Article type : Research Article

Identification of methylation changes associated with positive and negative growth deviance in Gambian infants using a targeted methyl sequencing approach of genomic DNA

Claire R. Quilter[1], Kerry M. Harvey[1], Julien Bauer[1], Benjamin M. Skinner[1,2], Maria Gomez[1], Manu Shrivastava[1], Andrew M. Doel[5,6], Saikou Drammeh[7] , David B. Dunger[3], Sophie E. Moore[5,6], Ken K. Ong[3,4,9], Andrew M. Prentice[6], Robin M. Bernstein[7,8], Carole A. Sargent[1] and Nabeel A. Affara.[1] (Corresponding author)

[1]Department of Pathology, University of Cambridge, Cambridge, UK

[1,2]School of Life Sciences, University of Essex, Colchester, UK

[3]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK

[4]Department of Paediatrics, University of Cambridge School of Clinical Medicine, Cambridge, UK

[5]Department of Women and Children's Health, King's College London, London, UK

[6]MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Banjul, The Gambia

[7]Growth and Development Lab, Department of Anthropology, University of Colorado, Boulder, CO, USA

[8]Institute of Behavioural Science, University of Colorado, Boulder, CO, USA

[9]Institute of Metabolic Science, Cambridge Biomedical Campus Cambridge

Corresponding Author: Telephone Office: 44 1223 333700
Professor Nabeel A. Affara Telephone Mobile: 44 780 182 1743
University of Cambridge e-mail: na106@cam.ac.ukDepartment of Pathology
Tennis Court Road
Cambridge CB2 1QP, United Kingdom

Claire Quilter current address:
East Midlands & East of England NHS Genomic Laboratory Hub
Genomics Laboratories
Cambridge University Hospitals NHS Foundation Trust
Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, United Kingdom

Maria Gomez current address:
Kennedy Institute of Rheumatology
University of Oxford
Roosevelt Drive
Oxford OX3 7FY
UK

Manu Shrivastava current address:
Manu Kunaal Shrivastava
Oxford University Hospitals
Oxford OX3 9DU
UK

Running Title: DNA methylation changes and positive and negative growth

Abbreviations

LAZ = Length for Age Z-score
DMR= Differentially Methylated Region
CTCF= CCCTC-binding factor
UGR= Uterine Growth Restriction
SGA= Small for Gestation Age
EWAS= Epigenome Wide Association Study
GWAS= Genome Wide Association Study
FDR= False Discovery Rate
MRC= Medical Research Council
SNP= Single Nucleotide Polymorphism
TSS= Transcription Start Site
MAF= Minor Allele Frequency
GAD= Genetic Association Database
DAVID= Database for Annotation, Visualization and Integrated Discovery
EBI= European Bioinformatics Institute
PCA= Principal Components Analaysis
NCBI= National Center for Bioinformatics Technology
OMIM= Online Mendelian Inheritance in Man
Trans-meQTL= Trans-acting methylation quantitative trait locus
Cis-meQTL= Cis-acting methylation quantitative trait locus

Cis-eQTM= Cis-acting quantitative trait methylation

Abstract

Low birthweight and reduced height gain during infancy (stunting) may arise at least in part from adverse early life environments that trigger epigenetic reprogramming that may favour survival. We examined differential DNA methylation patterns using targeted methyl sequencing of regions regulating gene activity in groups of rural Gambian infants: (a) low and high birthweight (DNA from cord blood (n=16 and n=20 respectively), from placental trophoblast tissue (n=21 and n=20 respectively) and DNA from peripheral blood collected from infants at 12 months of age (n=23 and n=17 respectively)), and, (b) the top 10% showing rapid postnatal length gain (high, n=20) and the bottom 10% showing slow postnatal length gain (low, n=20) based on z-score change between birth and 12 months of age (LAZ) (DNA from peripheral blood collected from infants at 12 months of age). Using BiSeq analysis to identify significant methylation marks, for birthweight, four differentially methylated regions (DMRs) were identified in trophoblast DNA, compared to 68 DMRs in cord blood DNA, and 54 DMRs in 12-month peripheral blood DNA. Twenty-five DMRs were observed to be associated with high and low length for age (LAZ) at 12 months. With the exception of five loci (associated with two different genes), there was no overlap between these groups of methylation marks. Of the 194 CpG methylation marks contained within DMRs, 106 were located to defined gene regulatory elements (promoters, CTCF binding sites, transcription factor binding sites and enhancers), 58 to gene bodies (introns or exons) and 30 to intergenic DNA. Distinct methylation patterns associated with birth weight between comparison groups were observed in DNA collected at birth (at the end of intrauterine growth window) compared to those established by 12 months (near the infancy/childhood growth transition). The longitudinal differences in methylation patterns may arise from methylation adjustments, changes in cellular composition of blood or both that continue during the critical postnatal growth period and in response to early nutritional and infectious environmental exposures with impacts on growth and longer-term health outcomes.

