GWAS of DNA Methylation Variation Within Imprinting Control Regions Suggests Parent-of-Origin Association

Miguel E. Rentería,[1,2] Marcel W. Coolen,[3,4] Aaron L. Statham,[3] R. Seong Min Choi,[1] Wenjia Qu,[3] Megan J. Campbell,[1] Sara Smith,[1] Anjali K. Henders,[1] Grant W. Montgomery,[1] Susan J. Clark,[3] Nicholas G. Martin,[1] and Sarah E. Medland[1]

[1]Queensland Institute of Medical Research, Brisbane, Queensland, Australia
[2]The University of Queensland, School of Psychology, Brisbane, Queensland, Australia
[3]Epigenetics Research Group, Cancer Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia
[4]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Imprinting control regions (ICRs) play a fundamental role in establishing and maintaining the non-random monoallelic expression of certain genes, via common regulatory elements such as non-coding RNAs and differentially methylated regions (DMRs) of DNA. We recently surveyed DNA methylation levels within four ICRs (*H19*-ICR, *IGF2*-DMR, KvDMR, and *NESPAS*-ICR) in whole-blood genomic DNA from 128 monozygotic (MZ) and 128 dizygotic (DZ) human twin pairs. Our analyses revealed high individual variation and intra-domain covariation in methylation levels across CpGs and emphasized the interaction between epigenetic variation and the underlying genetic sequence in a parent-of-origin fashion. Here, we extend our analysis to conduct two genome-wide screenings of single nucleotide polymorphisms (SNPs) underlying either intra-domain covariation or parent-of-origin-dependent association with methylation status at individual CpG sites located within ICRs. Although genome-wide significance was not surpassed due to sample size limitations, the most significantly associated SNPs found through multiple-trait genome-wide association (MQFAM) included the previously described rs10732516, which is located in the vicinity of the *H19*-ICR. Similarly, we identified an association between rs965808 and methylation status within the *NESPAS*-ICR. This SNP is positioned within an intronic region of the overlapping genes *GNAS* and *GNAS-AS1*, which are imprinted genes regulated by the *NESPAS*-ICR. Sixteen other SNPs located in regions apart from the analyzed regions displayed suggestive association with intra-domain methylation. Additionally, we identified 13 SNPs displaying parent-of-origin association with individual methylation sites through family-based association testing. In this exploratory study, we show the value and feasibility of using alternative GWAS approaches in the study of the interaction between epigenetic state and genetic sequence within imprinting regulatory domains. Despite the relatively small sample size, we identified a number of SNPs displaying suggestive association either in a domain-wide or in a parent-of-origin fashion. Nevertheless, these associations will require future experimental validation or replication in larger and independent samples.

■ **Keywords:** imprinting, *IGF2*, *H19*, *NESPAS*, KvDMR, GWAS, SNP, parent-of-origin, methylation

DNA methylation is a fundamental epigenetic modification implicated in many cellular processes, such as development, chromatin structure, maintenance of genomic imprinting, and *X* chromosome inactivation in females (Chow et al., 2005; Delaval & Feil, 2004; Robertson, 2005). DNA methylation status across specific loci in the genome also mediates the regulation of gene expression and maintains genome integrity (Weber & Schubeler, 2007). DNA methylation patterns are remarkably stable within functionally important regions (Kaminsky et al., 2009; Ushijima et al., 2003) but can be highly variable in other genomic regions (Fraga et al.,

2005; Mill et al., 2006; Petronis et al., 2003), even in genetically identical individuals (Coolen et al., 2011; Fraga et al., 2005; Heijmans et al., 2007). This suggests a possible involvement of environmental factors in the modulation of these changes. Interestingly, there is now increasing evidence that heritable factors may also play a role in the determination of DNA methylation patterns, at least at the genome-wide level (Bjornsson et al., 2008; Boks et al., 2009; Numata et al., 2012).

Imprinting control regions (ICRs) are differentially methylated regulatory elements located near clusters of imprinted genes that coordinate preferential gene transcription in a parent-of-origin fashion (Constancia et al., 2004; Li et al., 1993). ICRs entail segments of DNA rich in CpG dinucleotides, with the cytosine nucleotides displaying differential methylation depending upon parent-of-origin alleles. Whether the methylated/unmethylated state leads to silencing or not depends upon the default imprint state of the region (i.e., methylated or unmethylated; Lee et al., 2002). Abnormal DNA methylation within ICRs has been shown to result in loss of appropriate expression of imprinted genes. It is also associated with a number of diseases, such as Beckwith–Wiedemann syndrome (Smilinich et al., 1999; Weksberg et al., 2001, 2002), Silver–Russell syndrome (Eggermann et al., 2008; Gicquel et al., 2005), Angelman syndrome, and Prader–Willi syndrome (Zeschnigk et al., 1997). Due to the functional importance of genetic imprinting, it has been proposed that DNA methylation within ICRs might be under particular genetic regulation (Heijmans et al., 2007).

To address the impact of genetic sequence on DNA methylation variation, we recently performed a detailed study in a sample comprising whole-blood DNA from 128 pairs of monozygotic (MZ) and 128 pairs of dizygotic (DZ) human twins (Coolen et al., 2011). Twin studies have been particularly valuable in genetics research, and now offer an opportunity for the study of epigenetic variation as a dynamic quantitative trait. MZ pairs share the entirety of their genetic sequence and/or DZ pairs share approximately half of their genetic alleles, while both MZ and DZ pairs are usually subject to similar household conditions. Thus, by studying epigenetic differences in twins, it is possible to infer the proportion of epigenetic variation attributable to either environmental or genetic factors. We recently interrogated DNA methylation status of CpGs located within four ICRs (H19-ICR, IGF2-DMR, KvDMR, and NESPAS-ICR) and the non imprinted gene RUNX1 (Coolen et al., 2011). Details of interrogated domains and individual CpGs are given in Figure S1 of our previous study (Coolen et al., 2011). A high degree of variability in individual CpG methylation levels, notably at the *H19/IGF2* loci, was observed (see Figure 2 of Coolen et al., 2011). And, overall, a similar pattern of methylation variation between MZ and DZ twins was found, with imprinted regions displaying median methylation levels around the

expected 50%, and only a few single nucleotide polymorphisms (SNPs) showing increased mean methylation (*H19*-ICR_CpG02, *H19*-ICR_CpG03, *IGF2*-DMR_CpG05, and *NESPAS*-ICR_CpG14). Part of this DNA methylation plasticity seems to be clearly attributable to environmental and stochastic factors. However, concordant gains or losses of methylation were more common in MZ than DZ twin pairs, suggesting that de novo and/or maintenance methylation is influenced by the surrounding DNA sequence (see Table 1 of Coolen et al., 2011). This is in agreement with a previous longitudinal study, which reported familial clustering of methylation variation over time (Bjornsson et al., 2008). Importantly, we also observed significant intra- but not inter-domain covariation in methylation state across ICRs (see Figure 3 in Coolen et al., 2011) and showed that the rs10732516 (A/G) polymorphism, as well as the co-inherited rs2839701 (G/C) polymorphism — both located within the *IGF2/H19* locus — are strongly associated with increased hypermethylation of specific CpG sites in the maternal *H19* allele, which was later confirmed by clonal bisulfite sequencing analysis (see Figure 7 in Coolen et al., 2011). Additionally, we also performed a DNA sequencing analysis of all interrogated regions to exclude the possibility of bias introduced by genetic variation or mutations within primer binding sequences. Here, we expand our previous analyses to conduct a genome-wide search for other common polymorphisms that may be associated with methylation status within the ICRs of interest.

The identification of genetic variants that are associated with gene-specific DNA methylation could open a new avenue to the understanding of DNA methylation regulation. Genome-wide association studies (GWAS) of DNA methylation have been recently published and a number of methylation quantitative trait loci (mQTL) have been identified (Bell et al., 2011; Gibbs et al., 2010), demonstrating the usefulness of incorporating this approach into epigenetics research. However, by treating each individual CpG as an independent variable, such studies have ignored the underlying correlation that exists across neighboring methylation sites.

In order to maximize power to detect common SNPs associated with DNA methylation within the ICRs of interest, we have used the MQFAM algorithm implementation for PLINK (Ferreira & Purcell, 2009), a multivariate GWAS approach that employs canonical correlation analysis to account for cross-trait covariance (Ferreira & Purcell, 2009). Multiple-trait GWAS improves power to detect association signals with pleiotropic effects over sets of correlated traits, without an increase in the false discovery rate (Bolormaa et al., 2010), and has been applied to combinations of phenotypes such as blood lipid levels and gene expression (Inouye et al., 2010) or obesity, and osteoporosis measures (Liu et al., 2009).

