal/paternal effects to be distributed evenly over the genome and hence much more likely to be a *trans* effect than a *cis* effect.

In addition to supporting the existence of multiple imprinted genes that have yet to be characterized, we also found several instances of diverse imprinting patterns. In particular, a bipolar dominance pattern was observed among CpG sites near the insulin like growth factors and receptors *IGF1R*, *IGF2R* and *IGF2*, all of which are located on different chromosomes and are consistent with the hypothesis that maternal and paternal alleles are antagonistic with respect to growth[2]. Bipolar dominance patterns have been observed previously [6,15] and are hypothesized to occur when differentially imprinted genes are in tight linkage disequilibrium but exert opposing effects on the phenotype **[Figure 3]**. There were also other genes nearby CpG sites that exhibited bipolar dominance POEs patterns including *GNAS* which has been previously described to encode maternal, paternal and biallelic derived proteins[62].

In our analyses, we identified 50 loci within the 150 known imprinted loci (summarized in Supplementary Table 2) and 35 outside these regions, deemed novel. The fact that we did not detect all known imprinted loci could be for various reasons, including lack of statistical power, poor coverage of CpG sites in the HM450 array, or the fact that imprinted expression is not maintained in all cell types[30], and therefore we could not detect it in peripheral blood.

The strongest POEs that we identified outside known imprinted regions was on a CpG site close to the *Microtubule-Associated Protein 2* (*MAP2*) gene which plays an essential role in neurogenesis[52]. Genes located near CpGs where we also detected strong POEs included the *DEAD/H-Box Helicase 11* (*DDX11*) which is involved in rRNA transcription and plays a role in embryonic development[52,63], the *CCT6P4* pseudogene, and *QKI*, which is an RNA-binding protein that plays a key role in myelation[64]. Other interesting genes included *MOBP*, also involved in myelination, *CR1* which mediates cellular binding to particles and immune complexes that have activated complement, and *PCSK9*, an important gene in the metabolism of plasma cholesterol[65].

The identification of new imprinted genes is important in furthering our understanding of the role of imprinting in normal human development and disease. The gene enrichment analyses carried out pointed to multiple metabolic, developmental and mental disorders. Alzheimer's disease was among the top results, and was partially driven by the novel identified loci *MOBP* and *SPON1* with limited evidence in the literature to actually play a role in the disease's aetiology[66,67] and *CR1*, one of the most important risk genes[68]. The SNP lookup that we performed using Phenoscanner also highlighted the potential role of the candidate imprinted loci in growth, metabolism and mental health.

Assigning Allelic Transmissions in Unrelated Individuals

We were able to assign allelic transmissions at heterozygous individuals with moderate confidence at a handful of loci. For the remaining loci however, our predictive ability appeared to be very limited. Because of the presence of winner's curse, these figures are likely to represent an upper limit on the predictive ability of simple approaches to resolve allelic transmission. Nevertheless, our simulations indicate that in principle POEs could be detected even if allelic transmission can't be determined with certainty given large enough numbers of individuals with both EWAS and GWAS. Whilst there are no cohorts of this size that have this kind of information currently, it is possible that in the future as the cost of microarrays decrease, that these sorts of studies might be feasible, particularly in large scale population based cohorts like the UK Biobank where GWAS is already available[69]. Alternatively it may be possible to achieve enough power by combining cohorts with both GWAS and EWAS in a meta-analysis, as is currently being done as part of the Genetics of DNA Methylation Consortium (GoDMC). We note also that whilst we have performed power calculations using information of a single CpG per SNP, it is likely that power to detect POEs could be increased further by incorporating information from adjacent correlated CpG sites and SNPs in imperfect linkage disequilibrium.

Strengths and Limitations

To our knowledge, this is the first study to use SNP's POEs on human whole-genome DNA methylation to identify imprinted regions in the genome. With recent technological advances leading to decreasing sequencing costs, the current gold standard approach to identify imprinted genes is through RNA-seq – where it is possible to quantify the expression of heterozygote alleles[28,31,70]. However, this approach is still not cost effective as it is gene expression- and SNP-dependent; thus, imprinted genes with tissue-specific expression or lacking a heterozygous exonic SNP would be missed in the very small sample sizes that are common in these studies. Also, these studies usually require the genotyping or sequencing of parent-child trios in order to map the transmission of the alleles. In contrast, our approach uses large scale array data on SNPs and methylation to infer the transmission of the alleles even in absence of one parental genome. Our large sample size also provided us with greater statistical power to detect these imprinted regions.

Our approach however, does have its weaknesses. First, we were unable to directly assess whether the identified POEs directly affect the genes mentioned in this study. This is particularly problematic for the novel candidate imprinted loci where there is no prior functional work to back up our assertion. Unfortunately we could not use a more robust approach such as using eQTL information to map gene targets[71], as eQTLs databases are based on SNPs additive effects and not POEs[72]. Therefore, the fact that we mapped genes based on their distance to the CpG sites may have impacted our gene enrichment analyses.

The other important limitation is that we were not able to distinguish whether the allele inherited from the father or the mother is active or inactive (i.e. whether the maternal or paternal gene is silenced) as the POEs are relative, and DNA methylation seldom has a baseline of zero. For instance, taking Figure 2b as an example, we cannot distinguish between whether the DNA methylation baseline is ~0.65 and the maternally inherited minor allele increases DNA methylation while the paternally derived allele remains inactive or vice versa.

The Phenoscanner lookup had the limitation that all the GWAS we looked into report additive effects, so although we observed significant associations based on additive effects, we could not assess whether these SNPs also exert POEs on these phenotypes.

Conclusion

In conclusion, we report 35 novel genomic loci that exhibit parent of origin effects and consequently may be imprinted. We also show that the pattern of association at these loci remains stable from birth to adolescence. Although our approach does not replace traditional methods to detect genes subjected to imprinting, it is a convenient and cost-effective way to narrow down the search space and prioritize candidates. Consistent with what it is known about the biological role of imprinting, many of the identified loci were within or nearby genes with known effects on traits related to growth, development and behavior.

Acknowledgments

This work was supported by NHMRC Project Grant (APP1085130 to D.M.E) and a Medical Research Council program grant (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/4 and MC_UU_12013/8). The UK Medical Research Council and the Wellcome Trust (Grant refs: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. D.M.E is supported by an Australian Research Council Future Fellowship (FT130101709). Methylation data in the ALSPAC cohort were generated as part of the UK BBSRC funded (BB/I025751/1 and BB/I025263/1) Accessible Resource for Integrated Epigenomic Studies (ARIES, http://www.ariesepigenomics.org.uk). GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. This publication is the work of the authors and G.C.P will serve as guarantor for the contents of this paper.