Key Words: Birthweight, Stunting, DNA Methylation, Environmental Exposures

Introduction

About 45% of global deaths in children under 5 years of age are thought to be related to undernutrition (1). Children who survive early periods of undernutrition may suffer longer-term consequences, including stunting and other developmental deficits (2), which are major contributors to long term morbidity and mortality (3, 4). Although the prevalence of stunting declined in sub-Saharan Africa from 42% in 1990 to 32% in 2015, the numbers of affected individuals increased from 47 million to 58 million (5). Studies estimate that 20% of growth retardation starts *in utero* where under-nutrition in pregnancy increases the risks of intrauterine growth retardation (IUGR) and small for gestation age (SGA) infants, pre-term delivery (6) and long-term impaired immunity. It is hypothesized that an adverse early life environment and nutrition induce phenotypic adaptations through developmental plasticity (7) to favour survival in the short-term, but at the expense of lifelong effects on health (8, 9).

Nutritional interventions to improve child growth and adult health (10, 11) have had limited success, primarily for the lack of a clear understanding of optimal timing, target groups and the composition of supplements. The period of growth and development from conception to a child's second birthday (coined the first 1000 days) is one of the most critical windows of opportunity for interventions (12). There is a complex interplay between an individual's genetic constitution and the environment. Responses to extrinsic factors via modifications to the epigenome (which may include to both chromatin associated proteins and DNA bases) in the first 1000 days are believed to be important in establishing protective adaptations against the impact of under-nutrition and an adverse environment (thrifty phenotype) (13, 14). DNA methylation at CpG couplets is one of the most actively studied modifications to the epigenome.

A large meta-analysis of multiple epigenome wide association studies (EWAS) by the Childhood Epigenetics Consortium found methylation at 914 CpG sites associated with birthweight in whole blood DNA from healthy neonates, but < 1.3% persisted in children (2-13 years), <0.1% in adolescents (16-18 years) and none in adults (30-45 years) (15). The current study uses samples and data from a cohort of Gambian mother-infant pairs exhibiting high rates of maternal and child under-nutrition. Rural Gambian infants are small at birth relative to international standards, show positive growth patterns during the first few months of life and then enter a period of reduced growth marked by profound faltering until at least 24 months of age (16, 17). Schoenbuchner et al. (16) have suggested that stunting is an extreme adaptation to profound faltering episodes potentially arising from a complex interaction of malnutrition, infection and disease. Despite four decades of nutrition-sensitive and nutrition-specific interventions halving under-nutrition for young children from rural Gambia, substantial (30%) growth faltering remains (17), indicating a gap in our understanding of its complex aetiology.

Epigenetic studies carried out on Gambian populations have highlighted the importance of maternal nutrition and exposures and the effects of maternal nutritional supplementation in this highly seasonal environment. Many aspects of health and behaviour in rural Gambia are influenced by the annual seasonality with a single rainy 'hungry' season (late June – October) followed by a dry 'harvest' season (November-May/June) (18, 2). Specifically, there is evidence that seasonal variation in nutrition during the periconceptional period influences methylation status in postnatal infants at a number of loci (19), is related to methyl-donor nutrient content of the mother's diet (20, 21, 22, 23) and may be associated with an increase in both preterm and SGA infants (18). Periconceptional nutrition supplementation influences methylation changes in cord and postnatal infant blood DNA at CpG loci linked to genes associated with infection and immunity (24) and alters methylation at imprinted loci (25). Maternal exposure to aflatoxin B1 is also associated with DNA methylation changes at specific loci in Gambian infants (26).

The aim of the present study was to identify epigenetic marks that are established during the critical first 1000 days in a cohort of rural Gambian infants and explore how these may be associated with normal versus stunted growth outcomes in order to determine whether any targets for intervention are associated with pre- and/or post-natal periods of epigenetic modification. We used a targeted methyl sequencing approach of genomic DNA from placental trophoblast tissue, cord and infant (12 months of age) blood to identify methylation changes. These changes may be useful as biomarkers, highlighting genes influenced by exposures during embryonic and foetal development and early infancy, and identifying potential pathways through which these may influence growth outcomes at birth and in the first year of life.

Materials and Methods

Samples

The study was conducted among pregnant women and their infants living in the rural West Kiang region of The Gambia. Participants were recruited as part of the HERO-G (Hormonal Regulators of Growth) study. The study cohort was 238 newborns whose growth had been assessed longitudinally to 24 months of age. Table 1 summarises data associated with

the samples from individuals used in this study. The full HERO-G protocol is described elsewhere (27). Placentas from women who delivered at home were collected by trained field workers and immediately transported on ice to the nearby Medical Research Council (MRC) Unit The Gambia Keneba laboratory (within 20-30 minutes) and carefully processed to obtain trophoblast material following a standard protocol (see placenta sample collection protocol in supplementary materials). Placental samples each of 400mg were taken at four different evenly spaced locations, at least 2cm from the edge, and at consistent relative positions in each placenta to mitigate placental tissue heterogeneity. Samples were cut into four pieces, placed in RNAlater at a volume of 5 x tissue weight (Cat No 76106, Qiagen) and transported frozen on dry-ice to the UK for DNA extraction. After extraction samples from each of the four placental regions were pooled equimolarly. Cord blood and infant blood samples were collected into EDTA-lined tubes (BD Vacutainer, pink top) for DNA extraction in the UK. Ethical approval for the study was given by the joint Gambia Government/Medical Research Council (MRC) Unit The Gambia Ethics Committee (SCC 1313v3), with additional approval from the University of Colorado Institutional Research Board (protocol number 13-0441). Community approval was obtained from each participating village, and written, informed consent was obtained from each participating family. Samples for analysis were selected retrospectively from the study cohort representing (a) the highest 20% and lowest 20% birthweights and (b) according to the top and bottom 10% change in length-for-age (LAZ) from birth to 12 months. For the 12 month samples the male average age = 376.4 days, SD 9 days (366-409d) and females average age = 378.8 days, SD 10 days (367-413d). Table 2 summarises the number of samples analysed after quality testing for each tissue and test group and those that are common between groups.