Furthermore, as our previous analysis (Coolen et al., 2011) highlighted that SNP rs10732516 appears to have a

substantial impact on the DNA methylation status of the *H19*-ICR region only when inherited on the maternal allele, we also conducted genome-wide family-based association using the quantitative transmission disequilibrium test (QTDT) package to test for quantitative association of both paternally and maternally inherited SNP alleles with the DNA methylation status at each methylation site (Spielman & Ewens, 1996).

Materials and Methods

Subjects

The twin samples analyzed are part of a study on moliness and cognitive function (Bataille et al., 2000; Wright et al., 2001) and comprise 512 adolescent twins (70 female/female and 58 male/male MZ twin pairs; 25 female/female, 29 male/male, and 74 opposite sex DZ twin pairs), with a mean age of 14.15 years ($SD = 2.46$; range 12–22.85). The samples are predominantly (>95%) of northern European origin (mainly Anglo-Celtic). Zygosity of the twins was confirmed using microsatellite repeat marker testing. Written informed consent to participate in this study was given by all participants and by their parents, legal guardians, or caretakers when participants were under-aged. The study protocol was approved by the Queensland Institute of Medical Research Human Research Ethics Committee (Ethics approvals P193 and P455). Additionally, as part of the study, parents were asked to complete a series of questionnaires and to donate blood samples.

Genotyping

DNA was extracted from blood samples and SNP genotyping performed with Illumina HumanHap 610W Quad arrays (http://support.illumina.com/array/array_kits/human610-quad_beadchip_kit/documentation.ilmn) by deCODE Genetics (Reykjavík, Finland). Genotype data were screened through a series of quality control criteria including Mendelian errors, minor allele frequency (MAF) ≥ 1%, *p* value of a Hardy–Weinberg equilibrium (HWE) test ≥ 10^{-6}, SNP call rate > 95%, and Illumina Beadstudio GenCall score ≥ 0.7. We also screened for ancestral outliers by using principal component analysis (PCA; Price et al., 2006), comparing the genotyped data in the discovery sample with 16 global populations sourced from HapMap Phase 3 and northern European subpopulations from a previous study (McEvoy et al., 2009).

DNA Methylation Assay Design

MassCLEAVE™ assays against the genomic regions of interest were designed and tested using the AmpliconReport R-script that we described previously (Coolen et al., 2007). In brief, the method consists of a Polymerase chain reaction (PCR) amplification of bisulfite-converted DNA tagged with a T7-promoter, which is followed by the generation of a single-stranded RNA molecule and base-specific cleavage (3′ to either rUTP or rCTP) with RNase A. The result-

ing mixture of cleavage products with different length and mass is analyzed with MALDI-TOF mass spectrometry. Differences in original DNA methylation state are reflected in changes in nucleotide sequence after bisulfite treatment, and therefore will produce different fragment masses in the assay. The abundance of each fragment (signal/noise level in the spectrum) acts as an indicator of the amount of DNA methylation in the interrogated sequence (Coolen et al., 2007). The analyzed regions comprised imprinted regions *H19*-ICR (Takai et al., 2001), *IGF2*-DMR (Cui et al., 2003), KvDMR (Nakano et al., 2006; Smilinich et al., 1999), and *NESPAS*-ICR (Liu et al., 2005), and the promoter of the non-imprinted *RUNX1* gene (see Figure S1 in Coolen et al., 2011). In this study, we focus our analyses exclusively on the ICRs.

Genomic Bisulfite Treatment

DNA methylation measurements were performed on genomic DNA extracted from whole blood. Bisulfite treatment was carried out using the EZ-96 DNA Methylation-Gold Kit (Zymo Research: Cat No. D5008) according to the manufacturer's instructions. In brief, 500 ng DNA was used in the bisulfite reaction and incubated for 8 hours at 55 °C. After desulfonation and clean up, the bisulfite treated DNA was resuspended in 50 μL of which 2 μL was used in each PCR.

PCR-Tagging, In Vitro Transcription and Mass Spectrometry Analysis

The target regions were amplified in triplicate using the primer pairs and annealing temperatures (Ta) described in Table S1 of Coolen et al. (2011). The MassCleave methylation analysis was performed as described previously (Coolen et al., 2007). In brief, the PCR reactions were carried out in a total volume of 5 μL using 200 nM forward and reverse primer, 200 μM Deoxyribonucleotide triphosphate (dNTP), 1.5 mM MgCl₂ 1:100,000 dilution of SYBR Green (Invitrogen) and 0.35 U Platinum Taq DNA polymerase (Invitrogen) in 1 × PCR buffer without magnesium. PCR success was determined via melt curve analysis (ABI PRISM® 7700). Unincorporated dNTPs were dephosphorylated by incubation at 37 °C for 20 minutes in the presence of 1.7 μL H₂O and 0.3 U Shrimp Alkaline Phosphatase (SAP), followed by a heat-inactivation for 5 minutes at 85 °C. The triplicate PCR samples after SAP treatment were pooled and of this pool, 2 μL were used in a 7 μL transcription reaction, containing 3.14 mM DTT, 2.5 mM dCTP, 1 mM rUTP, 1 mM rGTP, 1 mM rATP, 20 U T7 R&DNA polymerase, and 0.09 mg/mL RNase A in 0.64 × T7 polymerase buffer (all reagents from SEQUENOM, San Diego). Transcription and digestion were performed in the same step at 37 °C for 3 hours. After the addition of 20 μL H₂O, conditioning of the phosphate backbone prior to MALDI-TOF MS was achieved by the addition of 6 mg CLEAN Resin (SEQUENOM, San Diego). Twenty-two nanoliters of the

cleavage reactions were robotically dispensed onto silicon chips preloaded with matrix (SpectroCHIPs; SEQUENOM, San Diego). Mass spectra were collected using a MassARRAY mass spectrometer (SEQUENOM).

Calculation of Methylation Ratios

We used a sensitive and high-throughput method for DNA methylation analysis that is quantitative to 5% methylation for each informative CpG residue (Coolen et al., 2007). Calculation of the DNA methylation ratios was performed using the R-script Analyze Sequenom Function (ASF; Coolen et al., 2007), which is an adaptation of the formula used by MassCLEAVE™ software. A full description of ASF has been published previously (Coolen et al., 2007). All statistical calculations were carried out using either Stata 9 (StataCorp LP, Texas, USA) or the free 'R' software package for statistical computing (http://www.R-project.org).

Multiple-Trait Association

Four parallel multiple-trait association were conducted with the MQFAM (Ferreira & Purcell, 2009) algorithm as implemented in PLINK (Purcell et al., 2007). Each study comprised CpGs contained within each ICR domain. *H19*-ICR (12 variables), *IGF2*-DMR (seven variables), KvDMR (11 variables), and *NESPAS*-ICR (11 variables). Descriptive statistics for these variables are presented in Table 1. MQFAM uses canonical correlation analysis (CCA) to measure the association between two sets of variables. To explain how CCA works, it is possible to make the following analogy with PCA: PCA is usually applied to one set (X) of possibly correlated traits to extract a number of independent variates (components) that explain as much variance in the original set, whereas CCA is applied to two sets of variables (X and Y) to extract a number of independent pairs of variates (U_i, V_i) that explain as much covariance between the two original sets. More details on the method can be found at the MQFAM support site (http://genepi.qimr.edu.au/staff/manuelF/multivariate/main.html). Data were adjusted to correct for any sex, age, age^2, sex*age, and sex*age^2 effects. Although PLINK does not account for family structure, MQFAM offers the possibility of running permutation testing to correct for relatedness. However, given that this analysis is slow and computationally intensive when many traits are analyzed at the same time, we performed a two-stage analysis in order to reduce computation time. In the first stage, association testing was performed on a filtered dataset which included only one individual per family ($N = 256$), hence making permutation correction unnecessary, and allowing us to identify a subset of potentially associated 'candidate' SNPs (cut-off was set at p value < .0001). Then, in the second stage of the analysis, we tested this subset of 'candidate' SNPs for association in the sample consisting of all DZ twins ($N = 256$ individuals from 128 pairs) and one MZ from each pair ($N = 128$ individuals),

resulting in a total sample of 384 individuals. We accounted for relatedness using 1 million permutations at each locus.