1. Barton, S.C., Surani, M.A. & Norris, M.L. Role of paternal and maternal genomes in mouse development. *Nature* **311**, 374-6 (1984).
2. Haig, D. & Graham, C. Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell* **64**, 1045-6 (1991).
3. Cordeiro, A., Neto, A.P., Carvalho, F., Ramalho, C. & Doria, S. Relevance of genomic imprinting in intrauterine human growth expression of CDKN1C, H19, IGF2, KCNQ1 and PHLDA2 imprinted genes. *J Assist Reprod Genet* **31**, 1361-8 (2014).
4. Barlow, D.P. Genomic imprinting: a mammalian epigenetic discovery model. *Annu Rev Genet* **45**, 379-403 (2011).
5. Cockett, N.E. *et al.* Polar overdominance at the ovine callipyge locus. *Science* **273**, 236-8 (1996).
6. Wolf, J.B., Cheverud, J.M., Roseman, C. & Hager, R. Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet* **4**, e1000091 (2008).
7. Girardot, M., Feil, R. & Lleres, D. Epigenetic deregulation of genomic imprinting in humans: causal mechanisms and clinical implications. *Epigenomics* **5**, 715-28 (2013).
8. John, R.M. Imprinted genes and the regulation of placental endocrine function: Pregnancy and beyond. *Placenta* (2017).
9. Perez, J.D., Rubinstein, N.D. & Dulac, C. New Perspectives on Genomic Imprinting, an Essential and Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain. *Annu Rev Neurosci* **39**, 347-84 (2016).
10. Bartolomei, M.S. & Ferguson-Smith, A.C. Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* **3**(2011).
11. Reik, W. Genomic imprinting and genetic disorders in man. *Trends Genet* **5**, 331-6 (1989).
12. Bassett, S.S., Avramopoulos, D. & Fallin, D. Evidence for parent of origin effect in late-onset Alzheimer disease. *Am J Med Genet* **114**, 679-86 (2002).
13. Francks, C. *et al.* Parent-of-origin effects on handedness and schizophrenia susceptibility on chromosome 2p12-q11. *Hum Mol Genet* **12**, 3225-30 (2003).
14. Lindsay, R.S., Kobes, S., Knowler, W.C., Bennett, P.H. & Hanson, R.L. Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians. *Diabetes* **50**, 2850-7 (2001).
15. Hoggart, C.J. *et al.* Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. *PLoS Genet* **10**, e1004508 (2014).
16. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868-74 (2009).
17. Perry, J.R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-7 (2014).
18. Day, F.R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* (2017).
19. Kim, J., Bretz, C.L. & Lee, S. Epigenetic instability of imprinted genes in human cancers. *Nucleic Acids Res* **43**, 10689-99 (2015).
20. Morison, I.M., Ramsay, J.P. & Spencer, H.G. A census of mammalian imprinting. *Trends Genet* **21**, 457-65 (2005).
21. Morcos, L. *et al.* Genome-wide assessment of imprinted expression in human cells. *Genome Biol* **12**, R25 (2011).
22. Morison, I.M., Paton, C.J. & Cleverley, S.D. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* **29**, 275-6 (2001).
23. Ruf, N. *et al.* Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. *Genomics* **87**, 509-19 (2006).

24. Luedi, P.P., Hartemink, A.J. & Jirtle, R.L. Genome-wide prediction of imprinted murine genes. *Genome Res* **15**, 875-84 (2005).
25. Shiura, H. *et al.* Paternal deletion of Meg1/Grb10 DMR causes maternalization of the Meg1/Grb10 cluster in mouse proximal Chromosome 11 leading to severe pre- and postnatal growth retardation. *Hum Mol Genet* **18**, 1424-38 (2009).
26. Henckel, A. & Arnaud, P. Genome-wide identification of new imprinted genes. *Brief Funct Genomics* **9**, 304-14 (2010).
27. Okae, H. *et al.* Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Hum Mol Genet* **21**, 548-58 (2012).
28. Wang, X. & Clark, A.G. Using next-generation RNA sequencing to identify imprinted genes. *Heredity (Edinb)* **113**, 156-66 (2014).
29. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816-31 (2012).
30. Baran, Y. *et al.* The landscape of genomic imprinting across diverse adult human tissues. *Genome Res* **25**, 927-36 (2015).
31. Metsalu, T. *et al.* Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. *Epigenetics* **9**, 1397-409 (2014).
32. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
33. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* **42**, 97-110 (2013).
34. Gaunt, T.R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* **17**, 61 (2016).
35. Relton, C.L. *et al.* Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol* **44**, 1181-90 (2015).
36. Touleimat, N. & Tost, J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* **4**, 325-41 (2012).
37. Naeem, H. *et al.* Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* **15**, 51 (2014).
38. Kemp, J.P. *et al.* Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet* **10**, e1004423 (2014).
39. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
40. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2011).
41. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
42. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10**, e1004234 (2014).
43. Mantey, C., Brockmann, G.A., Kalm, E. & Reinsch, N. Mapping and exclusion mapping of genomic imprinting effects in mouse F2 families. *J Hered* **96**, 329-38 (2005).
44. White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *.Econometrica*, 817-838 (1980).
45. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-7 (2005).
46. Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**, 765-9 (2004).

47. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
48. Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
49. Jourquin, J., Duncan, D., Shi, Z. & Zhang, B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* **13 Suppl 8**, S20 (2012).
50. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209 (2016).
51. Cohen, J. Statistical Power Analysis for the Behavioral Sciences (1988).
52. O'Leary, N.A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-45 (2016).
53. Court, F. *et al.* 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 (2014).
54. Daelemans, C. *et al.* High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genet* **11**, 25 (2010).
55. Do, C. *et al.* Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am J Hum Genet* **98**, 934-55 (2016).
56. Docherty, L.E. *et al.* 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 (2014).
57. Nakabayashi, K. *et al.* Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. *Hum Mol Genet* **20**, 3188-97 (2011).
58. Rochtus, A. *et al.* Genome-wide DNA methylation analysis of pseudohypoparathyroidism patients with GNAS imprinting defects. *Clin Epigenetics* **8**, 10 (2016).
59. Smeester, L. *et al.* Imprinted genes and the environment: links to the toxic metals arsenic, cadmium, lead and mercury. *Genes (Basel)* **5**, 477-96 (2014).
60. Yuen, R.K., Jiang, R., Penaherrera, M.S., McFadden, D.E. & Robinson, W.P. Genome-wide mapping of imprinted differentially methylated regions by DNA methylation profiling of human placentas from triploidies. *Epigenetics Chromatin* **4**, 10 (2011).
61. Sanchez-Delgado, M. *et al.* Human Oocyte-Derived Methylation Differences Persist in the Placenta Revealing Widespread Transient Imprinting. *PLoS Genet* **12**, e1006427 (2016).
62. Hayward, B.E., Moran, V., Strain, L. & Bonthron, D.T. Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc Natl Acad Sci U S A* **95**, 15475-80 (1998).
63. Sun, X. *et al.* The Warsaw breakage syndrome-related protein DDX11 is required for ribosomal RNA synthesis and embryonic development. *Hum Mol Genet* **24**, 4901-15 (2015).
64. Aberg, K., Saetre, P., Jareborg, N. & Jazin, E. Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia. *Proc Natl Acad Sci U S A* **103**, 7482-7 (2006).
65. Poirier, S. *et al.* The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2. *J Biol Chem* **283**, 2363-72 (2008).
66. Liu, Q.Y. *et al.* An exploratory study on STX6, MOBP, MAPT, and EIF2AK3 and late-onset Alzheimer's disease. *Neurobiol Aging* **34**, 1519 e13-7 (2013).
67. Sherva, R. *et al.* Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement* **10**, 45-52 (2014).
68. Zhu, X.C. *et al.* CR1 in Alzheimer's disease. *Mol Neurobiol* **51**, 753-65 (2015).
69. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
70. DeVeale, B., van der Kooy, D. & Babak, T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet* **8**, e1002600 (2012).

71. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
72. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).

Table 1. Summary of Loci identified. For each CpG site meeting experiment-wide significance, we show the SNP that produced the strongest pvalue for the POE term. If more than 1 CpG site was located at the same locus (based on distance), the one with the smallest P-value is shown. Minor alleles (MAF < 50%) were used as effect alleles (EA) while the major alleles were set to non-effect alleles (NEA). Effects are summarized as partial correlations (R) between the POE coding and methylation beta value at the CpG site. Parent-of-origin genotype coding was defined as -1 for heterozygotes where the minor allele was inherited from the father, 0 for homozygotes and 1 for heterozygotes where the minor allele was inherited from the mother. The gene reported is the one that is closest to the CpG site's position. P-values for the POE between the CpG and the SNP are shown for each time point. Novel loci are displayed in bold. In POE patterns "U" refers to a uniparental effect, "B" refers to a bipolar pattern, "P" to a polar underdominance pattern, and "?" to an uncharacterized pattern. A definition of the different types of methylation pattern is illustrated in Figure 2.

Loci	Gene	Chr	CpG BP	CpG ID	Best SNP	EA	NEA	MAF	R	Birth P	Child P	Adol P	Loci
1	*DRAXIN*	1	11783294	cg18285337	rs4845874	G	A	0.42	0.36	2.90E-20	1.26E-20	1.32E-22	U
2	***FAM231C***	**1**	**17053886**	**cg12648811**	**rs1977269**	**A**	**C**	**0.28**	**0.23**	**7.70E-09**	**3.15E-10**	**5.82E-10**	**U**
3	***PCSK9***	**1**	**55522104**	**cg13462158**	**rs2479418**	**G**	**A**	**0.46**	**0.42**	**1.03E-25**	**4.46E-29**	**1.94E-19**	**U**
4	*DIRAS3*	1	68516472	cg16682227	rs1430754	T	C	0.15	-0.4	1.71E-14	5.68E-17	6.57E-18	U
5	*MIR488*	1	177001903	cg18865685	rs16850689	T	C	0.16	-0.31	2.68E-09	3.02E-14	8.79E-15	U
6	***CR1***	**1**	**207669922**	**cg10021878**	**rs1830762**	**T**	**C**	**0.34**	**-0.3**	**3.54E-08**	**6.94E-17**	**1.73E-11**	**U**
	CR1L	**1**	**207843084**	**cg21106486**	**rs11118410**	**G**	**A**	**0.4**	**-0.28**	**8.10E-12**	**2.42E-15**	**7.91E-10**	**U**
7	***PGBD5***	**1**	**230468611**	**cg15363333**	**rs7414930**	**T**	**G**	**0.5**	**0.24**	**2.89E-03**	**3.97E-11**	**2.88E-07**	**U**
8	*LINC01115*	2	863946	cg01854967	rs4561699	A	G	0.25	-0.38	1.53E-13	6.72E-21	9.68E-19	U
9	***RAB11FIP5***	**2**	**73384389**	**cg01422370**	**rs6760964**	**G**	**C**	**0.38**	**-0.28**	**3.31E-11**	**3.82E-14**	**3.95E-13**	**U**
10	*REG3G*	2	79220881	cg14005019	rs283842	A	C	0.31	0.4	1.49E-14	7.01E-28	4.98E-17	U
11	***PLGLB1***	**2**	**87238274**	**cg02778467**	**rs4971967**	**A**	**G**	**0.46**	**0.26**	**5.51E-11**	**3.27E-12**	**3.02E-09**	**U**
	MIR4771-2	**2**	**87569142**	**cg01908508**	**rs4971967**	**A**	**G**	**0.46**	**-0.26**	**1.58E-10**	**2.01E-13**	**3.47E-12**	**U**
12	***SFT2D3***	**2**	**128453335**	**cg03738707**	**rs11681053**	**C**	**T**	**0.13**	**-0.23**	**4.58E-08**	**5.33E-09**	**2.11E-10**	**U**
13	***FAR2P1***	**2**	**130795713**	**cg00271807**	**rs13426973**	**C**	**T**	**0.41**	**0.36**	**2.55E-18**	**3.50E-21**	**8.20E-17**	**B**
14	***GPR39***	**2**	**133402827**	**cg07916022**	**rs3738842**	**A**	**G**	**0.48**	**0.32**	**1.07E-11**	**3.15E-19**	**1.31E-11**	**U**
15	***MAP2***	**2**	**210074276**	**cg09336323**	**rs10932287**	**T**	**C**	**0.49**	**-0.73**	**2.01E-118**	**6.25E-155**	**2.59E-146**	**U**
16	***MOBP***	**3**	**39543515**	**cg03054684**	**rs561543**	**A**	**G**	**0.18**	**-0.26**	**1.83E-06**	**1.58E-08**	**8.20E-11**	**U**
17	*FGF12-AS3*	3	192289245	cg17611045	rs10460805	C	T	0.41	-0.31	2.11E-12	8.71E-13	1.96E-18	U
18	*NAP1L5*	4	89619051	cg19151808	rs10428273	A	G	0.09	-0.92	0.00E+00	0	0	B
19	***GIMD1***	**4**	**107446698**	**cg20025135**	**rs5017898**	**C**	**G**	**0.17**	**0.36**	**6.08E-10**	**6.80E-12**	**2.35E-09**	**U**
20	*TRPC3*	4	122854405	cg16501140	rs13121031	C	G	0.12	0.26	1.84E-04	1.46E-11	0.000355	U
21	*AHRR*	5	421733	cg00976097	rs2672724	T	C	0.39	-0.25	1.77E-11	2.59E-08	3.92E-06	U
	SDHAP3	5	1594579	cg08778598	rs7734561	G	A	0.21	0.32	2.39E-15	8.72E-17	1.80E-15	U
22	***LOC105374727***	**5**	**37209440**	**cg00331501**	**rs11743146**	**A**	**C**	**0.12**	**0.25**	**9.08E-11**	**7.85E-13**	**2.77E-07**	**U**
23	*NUDT12*	5	102898648	cg09166085	rs7730302	T	C	0.16	0.42	2.84E-20	1.12E-20	5.00E-21	U
24	*RASGEF1C*	5	179588440	cg08453205	rs10078657	A	T	0.29	0.23	8.64E-09	5.79E-12	2.28E-07	U
	BTNL9	5	180487084	cg07774765	rs10054109	T	C	0.5	-0.26	4.01E-10	4.19E-14	2.77E-07	U
25	*FAM50B*	6	3849327	cg25195497	rs2239713	T	C	0.33	0.53	5.81E-36	4.55E-73	2.00E-73	B
26	*PLAGL1*	6	144329672	cg21526238	rs11155342	A	G	0.09	-0.68	4.01E-33	4.84E-50	5.11E-39	U
	HYMAI	6	144329732	cg21952820	rs6937531	T	G	0.27	-0.67	5.98E-62	1.92E-82	4.79E-69	B
27	*IGF2R*	6	160427501	cg08350488	rs8191738	A	G	0.08	-0.26	7.59E-34	2.57E-29	1.99E-17	B
28	***LOC102724152***	**6**	**164461074**	**cg19287610**	**rs7765982**	**T**	**C**	**0.46**	**-0.44**	**8.57E-23**	**1.71E-41**	**4.99E-38**	**U**
29	*WDR27*	6	170054730	cg19089141	rs3823464	A	G	0.19	-0.51	1.59E-37	7.08E-40	3.89E-39	U
30	*HECW1*	7	43151725	cg06096382	rs10226468	C	T	0.48	-0.21	5.39E-04	2.15E-10	6.70E-11	U
31	*GRB10*	7	50849639	cg09150232	rs6976501	G	A	0.22	-0.32	5.56E-01	1.19E-10	4.19E-22	U
32	***CCT6P3***	**7**	**64541387**	**cg04775059**	**rs10949962**	**G**	**T**	**0.42**	**0.49**	**2.62E-39**	**1.90E-43**	**2.09E-41**	**U**
	LOC441242	**7**	**65235879**	**cg00634984**	**rs2418470**	**G**	**A**	**0.42**	**0.41**	**1.98E-27**	**6.67E-22**	**1.33E-22**	**U**
33	*UPK3B*	7	76145632	cg16453056	rs10952936	T	C	0.39	-0.31	1.54E-08	2.33E-17	1.66E-15	U
	PMS2L11	7	76625569	cg15770687	rs7804646	C	A	0.17	0.18	2.27E-02	3.89E-10	9.64E-05	U
34	*MESTIT1*	7	130130383	cg26275543	rs17164989	T	C	0.11	-0.41	4.74E-28	5.55E-20	1.37E-21	U
	MEST	7	130132453	cg13986840	rs2301335	G	A	0.41	0.37	4.90E-12	2.62E-24	7.38E-19	U
35	*HTR5A-AS1*	7	154863381	cg09623773	rs732050	G	A	0.29	0.24	7.23E-08	8.35E-11	3.47E-09	U
36	*LOC100506585*	7	157919408	cg14141549	rs2335836	C	G	0.34	0.24	6.09E-08	2.22E-08	4.06E-10	U
	WDR60	7	158750244	cg12954512	rs6957744	A	C	0.18	-0.3	1.02E-12	2.39E-23	1.52E-13	U
37	*LOC401442*	8	832260	cg03494825	rs10110537	T	C	0.33	0.44	6.48E-13	1.03E-38	8.16E-39	U
	MYOM2	8	2075777	cg21847720	rs2280902	A	G	0.46	-0.27	5.83E-12	3.28E-10	3.01E-13	U
	LOC101927815	8	2591411	cg08242633	rs4875852	C	T	0.08	-0.29	7.23E-21	2.19E-18	6.68E-21	U
38	***CLDN23***	**8**	**8559999**	**cg06671706**	**rs1060106**	**G**	**A**	**0.24**	**0.45**	**3.87E-28**	**8.10E-25**	**1.77E-23**	**U**
39	*CHD7*	8	61626625	cg26441877	rs10957154	A	G	0.27	-0.24	1.43E-04	4.10E-08	9.55E-11	U