Nucleic Acid Extraction

DNA for DNA methylation studies was extracted from tissues using the Quick-DNA Mini Prep Plus kit (Cat No. D4068, Zymo Research). DNA extracted from blood followed the Biological Fluids and Cells protocol and DNA extracted from placenta followed the Solid Tissue protocol. DNA abundance and quality were determined after extraction using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Absorbance ratios (A260/A280 and A60/A230) were above the recommended 1.8. DNA from each sample was further quantified on a Qubit® Fluorometer using Qubit® dsDNA HS Assay kit (Cat. No. Q32854, Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Methyl-Seq library preparation

Methyl-Seq was performed using the SureSelect ^{XT} Methyl-Seq kit (Cat. No. G9651B, Agilent, Santa Clara, California, USA) according to the manufacturer's protocol (SureSelect ^{XT} Methyl-Seq Target Enrichment System for Illumina Multiplexed Sequencing protocol, version C.0, January 2015); this covers over 3.7 million individual CpG dinucleotide sequences covering CpG islands, CpG island shores, CpG island shelves, under-methylated regions, promoters, enhancers, transcription factors, CTCF binding sites, DNase 1 hypersensitive sites and DMRs. Three µg of DNA from each sample was initially sheared by Covaris sonication to 150-200bp in size and used to prepare genomic DNA libraries with the SureSelect ^{XT} Methyl-Seq Library Prep kit. After hybridization with the SureSelect ^{XT} Methyl-Seq capture library, targeted regions were isolated using complementary RNA baits. Isolated targets were bisulfite converted using the EZ DNA Methylation-Gold^{TM} (Cat. No. D5005, Zymo Research) which converts unmethylated Cytosines to Uracil, whilst methylated Cytosines are

unaltered. Subsequent PCR amplification creates an unmethylated CG→TA transition at unmethylated positions. Each sequence-modified, target-enriched library preparation was attached to a readable index (short DNA identifying code) by PCR. Libraries were quantified on a Bioanalyzer 2100 (Agilent, Santa Clara, California, USA) using the Agilent High Sensitivity DNA kit (Cat. No. 507-4626, Agilent, Santa Clara, California, USA) or using a 2200 TapeStation (Agilent, Santa Clara, California, USA) with High Sensitivity DNA ScreenTapes (Cat. No. 5067-5593, Agilent, Santa Clara, California, USA). Equimolar indexed libraries were multiplexed (in groups of 6) to a final concentration of 4nM in 20 µL nuclease free dH_2O or 10 mM Tris-Cl, pH 8.5 (Buffer EB, Cat. No. 19086, Qiagen) and run on a single flow cell on an Ilumina NextSeq 500 according to manufacturer's instructions using a 2x75bp paired end read kit giving a total read length of 150 (TG NextSeq® 500 kit High Output Kit v2, Cat. No. TG-160-2002, Illumina). To overcome colour imbalance inherent to low complexity in a bisulfite-converted genome, 10% of phiX genome was spiked into the reaction. Q30 scores of bases from NextSeq runs were within the threshold recommended by the manufacturer and depth of coverage was approximately 40x.

DNA Data Mapping

FastQC v0.11.4 (28) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to visualise the sequencing quality of the raw reads which were then trimmed using Trim Galore! v0.4.0 (29) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). This removes low quality bases (Qscore<20) starting from the 3' end of the read. After trimming, short reads are removed (<20 bases). supplementary figure 1 shows a typical example. The Bismark package uses Bowtie 2 alignment software v2.2.6 (30, 31) to align sequences to the reference genome (GRCh38/hg19 assemblies) and then methylation data were extracted employing default settings (31). Alignment mapping efficiency was in the region of 80% across all samples and is illustrated by supplementary figure 2. The bisulfite error rate, estimated from the methylation status at cytosines outside a CpG context was in the region of 1.0% (supplementary figure 3). Duplicated reads (removed using Bismark) were in the region of 20%. At each cytosine site, the methylation level was calculated as the ratio of the count of "C" (or the number of sequencing reads with methylated cytosine) to the count of "C" plus "T" (or the total number of reads covering that site). M-bias plots were generated after methylation data was extracted with Bismark to yield the percentage methylation across all reads and to identify any bias (e.g. bias at the end of the read due to drop in quality) arising from the position in the read of the cytosine residue being called. supplementary figure 4 illustrates an example (for infant bloods) of an Mbias plot illustrating the reduction of call quality at the 5'and 3'ends of the paired reads. This provides a guide to the extent of necessary sequence trimming (typically 4 bases removed from the 5'and 1 from the 3' end). Methylation information was then re-extracted and the output was processed and converted to a bedgraph.

Methylation Data Analysis

The resulting bed files from Bismark were used for further statistical analysis. Three comparison groups based on different growth criteria were examined: (a) high versus low birthweight babies sampled at birth for placenta and cord blood, (b) high versus low birthweight groups sampled at 12m for infant blood and (c) high versus low length-for-age based on change in Z

score (LAZ) between birth and 12 months sampled at 12months for infant blood. Differential methylation between groups was examined using BiSeq (31, 32). Only CpGs covered by at least 10 reads were included in the analysis.