Single-Trait Association

DNA methylation status was tested for standard single-trait association at the individual methylation unit level using MERLIN (Abecasis et al., 2002). Unlike PLINK, MERLIN corrects for relatedness by incorporating pedigree data. We also controlled for any age, sex, age^2, sex*age, and sex*age^2 effects. The total sample size for single-trait association was $N = 512$.

Family-Based Association Testing (Parent-of-Origin Ass)

We conducted family-based genome-wide testing for quantitative total association of both paternally and maternally inherited SNP alleles with DNA methylation levels at individual methylation sites across four ICRs using QTDT (Spielman & Ewens, 1996). QTDT is a software package for family-based linkage disequilibrium analyses of quantitative and discrete traits that can use all the information in a pedigree to construct powerful tests of association (Spielman & Ewens, 1996). The total sample size ($N = 1,024$) consisted of 512 twins (128 MZ and 128 DZ pairs) and their parents ($N = 512$).

Results

Multivariate Intra-Domain GWAS

We performed multivariate GWAS by testing 515,966 genotyped SNPs for association with intra-domain DNA methylation variation at four ICRs (*H19*-ICR, *IGF2*-DMR, KvDMR, and *NESPAS*-ICR) using the MQFAM (Ferreira & Purcell, 2009) algorithm implementation for PLINK (Purcell et al., 2007). MQFAM performs CCA to measure the association between two sets of variables by extracting the linear combination of traits that explain the largest possible amount of the covariation between the marker (SNP) and all traits. As detailed elsewhere (Coolen et al., 2011), the number of analyzed DNA methylation sites by region was 14 (*H19*-ICR), 8 (*IGF2*-DMR), 15 (KvDMR), and 17 (*NESPAS*-ICR). Data were adjusted for any age or sex effects (and their interactions) prior to association analysis. Due to method restrictions (PLINK is unable to account for relatedness of individuals in the sample), we structured our analysis in two stages. First, we tested for association in a sample subset containing only one individual per family selected at random ($N = 256$) and identified possible 'candidate' SNPs (those with p value < .0001). The candidate SNPs were then tested for association in a larger dataset consisting of 384 individuals (all DZ pairs, $N = 256$; and one MZ twin per family, $N = 128$; this is because although MQFAM permutation can account for DZ zygosity, it cannot adjust for genetically identical individuals).

After permutation correction, multivariate GWAS results (shown in Figure 1) revealed no SNPs with p values

TABLE 1

Descriptive Statistics for Analyzed CpG Sites

CpG site	Mean[a]	SD	Range
H19-ICR-CpG02	0.795	0.102	(0.467, 1.000)
H19-ICR-CpG03	0.629	0.153	(0.158, 1.000)
H19-ICR-CpG04	0.45	0.092	(0.213, 0.733)
H19-ICR-CpG05	0.406	0.085	(0.135, 0.695)
H19-ICR-CpG06	0.42	0.081	(0.190, 0.670)
H19-ICR-CpG08	0.498	0.09	(0.270, 0.830)
H19-ICR-CpG09-10	0.381	0.071	(0.217, 0.617)
H19-ICR-CpG11	0.502	0.098	(0.187, 0.827)
H19-ICR-CpG12	0.378	0.07	(0.130, 0.613)
H19-ICR-CpG21	0.451	0.121	(0.020, 0.797)
H19-ICR-CpG22	0.475	0.072	(0.247, 0.717)
H19-ICR-CpG23	0.432	0.087	(0.160, 0.723)
IGF2-DMR-CpG01-02	0.417	0.08	(0.163, 0.653)
IGF2-DMR-CpG03	0.395	0.094	(0.090, 0.710)
IGF2-DMR-CpG04	0.482	0.094	(0.170, 0.785)
IGF2-DMR-CpG05	0.64	0.118	(0.250, 1.000)
IGF2-DMR-CpG06	0.459	0.102	(0.150, 0.750)
IGF2-DMR-CpG07	0.507	0.11	(0.140, 0.856)
IGF2-DMR-CpG08	0.47	0.117	(0.120, 0.830)
KvDMR-CpG01	0.443	0.032	(0.333, 0.543)
KvDMR-CpG06	0.396	0.034	(0.285, 0.490)
KvDMR-CpG08_09	0.442	0.046	(0.310, 0.594)
KvDMR-CpG10-11-12	0.428	0.046	(0.275, 0.605)
KvDMR-CpG15	0.455	0.032	(0.350, 0.548)
KvDMR-CpG16	0.463	0.035	(0.353, 0.570)
KvDMR-CpG17-18	0.472	0.047	(0.305, 0.625)
KvDMR-CpG20	0.462	0.033	(0.360, 0.570)
KvDMR-CpG21	0.44	0.032	(0.335, 0.535)
KvDMR-CpG24	0.494	0.036	(0.375, 0.600)
KvDMR-CpG25	0.508	0.04	(0.380, 0.618)
NESPAS-ICR-CpG02	0.398	0.043	(0.276, 0.536)
NESPAS-ICR-CpG05-06-07	0.482	0.042	(0.369, 0.609)
NESPAS-ICR-CpG08-09	0.538	0.052	(0.359, 0.719)
NESPAS-ICR-CpG10-11	0.493	0.039	(0.380, 0.610)
NESPAS-ICR-CpG12-17	0.499	0.033	(0.398, 0.590)
NESPAS-ICR-CpG14	0.874	0.031	(0.780, 0.970)
NESPAS-ICR-CpG15-16	0.527	0.041	(0.398, 0.658)
NESPAS-ICR-CpG18	0.508	0.033	(0.410, 0.600)
NESPAS-ICR-CpG19	0.636	0.057	(0.496, 0.796)
NESPAS-ICR-CpG20	0.456	0.035	(0.346, 0.586)
NESPAS-ICR-CpG21	0.519	0.037	(0.383, 0.640)

Note: [a]Methylation ratios range between 0 and 1.

surpassing the genome-wide significance threshold. Table 1 shows a list of the most significantly associated SNPs with methylation within each ICR; rs4930103, the most significantly associated SNP with methylation at the *H19*-ICR, is in Linkage disequilibrium (LD) with SNPs rs2839701 and rs10732516 ($r^2 = 0.854$), and is part of the haplotype that was previously described (Coolen et al., 2011). A locus zoom view of this region is provided in Figure 2.

We tested the association of rs4930103 with individual CpG sites within *H19*-ICR by conducting standard single-trait association testing on the entire dataset ($N = 512$) using MERLIN (Abecasis et al., 2002). Our analyses revealed that association was at least nominally significant ($p \leq .05$) in five out of the twelve methylation sites surveyed, and that the three CpG sites with the most significant *p* values (*H19*-ICR_CpG02, $p = 3.582e\text{-}15$; *H19*-ICR_CpG03, $p = 1.118e\text{-}05$; and *H19*-ICR_CpG04, $p = 2.88e\text{-}04$; notation as in Coolen et al., 2011) cluster within a 50 bp region. This indicates that the observed association signal between

rs4930103 and methylation levels within the *H19*-ICR domain does not predispose methylation at all CpG sites of the domain, but it may be restricted instead to specific loci.

Multivariate GWAS at *NESPAS*-ICR identified rs965808 as the most significantly associated SNP. rs965808 is located within an intronic region of the *GNAS* and *GNAS-AS1* imprinted genes, which are regulated by *NESPAS*-ICR, clearly suggesting a possible regulatory effect of the association signal observed between rs965808 and methylation of the *NESPAS* domain (a locus zoom view of this region is also shown in Figure 2).

Other SNPs identified through multivariate GWAS include rs11897432, rs2412488, and rs2555155, associated with methylation at *H19*-ICR; rs10462794, rs11227306, rs9596905, rs4304977, rs1007190, and rs1004689, associated with methylation at *IGF2*-DMR; rs7644516, rs11933531, rs7027203, and rs3858526, associated with *Kv*DMR methylation; and rs17261688, rs3763558, rs724210, rs1022588, and rs965808, associated with methylation within *NESPAS*-ICR. Table 2 shows most significantly associated SNPs, their location, closest genes, *F* test statistic (an indicator of accumulated evidence of association), loadings per trait (an indicator of each trait contribution to total association), and *p* values before and after permutation correction. While some SNPs map within or near coding genes, it is not obvious the role of such genes in modulating methylation status, as previous reports in the literature are limited. Hence, the identification of novel genetic variants opens up new possibilities for the understanding of both pre-programmed and dynamic epigenetic mechanisms.