40	*TRAPPC9*	8	141359539	cg26135849	rs10091104	C	T	0.37	-0.35	2.48E-22	9.60E-24	5.77E-21	U
41	*SLC46A2*	9	115652824	cg07758904	rs13283782	T	C	0.14	0.36	1.11E-12	1.56E-14	2.41E-17	U
42	***BEND7***	**10**	**13481944**	**cg24686497**	**rs11258384**	**G**	**A**	**0.31**	**0.33**	**6.75E-09**	**8.23E-10**	**2.19E-12**	**U**
43	*PTCHD3*	10	27702528	cg18669948	rs4638210	T	C	0.23	-0.28	6.26E-11	2.64E-15	1.91E-10	U
44	***ANKRD30BP3***	**10**	**45694889**	**cg26510023**	**rs10793594**	**C**	**A**	**0.43**	**-0.43**	**8.26E-28**	**1.98E-34**	**4.09E-37**	**U**
45	*REEP3*	10	65733388	cg19573236	rs12570824	G	A	0.13	-0.19	1.70E-06	4.65E-11	2.86E-08	U
46	***CHST15***	**10**	**125751413**	**cg20250269**	**rs4929810**	**A**	**G**	**0.23**	**0.34**	**9.94E-08**	**4.00E-13**	**7.96E-14**	**U**
47	***GLRX3***	**10**	**131989849**	**cg11372818**	**rs11017128**	**G**	**A**	**0.15**	**0.33**	**4.74E-14**	**7.30E-22**	**6.46E-17**	**U**
48	*H19*	11	2022021	cg06982169	rs11042170	G	A	0.5	0.82	0	0	0	B
	IGF2	11	2171694	cg25742037	rs4320932	C	T	0.18	0.33	8.88E-15	1.76E-22	1.17E-16	U
	INS	11	2182618	cg25336198	rs3741212	A	G	0.35	-0.25	5.59E-04	9.75E-11	3.30E-05	U
	KCNQ1OT1	11	2721591	cg09518720	rs231356	T	A	0.29	-0.6	4.91E-52	2.11E-60	3.57E-70	B
	LOC100506082	11	4259129	cg26583619	rs11031595	C	T	0.09	0.26	6.60E-07	1.02E-08	1.05E-11	P
49	*RBMXL2*	11	7110083	cg23916104	rs7114066	G	C	0.42	0.25	8.33E-07	4.09E-12	2.96E-09	U
50	***SPON1***	**11**	**14281011**	**cg02886208**	**rs10766125**	**T**	**C**	**0.19**	**-0.24**	**1.77E-09**	**9.30E-15**	**1.79E-13**	**U**
51	*LINC00301*	11	60414615	cg02534760	rs2847216	G	A	0.34	0.29	6.84E-05	4.92E-14	8.39E-08	U
52	*DNAJB13*	11	73676012	cg25592907	rs605442	T	C	0.38	0.38	1.12E-21	4.40E-30	5.59E-26	U
53	***NAALAD2***	**11**	**89867911**	**cg14304817**	**rs10734123**	**A**	**G**	**0.34**	**0.32**	**0.000439**	**1.81E-11**	**0.000415**	**U**
54	*KLRB1*	12	9555480	cg13830619	rs10743781	T	C	0.07	0.19	6.27E-08	4.91E-11	3.77E-08	U
55	***DDX11***	**12**	**31272865**	**cg08537890**	**rs11051208**	**G**	**A**	**0.2**	**0.5**	**8.86E-42**	**6.09E-43**	**3.08E-51**	**U**
56	***ALG10***	**12**	**34506462**	**cg02590409**	**rs10466832**	**T**	**C**	**0.39**	**0.29**	**2.63E-14**	**2.81E-15**	**9.77E-11**	**U**
57	*WDR66*	12	122356390	cg21171335	rs10840631	C	T	0.49	-0.36	6.97E-20	5.70E-23	4.66E-19	U
58	*RB1*	13	48892244	cg11408952	rs9316395	A	T	0.29	0.62	2.91E-36	3.51E-73	1.59E-61	U
59	***CDC16***	**13**	**114965839**	**cg12584960**	**rs9562157**	**A**	**G**	**0.31**	**-0.24**	**1.41E-10**	**6.50E-13**	**2.08E-09**	**U**
60	***PCNX1***	**14**	**71606274**	**cg15816911**	**rs221900**	**T**	**C**	**0.45**	**-0.3**	**2.69E-10**	**4.07E-20**	**5.54E-12**	**U**
61	*DLK1*	14	101194748	cg18279536	rs1004573	C	G	0.14	0.63	1.62E-33	3.39E-106	5.92E-78	U
	MEG3	14	101294147	cg08698721	rs7156824	A	C	0.12	0.39	2.00E-20	6.16E-18	1.84E-28	U
	MIR370	14	101367300	cg16126137	rs1956128	A	T	0.24	0.39	9.89E-12	1.13E-32	1.73E-23	U
	MIR487B	14	101512612	cg19560831	rs10083406	C	A	0.17	-0.39	2.09E-09	7.36E-45	1.68E-25	U
	MEG9	14	101696245	cg04165845	rs17587049	A	C	0.21	-0.52	5.97E-22	3.65E-33	4.33E-41	U
62	***ELK2AP***	**14**	**106183770**	**cg10832239**	**rs4977158**	**G**	**T**	**0.03**	**0.06**	**3.56E-10**	**2.13E-11**	**>0.05***	**U**
	FAM30A	**14**	**106374384**	**cg10270204**	**rs17646414**	**T**	**C**	**0.03**	**-0.06**	**>0.05***	**1.37E-10**	**>0.05***	**U**
63	*NDN*	15	23931674	cg18602919	rs11161347	G	A	0.5	-0.24	6.78E-05	3.88E-10	9.67E-07	U
	PWRN4	15	24347063	cg14555733	rs12592162	A	G	0.12	-0.32	1.29E-18	2.72E-20	6.41E-16	U
	PWRN2	15	24506388	cg13749113	rs12911863	C	T	0.49	-0.61	1.34E-63	7.24E-84	1.88E-74	U
	PWRN3	15	24672032	cg26288595	rs8033671	T	C	0.32	0.49	2.34E-13	1.24E-39	8.18E-31	U
	PWRN1	15	24803245	cg03402443	rs6576317	A	C	0.43	0.3	2.67E-14	2.30E-16	4.58E-13	U
	SNRPN	15	25123688	cg01786704	rs12906774	C	G	0.29	-0.45	2.01E-27	1.69E-33	3.33E-33	U
64	***CHST14***	**15**	**40778778**	**cg04949308**	**rs11070275**	**C**	**T**	**0.4**	**-0.35**	**3.94E-17**	**1.14E-21**	**7.58E-17**	**U**
65	*IGF1R*	15	99408958	cg12553689	rs11247377	G	A	0.08	-0.37	8.97E-15	3.80E-20	2.77E-22	B
	TTC23	15	99789855	cg16052317	rs12911333	A	T	0.07	0.38	2.90E-08	8.18E-09	1.72E-11	U
	LOC102723335	15	101095730	cg03418136	rs12915921	T	C	0.23	0.37	1.48E-12	5.85E-10	8.06E-09	U
66	*MEFV*	16	3304449	cg08260052	rs224217	G	A	0.5	0.28	6.11E-05	1.66E-09	2.73E-15	U
	ZNF75A	16	3355951	cg04234063	rs220381	G	A	0.33	0.26	4.74E-10	9.71E-12	9.07E-07	U
	ZNF174	16	3464107	cg01330954	rs17136367	C	G	0.13	-0.23	1.11E-03	1.03E-10	0.00053	U
	ZNF597	16	3481970	cg02880119	rs171634	A	G	0.23	-0.74	5.38E-100	7.01E-162	3.57E-113	U
	NAA60	16	3507492	cg21433313	rs1690450	A	G	0.33	-0.25	6.90E-04	8.48E-12	2.25E-07	U
	LOC102724927	16	3988869	cg05351887	rs2531995	C	T	0.38	0.24	9.65E-08	1.01E-08	6.84E-11	U
67	***CHST5***	**16**	**75563489**	**cg02390813**	**rs2550886**	**C**	**T**	**0.25**	**-0.24**	**2.49E-08**	**9.30E-13**	**6.95E-08**	**U**
68	*SPATA33*	16	89740564	cg03605463	rs2115401	T	C	0.41	-0.58	1.02E-77	6.52E-86	1.01E-77	U
69	*LOC284241*	18	77376689	cg10929690	rs3786235	T	C	0.18	-0.34	1.71E-17	6.41E-22	6.01E-23	U
	KCNG2	18	77659695	cg05491587	rs12456484	G	C	0.41	-0.35	4.63E-19	8.67E-22	1.29E-20	U
	PARD6G-AS1	18	77905119	cg18973878	rs11659843	T	A	0.19	-0.32	7.55E-07	5.32E-21	1.95E-07	U
	PARD6G	18	77918588	cg07500432	rs3809927	G	C	0.4	0.27	4.31E-11	1.59E-08	2.29E-14	U
70	***FEM1A***	**19**	**4784940**	**cg22992730**	**rs3087692**	**A**	**G**	**0.18**	**-0.5**	**7.88E-23**	**6.27E-34**	**4.51E-38**	**U**
71	***ZNF564***	**19**	**12624832**	**cg01559901**	**rs4804712**	**T**	**G**	**0.24**	**0.28**	**4.75E-14**	**1.01E-11**	**1.03E-09**	**U**
72	***LINC00664***	**19**	**21666788**	**cg06405146**	**rs2562458**	**G**	**A**	**0.2**	**-0.34**	**1.57E-13**	**7.88E-18**	**1.12E-13**	**U**
	ZNF429	**19**	**21688905**	**cg21617916**	**rs7343175**	**C**	**T**	**0.18**	**-0.2**	**7.16E-09**	**3.08E-12**	**7.64E-10**	**U**
73	***HKR1***	**19**	**37795026**	**cg26585454**	**rs10404945**	**C**	**T**	**0.16**	**-0.21**	**9.37E-10**	**3.60E-10**	**6.81E-09**	**U**
74	*PLEKHA4*	19	49340593	cg26267310	rs16982311	T	C	0.17	0.37	1.63E-11	2.07E-10	2.16E-09	U
75	*ZNF331*	19	54042165	cg15545827	rs16984967	C	A	0.11	-0.79	1.22E-41	1.23E-103	1.46E-105	?