Detection of DMRs

Analysis was performed using R v 3.2.2 (33) and BiSeq version 1.18.0 (32, see review 34). BiSeq is designed specifically for targeted bisulfite sequencing data and includes features such as limiting high coverage, removing low coverage, spatial correlation, a multiple testing correction, visualization and genomic annotation. DMRs were detected by comparing birth weight categories (high v low) or length-for-age (LAZ) scores (high v low) and incorporated sex as a covariate. Briefly, sequences were grouped into clusters of adjacent CpG sites. CpG methylation often occurs in clusters and spatial correlation is a key characteristic of DNA methylation. As methylation is conserved across short distances, identification of these related regions reduces data dimensions and also increases detection power by borrowing nearby CpG information. BiSeq CpG clusters were defined as CpG sites covered in at least 25% of samples (defined as frequently covered CpG sites) with a maximum distance of 100bp between CpG sites within a cluster and with clusters containing at least 5 of these CpGs. To mitigate sequence overrepresentation distorting the data, sequences with greater than 90% of maximum coverage were removed. The methylation data was smoothed within CpG clusters using the smoothing algorithm ('predictMeth'). This estimates the true methylation level of each site in each sample. The methylation data were tested for both the test groups and resampled datasets under the null hypothesis that differences in methylation are random. The data from both were modelled by beta regression, with the group as the independent variable and the methylation probability as the dependent variable. A Wald test was used to confirm the parameters used in the beta regression could be included in the model and associated p-values were transformed into Z scores to allow DMRs to be detected (31). To account for multiple testing errors (multiple testing correction using the Benjamini-Hochberg method - 35), a two-step hierarchical procedure was employed. This first tests clusters, then individual CpG sites within those clusters. The two-step approach avoids loss of power by first testing at the cluster level and then the CpG in the cluster that showed a change in methylation and hence the number of CpGs needing correction is greatly reduced. A variogram was created under the null hypothesis, which estimates the correlation in methylation between two CpG sites within a cluster. This was plotted and smoothed, with a sill of 1 for all our tests and was combined with the Z scores of the test results of interest to estimate the correlation of Z scores between two locations in a cluster. Clusters without differentially methylated CpG sites were removed (FDR >= 0.1), before the remaining clusters were trimmed to indicate individual significant CpG sites (FDR <= 0.05). PCA analysis of methylation patterns determined from different DNA sequence runs did not reveal any batch effects.

Pyrosequencing

Validation of differentially methylated cytosines as detected by Methyl-Seq was performed by bisulfite pyrosequencing on the *ZFHX3* gene. Initially, PCR primers were designed using the Pyromark assay design SW 2.0 (Cat. No.9019077, Qiagen, USA) and were supplied by Sigma-Aldrich, UK. One of the primers was biotinylated and purified by HPLC. The primers were; *ZFHX3*: forward PCR primer GTTTTAATTTGATTGGGGGGAAAG, reverse PCR primer CCTTTAACAAACTAACCTCCTAACA, forward biotinylated sequencing primer TTTTTTTAAATGTAGATTTGAATT.

PCR amplification was performed with 10ng of bisulfite converted DNA using EpiTaq HS (Cat. No. R110A, TaKaRa Bio Inc, Japan). PCR was set up according to the manufacturers' instructions but the concentration of $MgCl_2$ varied between 15-25mM dependant on the primer set. Both methylated and unmethylated controls from the EpiTect PCR control DNA kit (Cat. No. 59695, Qiagen, USA) were run alongside. Thermal cycling conditions were performed using a touchdown programme with an annealing temperature range of 53°C - 62°C and cycle number range of 25-35, dependant on primer set. The PCR products were electrophoresed on a 3% agarose gel to check for product specificity. Pyrosequencing was then performed on the PyroMark Q24 Vacuum Workstation (Qiagen, USA) as described in the manufacturer's instructions. PyroMark CpG software Design 2.0 (Cat. No. 9019067, Qiagen, USA) was used in this assay and primers with the best quality score were selected. Bisulfite conversion was shown to be efficient for all samples as the fluorescence signal by cytosine in a non-CpG context was ≤ 1% of the signal produced by thymine.

Cellular Heterogeneity Assessment Between Sample Groups

For cord blood, cell composition was compared between low and high birthweight groups using overlaps with a cord blood cell type specific reference panel of 215,000 CpGs derived from the Illumina EPIC 850k array (ref: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6284779/). The reference panel set of CpG loci was used to find overlaps with the processed MethylSeq capture dataset. Co-methylation patterns extend up to several hundred base pairs across CpG clusters (36, 37). In order to obtain enough coverage for the regions covered by the EPIC reference set, we used intervals of 200 bases centred around the locations of the EPIC reference CpG set (updated in Human Genome - HG19). This yielded 14993 regions each containing CpG loci as present in the processed methyl capture sequence data set. Methylation calls were extracted from the processed sequence data as described above and the mean values in these regions were used to generate PCA plots and heatmaps to calculate the correlation values between experimental groups (using Pearson correlation). In the absence of a 12 month blood reference panel, the adult blood reference panel based on the Illumina Infinium HM450k and EPIC 850K methylation chips (38, 39) was used and processed in the same way for coverage across the Methylseq capture dataset. An interval of 200 bases yielded 33 regions each containing CpG loci (providing coverage for CD4 and CD8 lymphocytes, NK cells, neutrophils, B-cells and monocytes) that are present in the processed methyl capture sequence dataset to compare the 12 month groups.