To assess homogeneity of association signals across each ICR, we performed single-trait association using MERLIN for each suggestively associated SNP. rs2555155 and rs2412488, which were associated with methylation at *H19*-ICR were at least nominally significant in all and all but two sites, respectively. rs2555155 is located in an intronic region of genes *DNHD1* (Dynein heavy chain domain containing protein 1) and *FXC1* (Mitochondrial import inner membrane translocase subunit Tim9 B) in chromosome 11. On the other hand, rs2412488 is located within an intronic region of overlapping genes *FIP1L1* (Pre-mRNA 3′-end-processing factor *FIP1*) and *LNX1* (E3 ubiquitin-protein ligase) in chromosome 4. A Locus zoom view of these two regions is shown in Figure 3.

Parent-of-Origin Association With DNA Methylation

The second part of the analysis consisted of genome-wide screenings for evidence of parent-of-origin association between SNPs and DNA methylation at individual methylation sites. We used the genotypes and phenotypes from the whole twin sample ($N = 512$), as well as the genotypes of their parents ($N = 512$), to perform a family-based analysis with QTDT (Spielman & Ewens, 1996) to test for differences in association dependent upon paternal or maternal allelic
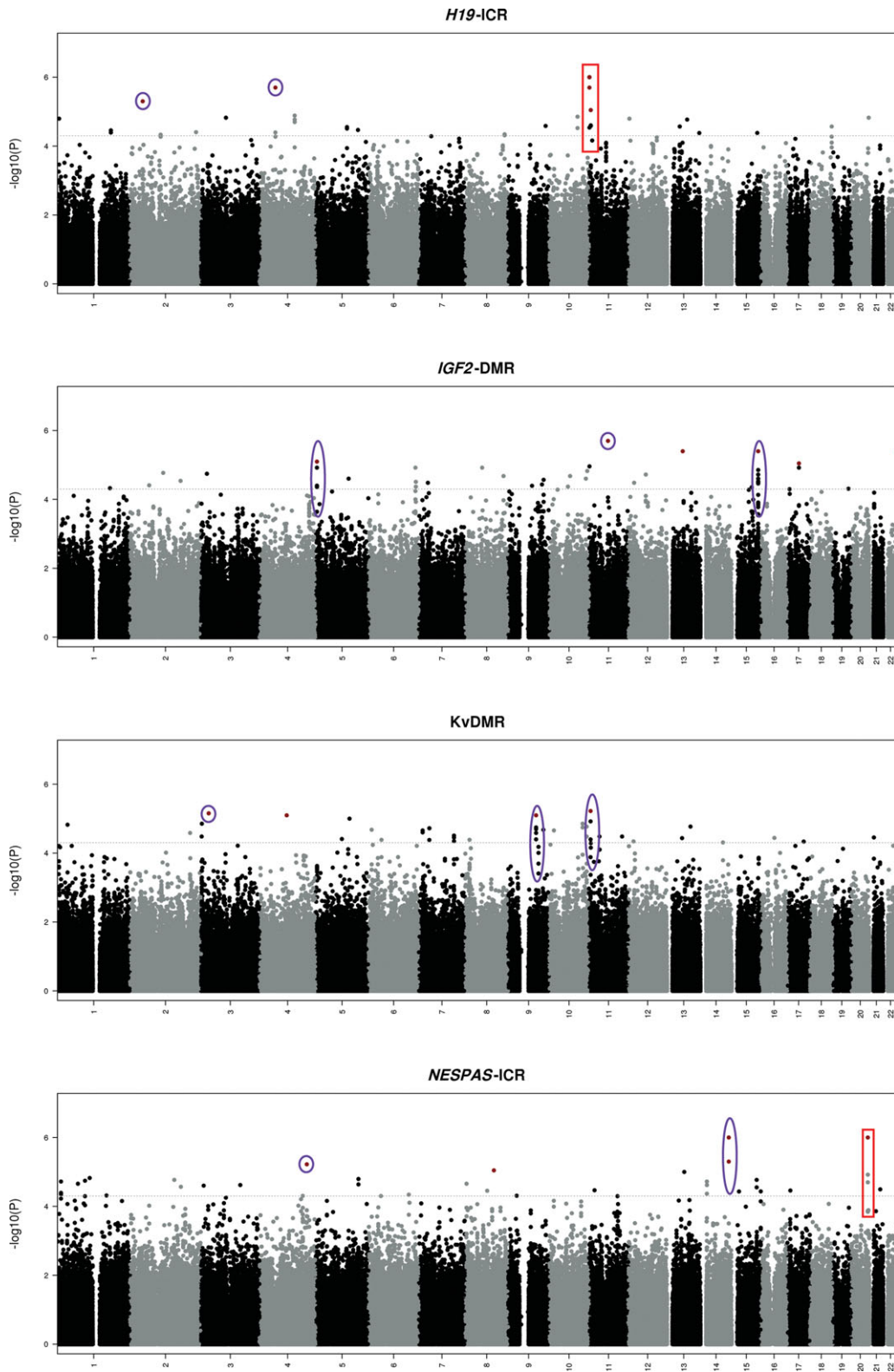
FIGURE 1

(Colour online) Manhattan plots showing results of multivariate GWAS for the four analyzed ICRs. The *x* axis indicates chromosome numbers while the *y* axis shows significance of association (i.e., $-\log_{10}(P)$ value). SNPs with a *p* value < 5e-05 appear above the gray horizontal line threshold and SNPs with a *p* value < 1e-05 are shown in red. Suggestive associations located in adjacent regions to the interrogated ICRs are highlighted in red, whereas suggestive associations in other regions are indicated in purple.
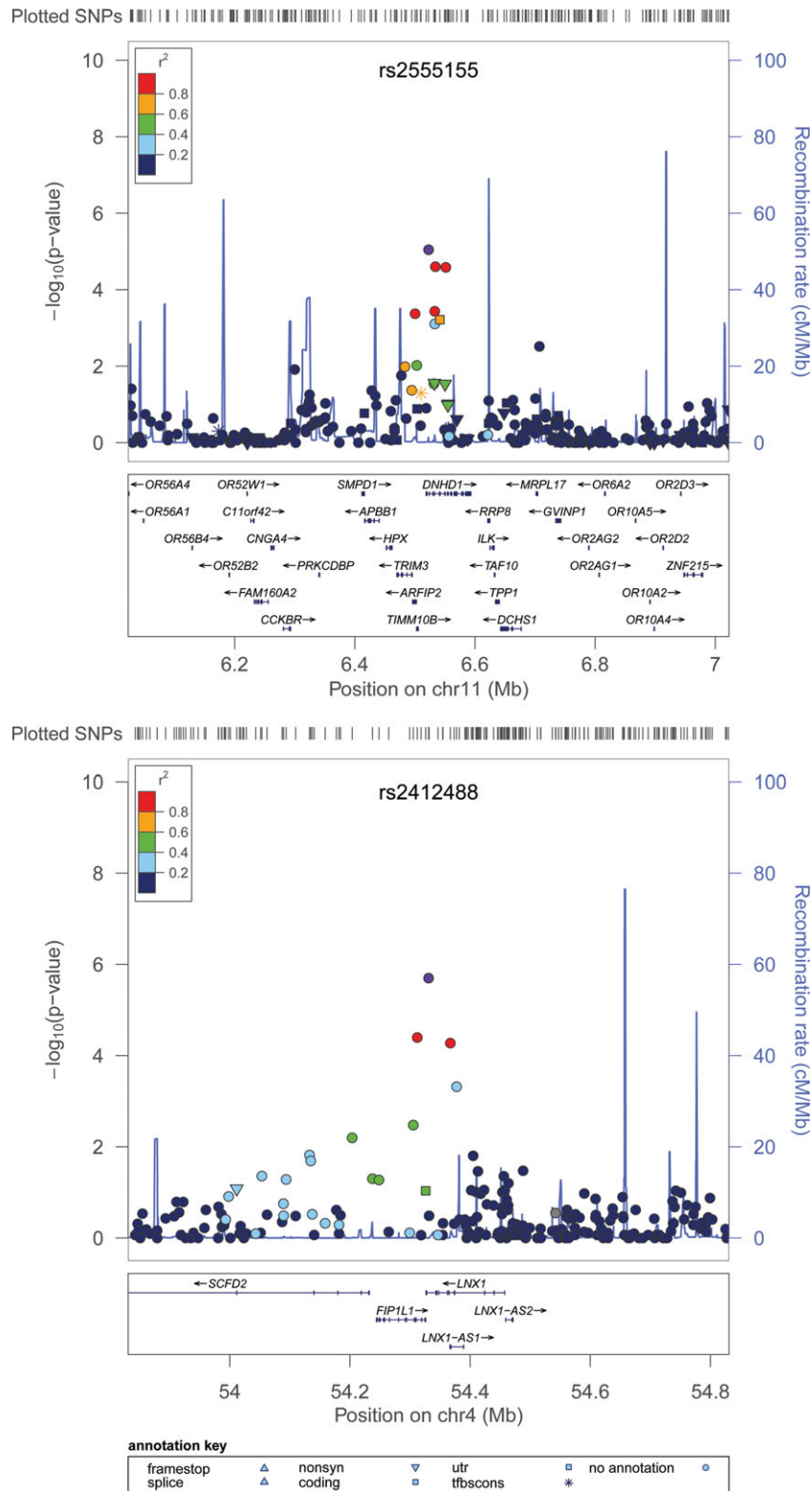
FIGURE 2