76	*PEG3*	19	57350503	cg07310951	rs2040857	C	T	0.24	-0.3	4.19E-09	3.77E-09	2.74E-18	U
	MIMT1	19	57376177	cg06627087	rs411808	C	T	0.2	-0.32	2.99E-16	5.72E-17	2.12E-12	U
	ZSCAN1	19	58566643	cg18075691	rs4801552	G	A	0.47	0.28	1.09E-13	1.56E-11	2.93E-12	U
	ZNF135	19	58570454	cg02473540	rs8104122	T	C	0.27	0.21	4.17E-10	0.000846	2.80E-05	U
77	*ACTL10*	20	32256071	cg13403462	rs6088244	T	C	0.48	-0.41	1.34E-17	4.16E-38	1.70E-25	U
78	*BLCAP*	20	36148954	cg14765818	rs2064638	G	A	0.22	-0.47	7.04E-30	3.65E-55	1.29E-39	U
	NNAT	20	36149455	cg21588305	rs2064638	G	A	0.22	-0.36	3.61E-10	1.15E-22	4.13E-15	U
79	*LINC00494*	20	47013841	cg25181043	rs7267199	G	T	0.41	-0.35	4.00E-21	1.21E-21	2.33E-16	U
80	*GNAS-AS1*	20	57426935	cg03606258	rs11699704	C	T	0.15	-0.86	5.06E-167	2.45E-164	2.14E-190	B
	GNAS	20	57427146	cg24617313	rs6015389	C	T	0.36	-0.88	2.29E-284	0	0	B
	LOC101927932	20	57463991	cg09885502	rs2057291	A	G	0.32	0.8	6.76E-147	2.66E-203	2.79E-161	B
81	*DSCR3*	21	38630234	cg11287055	rs2051399	T	C	0.25	-0.27	1.07E-15	3.42E-11	5.71E-12	U
82	***WRB***	**21**	**40757691**	**cg00606841**	**rs2244352**	**T**	**G**	**0.33**	**0.41**	**9.79E-08**	**7.12E-23**	**5.73E-30**	**U**
83	*PRMT2*	21	48081686	cg24877093	rs6518306	T	C	0.11	-0.35	2.96E-19	2.66E-15	1.73E-16	U
84	***SELENOM***	**22**	**31500896**	**cg21361322**	**rs11705137**	**C**	**T**	**0.49**	**-0.27**	**2.02E-07**	**3.26E-16**	**4.18E-10**	**U**
85	*SNU13*	22	42078707	cg11677105	rs4822052	A	G	0.19	0.52	8.51E-37	1.77E-60	1.81E-37	U
	CYP2D6	22	42524984	cg25452165	rs4147641	G	C	0.3	0.32	5.12E-17	9.97E-18	4.73E-18	U
	TCF20	22	42548783	cg15557168	rs2143139	G	C	0.32	-0.26	2.28E-10	7.67E-14	1.90E-10	U