CpG and Gene Annotation

Ensembl was used to annotate differentially methylated CpGs (based on hg38 version GRCh38 human genome build) to determine their location with respect to regulatory features. Ontologies, mutational and functional data of those genes associated with significant differentially methylated CpGs were determined by using the U.S. National Center for Biotechnology Information (NCBI; Bethesda, MD, USA; http://www.ncbi.nlm.nih.gov/) Gene, Online Mendelian Inheritance in Man (OMIM), PubMed databases and the Database for Annotation, Visualization and Integrated Discovery v6.7 (DAVID - http://david.abcc.ncifcrf.gov/) (40). Disease associations were determined by interrogating the Genetic Association Database (GAD) for complex diseases and the EBI GWAS Catalogue. PANTHER v14.0 (http://www.pantherdb.org) (41, 42) was used

to provide an overview of gene ontology (GO Terms) defining protein classes, cellular components, biological procesess and molecular functions of genes implicated by methylation marks.

Results

Quality Triage of Sample Cohorts

All samples underwent assessment to exclude maternal contamination and poor-quality samples. Maternal blood contamination of cord blood (for both sexes) was assessed using marker CpGs that are only methylated in adult blood DNA, and maternal contamination of placenta trophoblast samples from males was also flagged by examining the levels of Y DNA methylation dilution (43); see supplmentary figure 5a and b Poor quality and/or obvious outlier samples were identified by plotting a heatmap of the methylation data for each experimental group (see example of the methylation data from the birthweight cohort at 12 months of age in supplementary figure 6). The major component of variation was sex. Principal Component Analysis (PCA - done with and without inclusion of the sex chromosomes) matrices were also applied to a list of available variable information for the subjects contributing to each cohort and tissue sample (see supplementary table 1) to determine whether they had a significant effect on the variation in the data. Examples for sex (male/female), birthweight category (high/low) and season (dry/wet) are illustrated in supplementary figures 7a, b and c; only sex contributed significantly to variation in the data. Sample sets emerging from these analyses were re-analysed with sex as a covariate.

Assessment of Confounding Cellular Heterogeneity

There is no available cell type specific reference set for cord and adult blood to assess cell composition changes based on DMRs detected by Methyl-Seq data. Potentially confounding differences in cell-type composition between cord blood groupings (high and low birthweight) and infant blood groupings (high and low birthweight and length for age) were, therefore, assessed by using DMR regions within the capture DNA sequence data set that overlap within a 200 base-pair interval with the EPIC 850k cord blood and Infinium HM450k adult blood cell type specific CpG panels (see methods). The adult overlaps were used for the 12 month infant blood data in the absence of an age-related reference panel for this time point. The heatmap and PCA plots are shown in supplementary figures 8 and 9**.** For both cord and infant blood data, the heatmaps indicate high correlation between samples and no clear clustering according to comparison groups. PCA analysis indicates inter-individual differences in cellular composition. For the small number of probes from the adult blood reference panel present within the Methyl-Seq capture DNA sequence dataset, the analysis shows separation of individual samples into two groups; this may reflect the small number of probes available and their disproportionate weighting or variation in the rate of loss of nucleated erythrocytes between individuals. However, in all three PCA plots the variation between samples captured in PC1 and PC2 is distributed fairly uniformly across both experimental groups (high or low birthweight or tall or short height for age) indicating little difference in cellular composition between comparison groups to confound the determination of differential methylation values at the same time-point.

Differentially Methylated Loci Identified Through BiSeq Analysis

The total number of significant differentially methylated regions (DMRs) and the direction of median methylation change identified for each comparison group (placenta- birthweight, cord blood-birthweight, infant blood-birthweight at 12 months and infant blood-LAZ) by BiSeq analysis and the total number of significant CpGs they contain is shown in supplementary table 2a-d and summarized in table 3; 194 CpG loci in total. Each significantly differentially methylated CpG in each DMR was examined for the presence of single nucleotide polymorphism (SNP) directly in the CpG; these are shown in Supplementary table 2a-d. Apart from 3 SNP-containing CpGs, all the MAFs (minor allele frequencies) were <0.01. For the CpGs associated with implicated genes *RPS6KA2*, *PRSS3* and *GAR1*, the MAFs were <0.03, <0.11 and <0.02 respectively. These MAFs are at a level that would not significantly alter the estimation of methylation differences between test groups..

The distribution of CpGs between gene regulatory elements, gene bodies and intergenic regions is shown in table 4 (see supplementary tables 2a-d for full details on all DMRs and CpG locations). Figure 1 a-d provides an overview of the gene ontology (determined using PANTHER v14.0 available at http://www.pantherdb.org) characterizing genes implicated by differential methylation marks. The pie charts sumarise the distribution of this gene set across GO terms defining molecular functions, biological processes, cellular components and protein classes. It can be seen that certain GO categories predominate. For example, analysis of molecular function reveals that binding, catalytic activity, molecular function regulator and transcriptional regulator activity are most prominent. Detailed information on the genes and proteins in each of the GO categories can be obtained by uploading the gene lists to http://www.pantherdb.org from supplementary table 2a-b and interrogating each pie chart sector.