(Colour online) Locus zoom view of SNPs identified through multivariate GWAS located in adjacent regions to the interrogated ICRs. *H19*-ICR (rs4930103) and *NESPAS*-ICR (rs965808) displayed the most significant associations with intra-domain variation in DNA methylation. rs4930103 is in LD with SNPs rs2839701 and rs10732516, which we previously reported (see Coolen et al., 2011). On the other hand, rs965808 (purple dot in B) is a novel association located near the *GNAS* and *GNAS-AS1* imprinted genes, which are regulated by *NESPAS*-ICR.

FIGURE 3

(Colour online) Locus zoom view of SNPs identified through multivariate GWAS located in regions further apart from the *H19*-ICR. rs2555155 is located in chromosome 11, within an intronic region of genes *DNHD1* and *FXC1*, whereas rs2412488 is located in chromosome 4, within an intronic region of overlapping genes *FIP1L1* and *LNX1*.

TABLE 2

Characteristics of Most Significantly Associated SNPs at Each ICR Identified Through Multivariate GWAS

CHR	Minor/major	MAF	SNP	P_{hwe}	BP[a]	F test	p value[b]	p value[c]	Closest gene	Loadings[d]
						H19-ICR (11p15.5)				*H19*-ICR (CpG2, CpG3, CpG4, CpG5, CpG6, CpG8, CpG9-10, CpG11, CpG12, CpG21, CpG22, CpG23)
2	A/G	0.22	rs11897432	0.954	43817505	4.529	7.87E-07	5.00E-06	*THADA*	-0.4855, -0.6237, -0.3337, -0.5003, -0.2639, -0.2633, -0.3128, -0.1502, -0.2583, -0.173, -0.3326, -0.06849
4	A/G	0.29	rs2412488	0.953	54330867	4.50	8.94E-07	2.00E-06	*FIP1L1/LNX1*	-0.3328, -0.3147, -0.905, -0.7598, -0.7396, -0.6773, -0.8173, -0.5978, -0.6032, -0.4431, -0.6645, -0.5282
11	T/C	0.48	rs10769945	0.882	1985127	4.614	5.45E-07	2.00E-06	*MRPL23*, near H19/*IGF2*	-0.8035, -0.6038, -0.3786, -0.2034, -0.3181, -0.1789, 0.02524, -0.06634, -0.2161, -0.0735, -0.11, 0.09054
11	G/A	0.48	rs2285935	0.136	2014646	5.744	4.01E-09	1.00E-06	*H19*	-0.8484, -0.5631, -0.2756, -0.04986, -0.1874, -0.05344, 0.1527, 0.01536, -0.1399, -0.03018, -0.008095, 0.1266
11	T/C	0.32	rs217228	0.715	2014709	5.717	4.51E-09	1.00E-06	*H19*	-0.8296, -0.5523, -0.3212, -0.005915, -0.1488, -0.05715, 0.1298, -6.975e-05, -0.06985, -0.1397, 0.008058, 0.1274
11	C/T	0.49	rs2067051	0.109	2018168	8.642	1.46E-14	1.00E-06	*H19*	-0.8476, -0.5158, -0.276, -0.03118, -0.162, -0.07638, 0.1365, 0.06122, -0.1346, -0.08936, 0.008819, 0.1526
11	G/A	0.46	rs4930103	0.397	2024544	8.666	1.32E-14	1.00E-06	*H19*	-0.871, -0.5237, -0.3622, -0.1091, -0.2753, -0.157, 0.0489, -0.05559, -0.1923, -0.05659, -0.07786, 0.09102
11	G/A	0.47	rs2555155	0.117	6522804	4.085	5.32E-06	9.00E-06	*FXC1/DNHD1*	0.4656, 0.6446, 0.7284, 0.3774, 0.6593, 0.5153, 0.5563, 0.5073, 0.5028, 0.3861, 0.48, 0.6905
						IGF2-DMR (11p15.5)				*IGF2*-DMR (CpG1-2, CpG3, CpG4, CpG5, CpG6, CpG7, CpG8)
5	C/T	0.24	rs10462794	0.7842	4488245	5.465	5.41E-06	8.00E-06	*RP11-44503.2*	-0.1625, -0.3463, 0.3184, -0.1714, -0.4112, -0.2408, 0.1153
11	A/C	0.38	rs11227306	0.374	65578672	6.628	2.14E-07	2.00E-06	*CFL1/OVOL1*	0.1443, 0.5829, 0.665, 0.5412, 0.6529, 0.6974, 0.3126
13	A/G	0.06	rs9596905	0.834	54633373	5.76	2.37E-06	4.00E-06	intergenic/near-*LINC00458*	-0.09329, 0.1827, -0.2948, -0.2928, 0.2002, 0.3201, 0.4537
15	A/G	0.46	rs4304977	0.936	92806291	5.52	4.63E-06	4.00E-06	intergenic, *SLCO3A1/ST8SIA1*	-0.02869, -0.3286, -0.5867, -0.5518, -0.7945, -0.5065, -0.179
17	T/C	0.13	rs1007190	0.510	43038209	5.782	2.28E-06	9.00E-06	*C1QL1*	-0.3624, 0.1497, -0.5053, -0.3584, -0.04687, -0.231, 0.2853
22	G/T	0.32	rs1004689	0.467	48651997	5.876	1.72E-06	4.00E-06	intergenic	0.4287, 0.3246, 0.7321, 0.5124, 0.9169, 0.5351, 0.3182
						KvDMR (11p15.5)				KvDMR (CpG1, CpG6, CpG8-9, CpG10-11-12, CpG15, CpG16, CpG17-18, CpG20, CpG21, CpG24, CpG25)
3	G/A	0.12	rs7644516	1.000	25922285	4.331	4.13E-06	7.00E-06	*AC103588.1*	-0.04252, 0.4142, 0.007791, 0.3882, 0.4024, 0.08718, 0.3081, 0.1314, 0.2512, 0.2752, 0.3889
4	A/G	0.04	rs11933531	1.000	92415529	5.567	2.77E-08	8.00E-06	*FAM190A*	-0.05252, -0.09516, -0.2439, -0.02728, 0.1444, 0.391, 0.02211, -0.03128, -0.1318, -0.2175, 0.08515
9	C/T	0.35	rs7027203	0.040	96562563	4.243	5.87E-06	8.00E-06	intergenic	-0.5151, -0.3485, -0.251, 0.0467, -0.3155, -0.3545, -0.1084, -0.5354, -0.261, -0.1058, -0.05867
11	C/A	0.27	rs3858526	0.315	5959757	4.414	2.95E-06	6.00E-06	*TRIM5*	-0.06448, -0.6113, -0.05979, 0.023, -0.3942, -0.322, -0.4447, -0.3664, -0.2097, -0.2093, -0.1865

Miguel E. Rentería et al.

TABLE 2
Continued.