Novel loci are displayed in bold.

* Results where the test of association did not reach nominal significance (p-value >0.05) were not stored. CpG BP: CpG base pair position; Birth P, Child. P and Adol. P: P-value of SNP parent-of-origin effect on the CpG using DNA methylation measured at Birth, Childhood and Adolescence respectively.

Table 2. Top results of gene enrichment analysis based on gene ontology.

GO ID	Description	Genes in set	Observed	Expected	FDR
GO:0005158	Insulin receptor binding	31	4	0.11	6.64E-03
GO:0005159	Insulin-like growth factor receptor binding	15	3	0.05	1.54E-02
GO:0071514	Genetic imprinting	25	4	0.1	2.36E-02
GO:0008146	Sulfotransferase activity	51	3	0.17	4.31E-01
GO:0046883	Regulation of hormone secretion	252	6	1.01	9.67E-01
GO:0045840	Positive regulation of mitotic nuclear division	42	3	0.17	9.67E-01
GO:0032869	Cellular response to insulin stimulus	188	5	0.75	9.67E-01

FDR: False Discovery Rate; Observed: Number of genes observed; Expected: Number of genes expected under the null hypothesis (no enrichment).

Table 3. Gene enrichment analysis based ono DisGeNET and GLAD4U databases.

Disease Name	Genes in set	Obs	Exp.	FDR	Genes
Uniparental Disomy	57	15	0.25	0.00E+00	*KCNQ1OT1, ZNF597, H19, GRB10, MESTIT1, IGF2, MEST, NDN, PLAGL1, MEG3, HYMAI, SNRPN, PWRN1, NAA60, DLK1*
Beckwith-Wiedemann Syndrome	63	7	0.28	1.52E-05	*KCNQ1OT1, ZNF597, H19, IGF2, PLAGL1, ZNF135, NAA60*
Alzheimer's Disease	167	9	0.82	8.60E-05	***SPON1***, ***CR1***, *CYP2D6, HECW1, IGF1R, IGF2, IGF2R, INS,* ***MOBP***
Fetal Growth Retardation	64	6	0.28	3.33E-04	*H19, GRB10, IGF1R, IGF2, MEST, PLAGL1*
transient neonatal diabetes	16	4	0.07	4.08E-04	*KCNQ1OT1, INS, PLAGL1, HYMAI*
Growth Disorders	246	9	1.09	6.13E-04	*KCNQ1OT1, GNAS, H19, GRB10, IGF1R, IGF2, MEST, PLAGL1, CHD7,*
Small-for-dates baby	45	5	0.2	6.13E-04	*H19, IGF1R, IGF2, INS, DLK1*
Macroglossia	20	4	0.09	6.13E-04	*KCNQ1OT1, H19, PLAGL1, HYMAI*
Lewy Body Disease	19	4	0.09	6.79E-04	*IGF1R, IGF2, IGF2R, INS*
Hypothalamic Neoplasms	98	6	0.43	1.38E-03	*GNAS, NNAT, PLAGL1, MEG3, RB1, DLK1*
Facial Asymmetry	27	4	0.12	1.52E-03	*H19, IGF2, MEST, CHD7,*
Pituitary adenoma	107	6	0.47	1.70E-03	*GNAS, NNAT, PLAGL1, MEG3, RB1, DLK1,*
Pituitary Neoplasms	109	6	0.48	1.70E-03	*GNAS, NNAT, PLAGL1, MEG3, RB1, DLK1*
Endocrine System Diseases	489	11	2.17	1.70E-03	*GNAS, IGF1R, IGF2, INS, PLAGL1, MEG3, CHD7, HYMAI, TCF20, DLK1, DIRAS3,*
Central Nervous System Neoplasms	238	8	1.05	1.70E-03	*GNAS,* ***MAP2***, *NNAT, PEG3, PLAGL1, MEG3, RB1, DLK1*
Fetal Diseases	172	7	0.76	1.72E-03	*H19, GRB10, IGF1R, IGF2, IGF2R, MEST, DLK1*
Wilms Tumor	73	5	0.32	2.37E-03	*AIRN, KCNQ1OT1, H19, IGF2, DLK1*
Angiofibroma	12	3	0.05	2.37E-03	*H19, AHRR, RB1*
Nervous System Neoplasms	257	8	1.14	2.37E-03	*GNAS,* ***MAP2***, *NNAT, PEG3, PLAGL1, MEG3, RB1, DLK1*
Pregnancy	356	9	1.58	3.33E-03	*KCNQ1OT1, H19, IGF2, IGF2R, INS, MEST, PLAGL1, HYMAI, DLK1,*
Chromosome Disorders	448	10	1.98	3.33E-03	*DSCR3, KCNQ1OT1, H19, IGF2, MEST, NDN, PLAGL1, SNRPN,* ***WRB***, *DLK1*
Embryo Loss	53	4	0.23	1.01E-02	*MEG9, KCNQ1OT1, PLAGL1, DLK1*
Pain, Postoperative	21	3	0.09	1.15E-02	*CYP2D6,* ***ZNF429***, ***LINC00664***
Pituitary Diseases	109	5	0.48	1.28E-02	*GNAS, NNAT, PLAGL1, MEG3, DLK1*
Nephroblastoma	16	3	0.08	1.33E-02	***PCSK9***, *H19, IGF2*
Adrenocortical carcinoma	17	3	0.08	1.33E-02	*IGF1R, IGF2, RB1*
Prader-Willi Syndrome	192	6	0.85	2.11E-02	*NDN, SNRPN, PWRN1, PWRN2, TRAPPC9, DLK1*
Acromegaly	27	3	0.12	2.20E-02	*GNAS, INS, PLAGL1*
Cryptorchidism	69	4	0.31	2.29E-02	*H19,* ***DNAJB13***, *CHD7, PWRN1*
Somatotroph adenoma	28	3	0.12	2.29E-02	*GNAS, PLAGL1, MEG3*
Obesity	384	8	1.7	2.59E-02	*PCSK9, GNAS, IGF2, INS, NDN, SNRPN, PWRN1, DLK1*
Corticotroph adenoma	30	3	0.13	2.63E-02	*GNAS, MEG3, DLK1*
Parkinson Disease	109	5	0.54	2.83E-02	*CYP2D6, IGF1R, IGF2, IGF2R, INS*
Hypoglycemia	80	4	0.35	3.51E-02	*KCNQ1OT1, H19, IGF2, INS*
Rhabdomyosarcoma, Alveolar	34	3	0.15	3.51E-02	*IGF1R, IGF2,* ***FEM1A***
Neonate	311	7	1.38	3.51E-02	*GNAS, H19, IGF2, INS, PLAGL1, CHD7, HYMA*