Very few of the differentially methylated CpGs found in DMRs identified from the cord blood comparisons are found in the 12 month infant blood comparisons; (a) of the four closely linked CpGs associated with *TNXB,* one (upstream intergenic) is differentially methylated in the cord blood birthweight group and the remaining three (within intron 1 of the gene) in the 12 month infant blood birthweight group and (b) the intergenic CpG upstream of the *HLX* gene is differentially methylated in both the 12-month birthweight and 12-month length for age groups.

CpG Loci Showing 5% or Greater Methylation Change

We have chosen to focus on those marks that show 5% or greater methylation change. Figure 2 summarises the CpGs that have been located to regulatory features (promoters, CTCF binding sites, transcription factor binding sites – 56 CpG loci in total) and table 6 those located to gene bodies (introns and exons) and closely linked intergenic regions (64 CpG loci in total). Figures 2 and 3 also present the locations of CpGs (based on hg38 release 85 from ENSEBL) with respect to the Transcription Start Site (TSS) of genes implicated by location (93 in total), the median p-value for the DMR corrected for multiple testing, direction and change in median methylation value, GWAS disease associations and a short vignette summarizing any mutational data and functional studies of implicated genes culled from the various databases outlined in the materials and methods. Finally, figures 2 and 3 flag whether any of the DMR associated genes are also subject to Trans or Cis-meQTLs (genetic variation that influences methylation at CpG sites adjacent to implicated genes) and/or Cis-eQTMs

(variation of methylation that influences expression of an adjacent gene) collated in the Bios QTLBrowser held at www.genenetwork.nl/biosqtlbrowser (85, these are marked in red in the first column). This data was based on the analysis of cohorts from the Dutch population and may only partially reflect genetic variation in the Gambian population with Trans or Cis-eQTL effects.

The implicated gene names in figures 2 and 3 are colour coded according to categories of gene function/disease revealed by functional and/or mutation analysis (see legends) and figure 4 summarises the numbers of implicated genes found in associated disease categories bearing the same colour coding. It is immediately clear that neurological, growth and development and oncological disorders are the most prominent amongst the implicated genes showing 5% or greater methylation change.

Replication of Findings in Other Methylation Studies

Identification of a significant proportion of the same implicated genes reported in related studies provides strong validation of the findings reported here. Highlighted in red bold in supplementary table 2a-d are the DMR-implicated genes that are also documented in the recent large meta-analysis of multiple EWAS by the Childhood Epigenetics Consortium examining DNA methylation associated with birthweight (15). When all genes (4848, representing about 19% of the estimated 25,000 genes in the human genome) from the consortium study associated with 8170 CpGs significant after FDR correction for multiple testing are screened, 62 DMR implicated genes from the current study show a match (of which 34 show >5% methylation change). If this is restricted to those genes (729; about 2.9% of human genes) associated with 914 CpG loci surviving Bonferroni correction ($p<1.06E-7$), then 11 matches are found (marked with a red asterisk of which 10 show >5% methylation change in the current study). The vast majority of CpGs in the meta-analysis are located within or closely linked to genes. Thus taking the number of genes in the genome as 25,000, the approximate probability of a match by chance for any given DMR-implicated gene in the present study is 0.19 (4848/25000) for all genes and 0.029 (729/25000) for those associated with the 914 CpG loci. The probability that these matches have occurred by chance for 62 and 11 genes is 0.19^{-62} and 0.029^{-11} respectively.

Comparisons have been made to two further studies examining the impact of gestational age (86, where some of the data is subsumed in the large meta-analysis mentioned above) and smoking on birthweight (87); these share, respectively, 53 (marked with a green asterisk in supplementary table 2a-d) and 11 (marked with a blue asterisk) genes associated with differential methylation identified in the current study. These two studies identify a further 22 DMR associated genes that overlap with our findings, bringing the replication in other related studies to 84 (48%) of the 173 implicated genes we have documented (The asterisks in tables 5 and 6 mark which of those shared genes show 5% or greater methylation change – in total 49 of the 93 in tables 5 and 6 between the three birthweight-related studies).

The genes *ZNF678*, *VTRNA2-1*, *SCRIB* and *TNXB* match those reported in other studies on maternal exposures and differential methylation of associated CpG loci in Gambian infants (22, 26 – marked with black triple asterisks in

supplementary table 2a-d) and *SEMA3B*, *ARID1B* and *HOXA10* from the Cambridge Baby Growth Study (88 – marked with black double asterisks).

The high degree of replication observed in related studies provides robust validation of the findings reported here. Pyrosequencing analysis of the methylation mark associated with the *ZFHX3* gene was performed to illustrate an example experimental confirmation of methylseq derived methylation data. Table 5 summarises the data for several individuals selected from the high and low groups for the 12 month LAZ comparison. The results in table 5 show good concordance between the two methods in both the quantum and direction of change when compared to the median methylation value change derived from the group comparisons by Biseq analysis of methylseq data.

Discussion

It has been suggested that epigenetic changes may be involved in the mechanism of reprogramming induced by under-nutrition, infection and adverse environmental exposures, although, it is not clear whether these are primary or secondary events in the chain of causality. This paper has used extremes of variation in birth weight and subsequent gains in length to examine associated methylation changes in DNA from trophoblast and cord blood DNA from small and large babies, and blood DNA from 12 month old infants analysed both according to their size at birth and their change in length from birth to 12 months (LAZ).