CHR	Minor/major	MAF	SNP	P_{hwe}	BP[a]	F test	p value[b]	p value[c]	Closest gene	Loadings[d] **NESPAS-ICR (CpG2, CpG5-6-7, CpG8-9, CpG10-11, CpG12-17, CpG14, CpG15-16, CpG18, CpG19, CpG20, CpG21)**
										NESPAS-ICR (20q13.32)
4	G/A	0.12	rs17261688	1.000	161028817	4.386	3.29E-06	1.00E-06	Intergenic	0.01868, 0.2265, 0.0007274, 0.6085, -0.1569, -0.1646, 0.4127, 0.4256, 0.1548, 0.1586, 0.2464
8	A/C	0.11	rs3763558	0.765	97796731	4.68	1.01E-06	1.00E-06	PGCP	-0.4908, -0.461, -0.05607, -0.3733, 0.2106, 0.331, -0.4525, -0.2191, -0.2118, -0.09158, -0.03833
14	C/T	0.33	rs724210	0.362	97523745	4.729	8.27E-07	1.00E-06	intergenic	-0.2366, -0.13, 0.2702, -0.6275, -0.01265, -0.1335, -0.131, -0.1713, -0.396, -0.3241, -0.05297
14	C/T	0.34	rs1022588	1.000	97529354	4.31	4.47E-06	1.00E-06	intergenic	-0.2957, -0.1535, 0.2584, -0.5947, -0.01368, -0.2264, -0.1567, -0.2156, -0.4092, -0.3785, -0.04211
20	**G/T**	**0.22**	**rs965808**	**0196**	**57408426**	**5.964**	**5.47E-09**	**1.00E-06**	**GNAS, GNAS-AS1**	**-0.1193, -0.004594, -0.1667, 0.1185, -0.209, -0.08219, 0.5897, -0.1618, -0.0384, -0.1876, 0.06337**

Note: SNPs located in neighbouring regions to relevant DMRs are highlighted in bold. [a] Coordinates according to GRCh37.p5; [b] p value before correcting for relatedness; [c] p value corrected for relatedness; [d] Loadings reflect contribution of individual traits (independent CpGs within the same DMR) to multivariate association with the given SNP.

origin. This was to expand on our previous finding where rs10732516 and co-inherited rs2839701 show a strong association over the DNA methylation status of specific CpG sites within the *H19*-ICR, but only when inherited from the maternal allele. A summary of most significantly associated SNPs is shown in Table 3.

As expected, genome-wide parent-of-origin association analysis also detected the previously described association of rs4930103 within the *H19*-ICR region haplotype that contains binding sites for the insulator protein CTCF (Coolen et al., 2011). *H19*-ICR-CpG02 showed the highest parent-of-origin association with rs4930103 ($p = 5.00e\text{-}16$). The A allele displayed an association effect (β) of -0.129 standard deviations (*SD*), when inherited from the paternal side, compared with 1.189 *SD* when inherited from the maternal side This is equivalent to -0.01 and 0.12 (in a scale of 0 to 1.0) in methylation ratio units, respectively. Likewise, SNP rs2285935 (in LD with rs4930103, $r^2 = 0.854$) also reached genome-wide significance association with *H19*-ICR-CpG04 ($p = 2.00e\text{-}11$), whereas rs2067051 was the most significant SNP associated with methylation at *H19*-ICR_CpG06 and *H19*-ICR_CpG22 (1.00e-08).

SNP rs10012307, located in an intergenic region of chromosome 4 near *TERF1P3* (telomeric repeat binding factor 1 pseudogene 3), was associated with methylation at *H19*-ICR_CpG_09-10. Interestingly, varying degrees of significance were observed for SNPs in this region, going from no association to suggestive association to genome-wide significant. Figure 4 shows variation in parent-of-origin association results observed across different CpGs of the *H19*-ICR highlighting regions around rs10012307 and rs4930103.

The most significantly SNP associated with methylation at *IGF2*-DMR_CpG01 identified through parent-of-origin association was rs4819833 (6.00e-08), which mapped on chromosome 22 neighboring the *SEPT5*, *GP1BB*, and *AC000093.3.1* genes. Furthermore, SNP rs7931462, located within *RPL26P31* (ribosomal protein L26 pseudogene 31) gene in chromosome 11 showed association with DNA methylation at KvDMR_10-11-12 ($p = 2.00E\text{-}09$). Currently, the function of these genes is unknown.

Finally, methylation ratios at KvDMR_21, NES-PAS_CpG08_09, and NESPAS_CpG02 showed suggestive associations with rs10497324, rs10934011, and rs13135284, which locate in intergenic regions of chromosomes 2, 3, and 4, respectively ($p = 2.00e\text{-}09$, 1.00e-09, and 5.00e-08).

Discussion

Unlike the genome, which is determined at the time of conception and remains nearly unchanged across all cells of the body during the life span of an individual, the epigenome is greatly variable from tissue to tissue (Feinberg, 2007; Mill et al., 2008; Petronis et al., 2003). Dynamic changes can result either from carefully directed epigenetic reprogramming processes during development or from

TABLE 3

Most Significantly Associated SNPs Identified Through Univariate Parent-of-Origin GWAS

Methylation site	SNP	Chr	All	MAF	P_{hwe}	p Beta[a]	m Beta[b]	χ^2	p value[c]	Associated gene
H19-ICR-CpG02	**rs4930103**	**11**	**A**	**0.49**	**0.40**	**-0.129**	**1.189**	**70.289**	**5.00E-16**	**H19**
H19-ICR-CpG04	**rs2285935**	**11**	**A**	**0.48**	**0.14**	**-0.508**	**1.254**	**49.306**	**2.00E-11**	**H19/MIR675**
H19-ICR_CpG06	**rs2067051**	**11**	**T**	**0.49**	**0.11**	**-0.477**	**1.123**	**36.385**	**1.00E-08**	**H19**
H19-ICR_CpG_09-10	rs10012307	4	T	0.06	0.65	0.295	-1.693	35.015	2.00E-08	intergenic
H19-ICR_CpG22	**rs2067051**	**11**	**T**	**0.49**	**0.11**	**-0.548**	**1.021**	**33.004**	**7.00E-08**	**H19**
IGF2-DMR_CpG01	rs4819833	22	C	0.34	0.42	0.486	-0.982	33.184	6.00E-08	SEPT5, GP1BB, TBX1, AC000093.3
KvDMR_10-11-12	**rs7931462**	**11**	**G**	**0.02**	**0.32**	**-0.68**	**-0.059**	**40.179**	**2.00E-09**	**RPL26P31**
KvDMR_21	rs10497324	2	A	0.04	0.32	-0.016	0.639	40.34	2.00E-09	intergenic
NESPAS_CpG02	rs13135284	4	C	0.26	0.52	-0.589	-0.019	33.64	5.00E-08	intergenic
NESPAS_CpG08_09	rs10934011	3	G	0.21	0.39	-0.827	0.507	41.255	1.00E-09	intergenic

Note: SNPs located in neighboring regions to relevant DMRs are highlighted in bold.
[a]Paternal effect;[b]Maternal effect;[c]p value (correcting for relatedness).

environmental exposure, aging or stochastic factors (Horsthemke, 2006; Schanen, 2006; Schneider et al., 2010; Siegmund et al., 2007; Waterland & Jirtle, 2003).

While the dynamic nature of the epigenome is widely acknowledged, little is known about what drives this variation. Some studies have highlighted the importance of the environment while others have focused on the effects of genetic variation. In the present study, we applied two different types of genome-wide scans to look for genetic variants that contribute either to individual variation in the DNA methylation status within four imprinting domains or to single CpG methylation variation in a parent-of-origin fashion.

The most significantly associated SNPs found through multivariate GWAS lie in the neighboring regions of the analyzed imprinting regions: rs2067051 with *H19*-ICR and rs965808 with *NESPAS*-ICR, which provides support for our approach. The *H19* region contains binding sites for CTCF, a critical zinc-finger CCCTC-binding factor, and we speculate that this polymorphism may directly affect the binding affinity of the insulator protein CTCF to this region (Coolen et al., 2011). Notably, we were also able to detect SNPs located further from the interrogated ICRs that could be associated with methylation variation across CpGs of certain ICRs. It is worth noting that the significance of most significantly associated SNPs was of comparable level, regardless of their location (i.e., neighboring the interrogated regions or on different chromosomes). However, the role of SNPs located on different chromosomes is not evident, given that most of such variants lie within intronic regions or functionally uncharacterized intergenic regions.

SNP rs2067051 is of particular interest given that two prior independent studies (Adkins et al., 2010b; Petry et al., 2011) found association between the haplotypes that contain rs2067051 and birth weight. A third study that looked specifically at the association of birth weight with polymorphisms within and around the *IGF2*, *H19*, and *IGF2R* genes found that rs2067051, together with rs2251375 and rs4929984, is strongly associated with birth weight. Subsequent analysis also determined that association of the

maternal genotype with newborn birth weight was due to parent-of-origin effects and not to direct maternal or direct newborn effects (Adkins et al., 2010b). *IGF2* is believed to be a major fetal growth factor, and *H19*-ICR plays a crucial role in regulating transcript levels of both *H19* and *IGF2*. Hence, altogether, these studies support the link between parent-of-origin DNA methylation at *H19*-ICR and birth weight. Interestingly, maternally transmitted alleles within the *GNAS* gene, which is regulated by NESPAS-ICR, had also been associated previously with birth weight in male newborns (Adkins et al., 2010a). This was later supported by a recent study which analyzed ~20,000 CpG sites and found that methylation of genes involved in metabolism and biosynthesis was associated with birth weight (Gordon et al., 2012). However, such study did not include analysis of parent-of-origin effects and was performed in a significantly smaller dataset ($N = 68$; Gordon et al., 2012).