Diabetes, Gestational	83	4	0.37	3.76E-02	*IGF2, INS, TCF20, DLK1*
Chromosome Aberrations	418	8	1.85	3.80E-02	*H19, MEST, PLAGL1, MEG3, HYMAI, RB1, SNRPN, DLK1*
Rhabdomyosarcoma	87	4	0.39	4.25E-02	*H19, IGF1R, IGF2, DLK1*
Hydatidiform Mole	38	3	0.17	4.35E-02	*ZNF597, NAP1L5, HYMAI*
Russell-Silver syndrome	6	2	0.03	4.63E-02	*H19, IGF2*

Novel loci are displayed in bold.

FDR: False Discovery Rate; Obs: Number of genes observed; Exp: Number of genes expected under the null hypothesis (no enrichment).

Table 4. Phenoscanner lookup of significant SNPs exerting POEs.

Phenotype	**SNP**	**Phenotype P**	**CpG Chr**	**CpG BP**	**CpG ID**	**Birth P**	**Child. P**	**Adol. P**	**Closest Gene**
Hair color red	rs258319	5.70E-34	16	89740564	cg03605463	1.97E-76	4.71E-79	3.55E-79	*SPATA33*
Height	rs6088244	2.90E-19	20	32255988	cg14921437	1.58E-19	4.16E-38	2.41E-30	*ACTL10*
Height	rs4320932	2.30E-12	11	2171694	cg25742037	8.88E-15	1.76E-22	1.17E-16	*IGF2*
Fasting glucose	rs6976501	8.96E-12	7	50849723	cg16349612	2.31E-05	1.19E-10	4.19E-22	*GRB10*
Height	rs2057291	2.20E-10	20	57463991	cg09885502	6.8E-147	2.6E-203	2.8E-161	*LOC101927932*
Height	rs2735469	2.50E-10	11	2022021	cg06982169	4.25E-31	1.01E-42	1.36E-35	*H19*
Type II diabetes	rs231356	3.70E-10	11	2721591	cg09518720	4.91E-52	2.11E-60	3.57E-70	*KCNQ1OT1*
BMI	rs2531995	3.94E-10	16	3988700	cg05351887	9.65E-08	1.01E-08	6.84E-11	*LOC102724927*
Mean corpuscular hemoglobin concentration MCHC	rs2064792	2.26E-09	6	164461131	cg26079810	2.82E-26	2.57E-41	1.95E-38	*LOC102724152*
Age at menopause	rs12544305	8.80E-09	8	61626625	cg26441877	0.000143	2.71E-09	9.55E-11	*CHD7*
Red blood cell traits combined	rs2238439	1.97E-08	16	3988869	cg05351887	2.07E-09	3.30E-10	6.89E-07	*LOC102724927*
Serum metabolite mass spec peak 6359 mz	rs10829705	2.93E-08	10	131989849	cg11372818	2.60E-13	1.52E-20	1.04E-19	*GLRX3*
Schizophrenia	rs2143139	4.57E-08	22	42524984	cg25452165	1.99E-10	7.67E-14	1.84E-11	*CYP2D6*
Type 1 diabetes	rs3741206	6.33E-08	11	2171694	cg25742037	2.47E-06	1.56E-11	3.72E-13	*IGF2*
Height	rs11743146	3.80E-07	5	37209440	cg00331501	9.08E-11	7.85E-13	2.77E-07	*LOC105374727*
log(eGFR creatinine)	rs6593140	5.00E-07	7	50849723	cg16349612	5.35E-18	5.74E-16	2.35E-18	*GRB10*
Nicotine dependence	rs171634	5.08E-07	16	3481970	cg02880119	5.38E-100	7.01E-162	3.57E-113	*ZNF597*
Body mass index BMI	rs221892	6.96E-07	14	71606274	cg15816911	1.69E-10	6.15E-20	3.60E-12	*PCNX1*
Spine bone mineral density BMD	rs2244352	9.23E-07	21	40757691	cg00606841	9.79E-08	7.12E-23	5.73E-30	*WRB*
Height	rs6976501	1.00E-06	7	50849639	cg09150232	0.556	1.19E-10	4.19e-22	*GRB10*
Schizophrenia	rs4798923	1.07E-06	18	77659695	cg05491587	9.34E-18	8.38E-21	3.61e-20	*KCNG2*
Height	rs2407093	1.30E-06	13	48892244	cg11408952	3.00E-29	3.36E-60	1.32e-52	*RB1*
log(eGFR creatinine)	rs6593140	1.30E-06	7	50849639	cg09150232	0.00781	1.59E-12	2.35e-18	*GRB10*
Hip circumference in females	rs2531995	1.40E-06	16	3988869	cg05351887	9.65E-08	1.01E-08	6.84e-11	*LOC102724927*
Height	rs2531995	1.87E-06	16	3988869	cg05351887	9.65E-08	1.01E-08	6.84e-11	*LOC102724927*
Serum creatinine	rs6593140	1.90E-06	7	50849639	cg09150232	0.00781	1.59E-12	2.35e-18	*GRB10*
Schizophrenia	rs2240341	2.14E-06	14	71606274	cg15816911	1.34E-11	6.07E-17	1.79e-10	*PCNX1*
Weight	rs2238439	2.22E-06	16	3988694	cg01971227	1.48E-06	3.30E-10	6.22e-06	*LOC102724927*

Phenotype P: P-value of the association between phenotype and SNP; CpG BP: CpG base pair position; Birth P, Child. P and Adol. P: P-value of SNP parent-of-origin effect on the CpG using DNA methylation measured at Birth, Childhood and Adolescence respectively.