The methylation marks found at birth and those at 12 months in relation to birthweight show little longitudinal persistence (see tables 5 and 6 and supplementary table 2a-d). This suggests ongoing epigenetic adjustments, significant changes in blood cellular composition (such as the loss of nucleated erythrocytes – 89) or both in the critical postnatal growth period and the subsequent infancy-childhood growth transition (ICT) (90). Nonetheless, what does persist at 12 months is different, almost completely non-overlapping methylation patterns (not confounded by cellular composition differences) between the high and low birthweight comparison groups and the comparison groups showing rapid or slow post-natal height gain. These two distinct methylation patterns may reflect different interactions with nutritional, infectious and other environmental exposures during the post-natal growth phase potentially associated with negative or positive growth trajectories or a combination of both. Thus, any continued challenges (such as those provoked by under nutrition and infection) to homeostasis during the development period may trigger epigenetic programming and shift the timing and duration of these periods of growth. The study reported by Bernstein et al. (91) has revealed an accelerated transition to a childhood pattern of growth in Gambian compared to UK infants. A later transition, observed in UK infants, extends the high growth rate experienced during the infancy stage. This is reduced in Gambian infants, potentially impacting on growth outcomes in childhood while diverting energy into other processes critical for responses to acute infectious challenges; later developmental stages in this population offer an extended window for catch-up growth.

Over half (54.3%) of the identified methylation marks are located in gene regulatory elements, 30.3% in gene bodies and the remaining 15.4% in intergenic regions closely linked to implicated genes. Alteration of gene activity by methylation of

implicated genes may occur by impacting the functionality of cis transcriptional regulatory elements or changing chromatin conformation and accessibility to the transcriptional machinery. Several of the methylation marks are found in the binding site (an estimated 326,000 in the human genome) for the multifunctional CTCF zinc finger protein. This protein plays a key regulatory role through a number of varied functions that include influencing chromatin architecture (binding at chromatin domain boundaries and the formation of chromatin loops), binding to promoters, enhancers and within gene bodies and recruitment of transcription factors. The protein can also act as an insulator, blocking long-range promoter-enhancer interactions (for review see 92). Of particular relevance is the observation that methylation at CTCF binding sites in imprinted regions can disrupt the binding of the CTCF protein and its insulator activity (93) and, more generally, at many other methylated sites outside imprinted regions (94). From the annotation associated with each of the CpG loci covered by this methyl-seq capture set, almost all the methylation marks described in this study are in regions containing DNAse 1 hypersensitive Sites (DHS –markers of DNA regulatory regions and transcriptionally active open chromatin) described by the ENCODE (Encyclopedia of DNA Elements - 95) project. The ENCODE project has shown that a small proportion (~ 5%) of DHSs are found in TSS (Transcirption Start Site) regions, that most are located in introns and intergenic DNA and that there is cell type specificity in the distribution of DHSs. This indicates that the majority of methylation marks reported in our study are potentially in areas of remodelled open chromatin associated with transcriptional activity and may influence target gene activity possibly by altering chromatin architecture. Tables 5 and 6 also indicate that a number of the DMR-associated genes showing 5% or greater methylation change are subject to trans and/or cis genetic variation (Trans-meQTL and Cis-meQTL) that impacts the level of methylation of closely linked CpG loci; in some cases these methylation changes affect gene expression (Cis-eQTM). One consequence of this polymorphism in the genetic modulation of methylation marks is that it is likely to lead to a diversity of methylation responses to environmental exposures in different populations. Thus interaction between environmental exposures, genetic background and modulation of methylation patterns will have to be assessed for each study population.

Distribution of implicated genes across GO term categories demonstrates that they encompass biochemical and biological functions that include signalling or interaction with signalling pathways; interacting with or acting as receptors; constituents of or interacting with the extracellular matrix; deposition of connective tissue; structure and function of the actin cytoskeleton; trafficking across cellular membranes; cell cycle control and cellular growth; transcription regulation; metabolic regulation (see figure 1). Biological functions revealed by functional studies, animal models and mutation analysis primarily highlight roles in neurological, growth and developmental, neoplastic and immunological dysfunction (see figures 5, 6 and 7and supplementary table 2a-d for details). The precise impact of the methylation changes on the expression of implicated genes, however, is unknown and awaits more detailed functional analysis. Nevertheless, the location of these methylation marks within appropriately positioned regulatory elements and gene bodies or in close intergenic linkage to implicated genes, encourages their consideration as biomarkers associated with and the genetic pathways within which they are active in as potential contributors to variation in prenatal and post-natal growth , subsequent outcomes in later life and as possible intervention targets.