Genetic imprinting generally indicates that a selective advantage exists for the strict maintenance of allelic silencing of a gene or a discrete set of genes. Results from recent studies suggest that only a fraction of imprinted genes have been documented. For instance, Gregg et al. (2010b) described over 1,300 loci with parental bias in the expression of individual genes and of specific transcript isoforms in the mouse brain. The differences between different brain regions were just as striking. For example, they reported preferential maternal contribution to gene expression in the developing brain and a major paternal contribution in the adult brain. In another paper, Gregg et al. (2010a) also described imprinting differences between males and females. Future studies could test whether parent-of-origin genetic effects on methylation also contribute to regulate individual variation in gene expression levels of these newly discovered imprinted genes.

We acknowledge that our study presents some limitations. First, the relatively moderate associations between SNPs and DNA methylation, together with the relatively small sample size of the study, provide a limited statistical power to detect associations beyond the genome-wide significance threshold. However, by incorporating
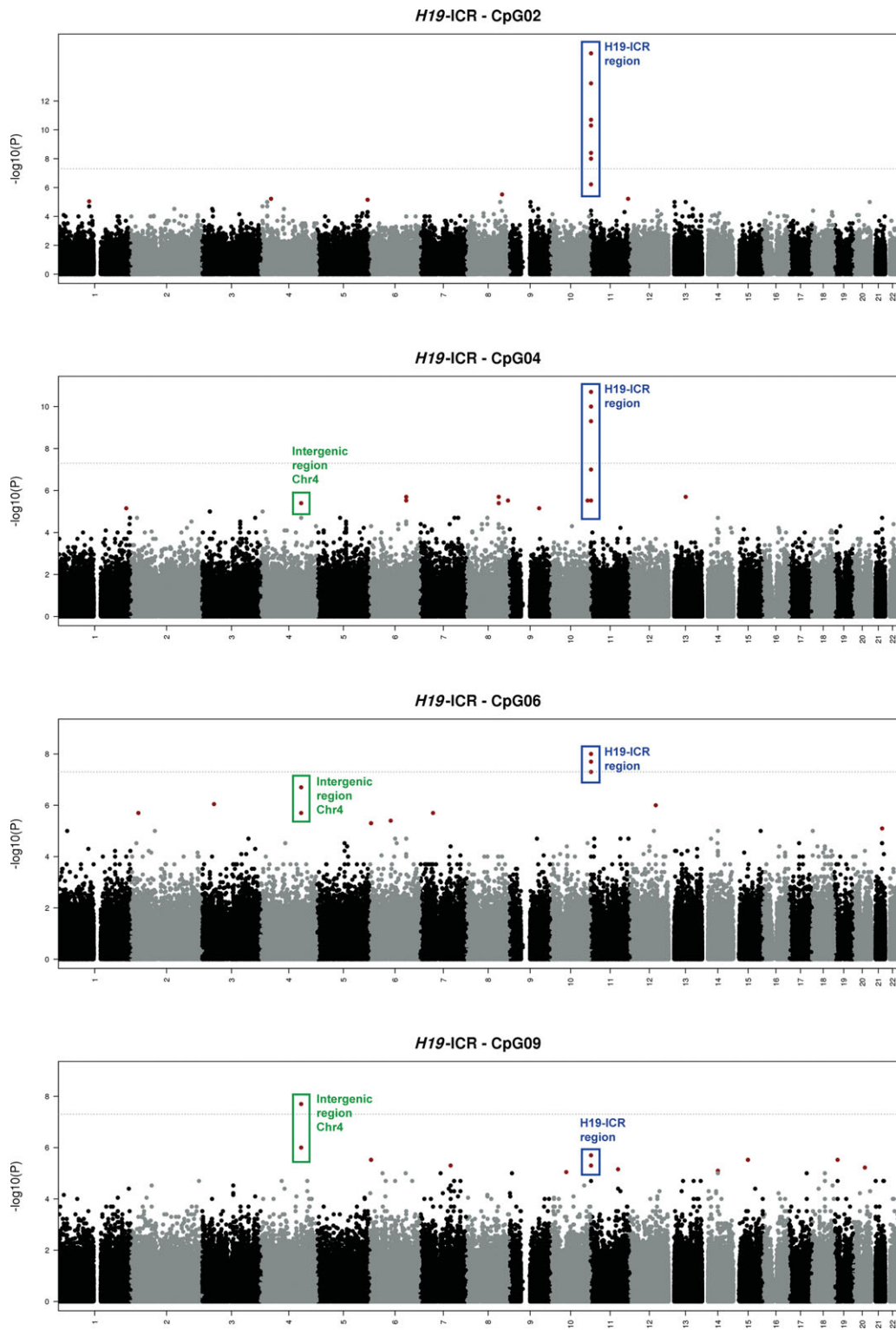
FIGURE 4

(Colour online) Manhattan plots showing results for parent of origin GWAS within the *H19*-ICR. The *x* axis shows chromosome numbers while the *y* axis indicates significance of association (i.e., −log$_{10}$(P) value). SNPs neighboring the H19-ICR region are highlighted in blue, whereas region highlighted in green indicates a region located in chromosome 4 for which varying degrees of significance were observed, and the GWAS significance threshold (rs2412488) was surpassed with *H19*-CpG09. The horizontal line indicates the 5e-08 threshold and SNPs with a *p* value equal or lower than 1e-05 are colored in red.

multivariate and family-based models of association, we were able to make a better use of the evidence contained in both the genotype and phenotype data. Furthermore, the fact that we were able to detect associations between ICRs and SNPs located in neighboring regions, via a hypothesis-free approach such as GWAS, also reduces uncertainty about the validity of our results, and makes it plausible that associations found in other regions are also valid. Our analysis also revealed that heterogeneity of associations across CpGs within the same ICR is rather high. As shown by the loading factors observed in multivariate association analysis (Table 2), a given genetic polymorphism might affect DNA methylation at some but not all CpGs within a region (i.e., an ICR). Our approach was designed to only detect those SNPs that display either a consistent association across all or the majority of the surveyed methylation sites within a region, or large parent-of-origin-dependent variation. Thus, it was not able to detect SNPs that have a significant association on specific individual CpGs.

Conclusions

With respect to previous studies, our analysis is the first to apply both multiple-trait GWAS and genome-wide parent-of-origin association approaches. Despite the relatively small sample size, we identified SNPs with *p* values that display suggestive associations. These include two SNPs located near the surveyed regions: the previously reported rs2067051 (associated with methylation at *H19*-ICR) and the novel rs965808 (associated with methylation at *NESPAS*-ICR). Furthermore, a number of other potentially associated SNPs were identified, a number of which display parent-of-origin association. Overall, our study provides a proof of principle for the use of alternative GWAS approaches to epigenetics. However, due to the customized design of the study, no replication cohort could be found.

Future studies with larger sample sizes should also include comparisons between methylation status across different tissues, and family-based studies could provide exciting information about the mechanisms underlying the parent-of-origin associations with DNA methylation and imprinting. However, an immeasurable complexity involving genetic and environmental factors is anticipated.

Acknowledgments

Authors' Contributions

MWC, MER, SEM, and SJC conceived and designed the experiments. MWC, ALS, WQ, SS, and MJC performed the experiments. MER, MWC, and ALS analyzed the data. AKH, GWM, and NGM contributed reagents, materials, and analysis tools. MER, SMC, SJC, NGM, and SEM wrote the paper.

Competing Interests

The authors declare no competing interests.

References

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin — Rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, *30*, 97–101.

Adkins, R. M., Krushkal, J., Magann, E. F., Klauser, C. K., Morrison, J. C., Ramsey, R., & Somes, G. (2010a). Association of maternally inherited GNAS alleles with African-American male birth weight. *International Journal of Pediatric Obesity*, *5*, 177–184.

Adkins, R. M., Somes, G., Morrison, J. C., Hill, J. B., Watson, E. M., Magann, E. F., & Krushkal, J. (2010b). Association of birth weight with polymorphisms in the IGF2, H19, and IGF2R genes. *Pediatric Research*, *68*, 429–434.

Bataille, V., Snieder, H., MacGregor, A. J., Sasieni, P., & Spector, T. D. (2000). Genetics of risk factors for melanoma: An adult twin study of nevi and freckles. *Journal of the National Cancer Institute*, *92*, 457–463.