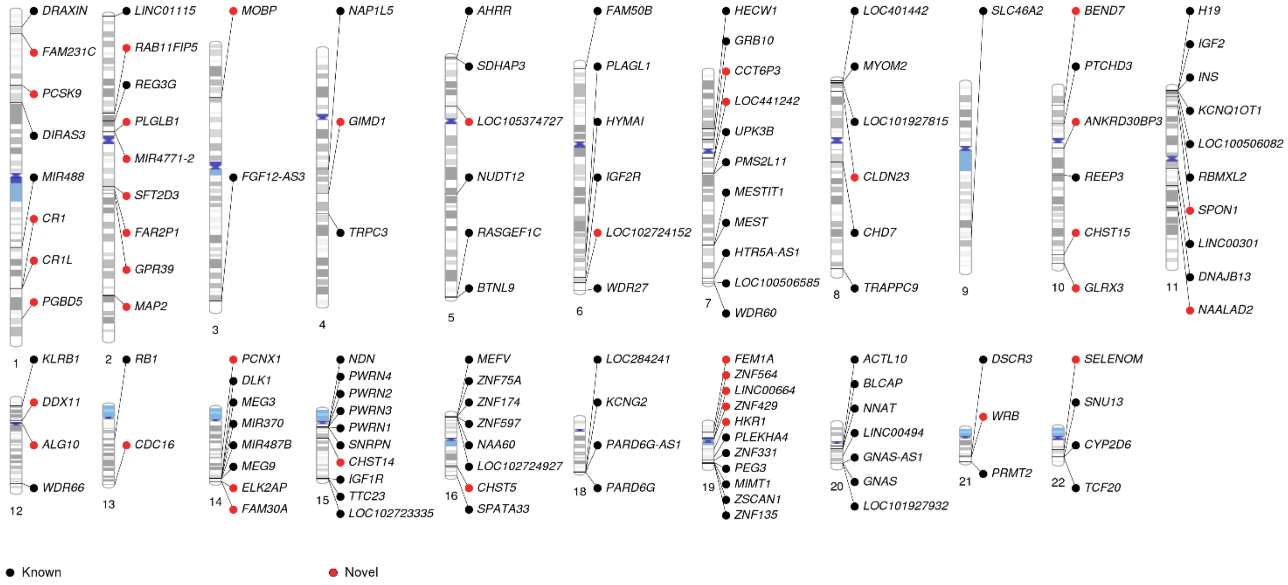

Figure 1. Candidate imprinted loci.

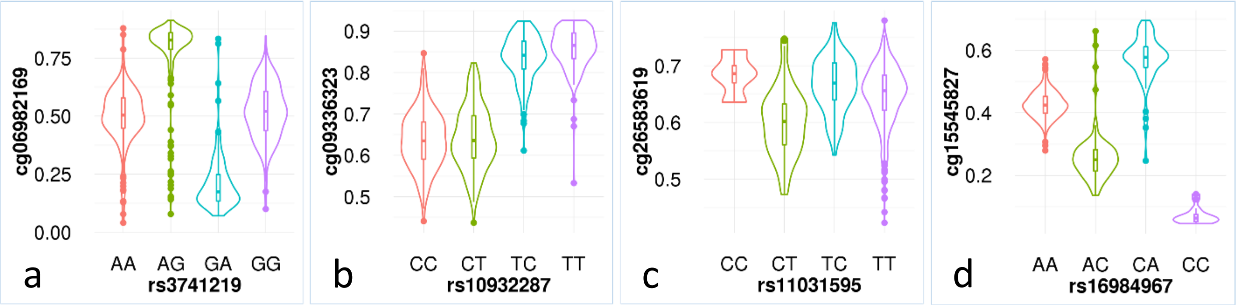

Figure 2. Patterns of parent-of-origin effects. Violin plots showing four patterns of CpG methylation observed in this study: (a) Bipolar dominance pattern observed at a CpG site near *H19* where one heterozygous genotype has a larger mean phenotypic value than the two homozygotes and the other heterozygote has a smaller mean value; (b) The canonical pattern of imprinting observed at a CpG site near *MAP2*, where one of the alleles leads to a larger phenotypic value than the other and one of the chromosomes is putatively silenced; (**c**) Underdominance pattern at a CpG site near *LOC100506082*, where one of the heterozygotes has a lower phenotypic value than the rest of the genotype groups; (**d**) Uncharacterized pattern at a CpG site near *ZNF331*.

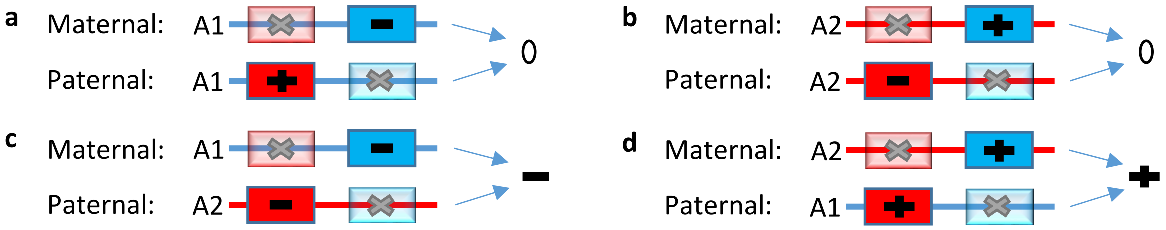

Figure 3. A Mechanism that generates a bipolar dominance pattern. Each of the panels in the figure displays the same two SNPs (in blue and in red) which are in high LD with each other on two different haplotypes (A1 and A2). In the case of the A1 haplotype, the allele encoded by the red SNP has a positive effect on the phenotype while the allele encoded by the blue SNP has a negative effect. In the case of the A2 haplotype, the allele encoded by the red SNP has a negative effect on the phenotype while the allele encoded by the blue SNP has a positive effect. In this example, genomic imprinting results in the red SNP being inactive in the chromosome inherited by the mother and the blue SNP being inactive in the chromosome inherited by the father. In panels (**a**) and (**b**), individuals who receive two copies of either haplotype A1 or haplotype A2 have a mean phenotype of 0. In panel (**c**) the effect on phenotype is negative as haplotype A1 is inherited from the mother and haplotype A2 from the father. In panel (**d**) the overall effect is positive as haplotype A2 is inherited from the mother and haplotype A2 from the father.