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., . . . Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology*, *12*(1), R10.

Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, T., Cui, H., . . . Feinberg, A. P. (2008). Intra-individual change over time in DNA methylation with familial clustering. *JAMA*, *299*, 2877–2883.

Boks, M. P., Derks, E. M., Weisenberger, D. J., Strengman, E., Janson, E., Sommer, I. E., . . . Ophoff, R. A. (2009). The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*, *4*(8), e6767.

Bolormaa, S., Pryce, J. E., Hayes, B. J., & Goddard, M. E. (2010). Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science*, *93*, 3818–3833.

Chow, J. C., Yen, Z., Ziesche, S. M., & Brown, C. J. (2005). Silencing of the mammalian X chromosome. *Annual Review of Genomics and Human Genetics*, 6, 69–92.

Constancia, M., Kelsey, G., & Reik, W. (2004). Resourceful imprinting. *Nature*, 432, 53–57.

Coolen, M. W., Statham, A. L., Gardiner-Garden, M., & Clark, S. J. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: Critical evaluation and improvements. *Nucleic Acids Research*, 35, e119.

Coolen, M. W., Statham, A. L., Qu, W., Campbell, M. J., Henders, A. K., Montgomery, G. W., . . . Clark, S. J. (2011). Impact of the genome on the epigenome is manifested in DNA methylation patterns of imprinted regions in monozygotic and dizygotic twins. *PLoS One*, 6, e25590.

Cui, H., Cruz-Correa, M., Giardiello, F. M., Hutcheon, D. F., Kafonek, D. R., Brandenburg, S., . . . Feinberg, A. P. (2003). Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science*, 299, 1753–1755.

Delaval, K., & Feil, R. (2004). Epigenetic regulation of mammalian genomic imprinting. *Current Opinion in Genetics and Development*, 14, 188–195.

Eggermann, T., Eggermann, K., & Schonherr, N. (2008). Growth retardation versus overgrowth: Silver–Russell syndrome is genetically opposite to Beckwith–Wiedemann syndrome. *Trends in Genetics*, 24, 195–204.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447, 433–440.

Ferreira, M. A., & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics*, 25, 132–133.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., . . . Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 10604–10609.

Gibbs, J. R., van der Brug, M. P., Hernandez, D. G., Traynor, B. J., Nalls, M. A., Lai, S. L., . . . Singleton, A. B. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genetics*, 6, e1000952.

Gicquel, C., Rossignol, S., Cabrol, S., Houang, M., Steunou, V., Barbu, V., . . . Le Bouc, Y. (2005). Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver–Russell syndrome. *Nature Genetics*, 37, 1003–1007.

Gordon, L., Joo, J. E., Powell, J. E., Ollikainen, M., Novakovic, B., Li, X., . . . Saffery, R. (2012). Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Research*, 22, 1395–1406.

Gregg, C., Zhang, J., Butler, J. E., Haig, D., & Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. *Science*, 329, 682–685.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., & Dulac, C. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, 329, 643–648.

Heijmans, B. T., Kremer, D., Tobi, E. W., Boomsma, D. I., & Slagboom, P. E. (2007). Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Human Molecular Genetics*, 16, 547–554.

Horsthemke, B. (2006). Epimutations in human disease. *Current Topics in Microbiology and Immunology*, 310, 45–59.

Inouye, M., Silander, K., Hamalainen, E., Salomaa, V., Harald, K., Jousilahti, P., . . . Peltonen, L. (2010). An immune response network associated with blood lipid levels. *PLoS Genetics*, 6(9), e1001113.

Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., Oh, G. H., Wong, A. H., . . . Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics*, 41, 240–245.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., . . . Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development*, 129, 1807–1817.

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, 366, 362–365.

Liu, J., Chen, M., Deng, C., Bourc'his, D., Nealon, J. G., Erlichman, B., . . . Weinstein, L. S. (2005). Identification of the control region for tissue-specific imprinting of the stimulatory G protein alpha-subunit. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 5513–5518.

Liu, Y. Z., Pei, Y. F., Liu, J. F., Yang, F., Guo, Y., Zhang, L., . . . Deng, H. W. (2009). Powerful bivariate genome-wide association analyses suggest the SOX6 gene influencing both obesity and osteoporosis phenotypes in males. *PLoS One*, 4, e6827.

McEvoy, B. P., Montgomery, G. W., McRae, A. F., Ripatti, S., Perola, M., Spector, T. D., . . . Visscher, P. M. (2009). Geographical structure and differential natural selection among North European populations. *Genome Research*, 19(5), 804–814.

Mill, J., Dempster, E., Caspi, A., Williams, B., Moffitt, T., & Craig, I. (2006). Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 141B, 421–425.

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., . . . Petronis, A. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *American Journal of Human Genetics*, 82, 696–711.

Nakano, S., Murakami, K., Meguro, M., Soejima, H., Higashimoto, K., Urano, T., . . . Oshimura, M. (2006). Expression profile of LIT1/KCNQ1OT1 and epigenetic status at the KvDMR1 in colorectal cancers. *Cancer Science*, 97, 1147–1154.

Numata, S., Ye, T., Hyde, T. M., Guitart-Navarro, X., Tao, R., Wininger, M., . . . Lipska, B. K. (2012). DNA methylation signatures in development and aging of the human

prefrontal cortex. *American Journal of Human Genetics*, *90*(2), 260–272.

Petronis, A., Gottesman, I. I., Kan, P., Kennedy, J. L., Basile, V. S., Paterson, A. D., & Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: Clues to twin discordance? *Schizophrenia Bulletin*, *29*, 169–178.

Petry, C. J., Seear, R. V., Wingate, D. L., Acerini, C. L., Ong, K. K., Hughes, I. A., & Dunger, D. B. (2011). Maternally transmitted foetal H19 variants and associations with birth weight. *Human Genetics*, *130*, 663–670.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, *6*, 597–610.

Schanen, N. C. (2006). Epigenetics of autism spectrum disorders. *Human Molecular Genetics*, *15*, R138–R150.

Schneider, E., Pliushch, G., El Hajj, N., Galetzka, D., Puhl, A., Schorsch, M., . . . Haaf, T. (2010). Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. *Nucleic Acids Research*, *38*, 3880–3890.

Siegmund, K. D., Connor, C. M., Campan, M., Long, T. I., Weisenberger, D. J., Biniszkiewicz, D., . . . Akbarian, S. (2007). DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS One*, *2*(9), e895.

Smilinich, N. J., Day, C. D., Fitzpatrick, G. V., Caldwell, G. M., Lossie, A. C., Cooper, P. R., . . . Higgins, M. J. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith–Wiedemann syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 8064–8069.

Spielman, R. S., & Ewens, W. J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. *American Journal of Human Genetics*, *59*, 983–989.

Takai, D., Gonzales, F. A., Tsai, Y. C., Thayer, M. J., & Jones, P. A. (2001). Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Human Molecular Genetics*, *10*, 2619–2626.

Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T., & Miyamoto, K. (2003). Fidelity of the methylation pattern and its variation in the genome. *Genome Research*, *13*(5), 868–874.

Waterland, R. A., & Jirtle, R. L. (2003). Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Molecular and Cellular Biology*, *23*, 5293–5300.

Weber, M., & Schubeler, D. (2007). Genomic patterns of DNA methylation: Targets and function of an epigenetic mark. *Current Opinion in Cell Biology*, *19*, 273–280.

Weksberg, R., Nishikawa, J., Caluseriu, O., Fei, Y. L., Shuman, C., Wei, C., . . . Squire, J. (2001). Tumor development in the Beckwith–Wiedemann syndrome is associated with a variety of constitutional molecular 11p15 alterations including imprinting defects of KCNQ1OT1. *Human Molecular Genetics*, *10*, 2989–3000.

Weksberg, R., Shuman, C., Caluseriu, O., Smith, A. C., Fei, Y. L., Nishikawa, J., . . . Squire, J. (2002). Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith–Wiedemann syndrome. *Human Molecular Genetics*, *11*, 1317–1325.

Wright, M., De Geus, E., Ando, J., Luciano, M., Posthuma, D., Ono, Y., . . . Boomsma, D. (2001). Genetics of cognition: Outline of a collaborative twin study. *Twin Research*, *4*, 48–56.

Zeschnigk, M., Schmitz, B., Dittrich, B., Buiting, K., Horsthemke, B., & Doerfler, W. (1997). Imprinted segments in the human genome: Different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Human Molecular Genetics*, *6*, 387–395.