In total, eighty-four genes implicated by DMRs (shown in red bold and flagged by green and blue asterisks- see supplementary table 2a-d) are shared with DMR associated genes reported in the large array-based meta-analysis of multiple EWAS by the Childhood Epigenetics Consortium and two further related studies (15,86,87). This demonstrates concordance with a substantial proportion (48%) of the genes documented in the current study and provides robust validation of the Biseq analysis of methylseq data. Eleven matched genes are associated with CpG loci surviving stringent Bonferroni correction in the Kuppers et al. study (15) (marked with a red asterisk in supplementary table 2a-d). Differences in genetic background, environmental exposures and nutrition between populations contributing to different studies could lead to methylation changes at different CpG loci but still affect DMRs associated with the same implicated genes. In the case of *MAD1L1* and *NFIX*, differential methylation has been detected at the same Bonferroni significant CpG sites that are reported in the meta-analysis 15). *MAD1L1* (a component of the mitotic spindle-assembly checkpoint) has a role in cell cycle control and tumour suppression and methylation levels have been strongly correlated with hepatocellular carcinoma (96). It is also a susceptibility gene for bipolar disorder and schizophrenia with a risk allele linked to reward systems in healthy adults (97). *NFIX* is most highly expressed in brain, fat and prostate, is linked to cancer (DNA hypermethylation associated with lung adenocarcinoma - LUAD) (98), muscle development and dystrophies (99). Interestingly, 19p13 microduplications encompassing *NFIX* are responsible for intellectual disability, short stature and small head circumference (100).

Three implicated genes match those flagged by methylation changes found in DNA from babies in the Cambridge Baby Growth Study investigating the effects of maternal gestational diabetes or intrauterine growth retardation (88). *ARID1B* and *SEMA3B* are potential tumour suppressor genes. *ARID1B* is a chromatin remodelling factor and individuals with *ARID1B*-related disorder have many phenotypic features including slow growth (101). The third gene is *HOXA10* (homeobox A10), whose expression is down regulated in endometriosis but in late gestation is required for proper placental differentiation and function (100).

Implicated genes *ZNF678* (a zinc finger gene), *VTRNA2-1, SCRIB and TNXB* have been reported in other Gambian-based studies investigating periconceptional nutritional exposures associated with differential methylation (11, 22, 26). *VTRNA2a-1* is a non-coding RNA gene that functions as a tumour suppressor (47, 48, 49, 50) and is an imprinted locus (51, 52). *SCRIB* (a scaffold protein found at epithelial adherens junctions and neuronal presynaptic compartments) can act as a tumour suppressor gene and has been shown to be mutated in severe neural tube defects (see OMIM entry 607733). *TNXB* is an extracellular matrix glycoprotein thought to function in matrix maturation during wound healing. Different pathogenic alleles give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux –VUR (65), both of which involve alterations to collagen deposition in the extracellular matrix.

The genes *DLK1* and *MEG9* (LINC00584 - long intergenic non-coding RNA) are worthy of further comment given their location within an important imprinted region on chromosome 14 at 14q32. As revealed by maternal and paternal Uniparental Disomy (UPDm and UDPp, respectively) and genetic and functional studies of individual genes encompassed within the locus (for review see OMIM entries 601038, 60563, 611896, 172690, 613648 and ref 103), the region has a major impact on growth and development. The 14q23 locus is complex with a cluster of maternally and paternally imprinted genes, non-

coding snoRNAs (small nucleolar organiser RNA), miRNAs (microRNAs), LncRNAs (long non-coding RNAs) and LINC RNAs under the control of an intergenic differentially methylated region (IG-DMR) (104). Three genes (*DLK1*, *RTL1* and *DIO3*) are all expressed from paternal alleles. *DLK1*, containing six epidermal growth factor repeats, has reduced plasma levels in women bearing small for gestational age babies (105), is an inhibitor of adipocyte differentiation (106) and shows genetic association with age of menarche (107, 108, 107, 110). *RTL1* is essential for maintenance of foetal capillaries and potentially involved in formation of the chorioallantoic placenta (111), while *DIO3* (Thyoroxine Deiodinase Type III) is essential for the maturation and function of the thyoroid axis (112). A further four genes (*MEG3*, *RTL1as*, *MEG8* and *MEG9*) are all expressed from maternal alleles. *MEG3* is a LncRNA affecting growth and development in *Meg3* knock-out mice (113); *RTL1as* is an anti-sense transcript to the paternally expressed gene *RTL1* and encodes a number of microRNAs that may regulate the expression of *RTL1* (105); *MEG8* is a LncRNA involved in the regulation of trophoblast proliferation and invasion, and implicated in spontaneous early abortion (114) and *MEG9,* a LINC RNA involved in megakaryocyte differentiation and angiogenesis (70) shows genetic association with body mass index and age of menarche (107).

The UPDm (no paternal transcripts: Temple Syndrome) phenotype is characterised by pre-and postnatal growth retardation, neonatal hypotonia, precocious puberty and facial dysmorphism. The UPDp (no maternal transcripts: Kagami-Ogata Syndrome) phenotype is characterised by severe growth retardation, skeletal abnormalities, facial anomalies and abdominal muscular defects. Trans-regulation by maternally expressed small non-coding RNAs from the 14q32 region on the activity of other genes in the genome is likely to contribute to these complex phenotypes (105). On the maternal chromosome *DLK1* is silenced. The current study shows a 6% methylation difference of a *DLK1* DMR (higher in high birthweight than low birthweight babies). In contrast, *MEG9* is silenced on the paternal chromosome and shows a 22% methylation difference of a *MEG9* DMR (higher in high birthweight than in low birthweight babies). It is not clear what the impact of these methylation marks is on expression levels as they lie outside the immediate promoter within the gene body. Nevertheless, given that both methylation marks are in DMRs containing DHSs marking potentially open chromatin, it is reasonable to suggest that alteration of the methylation landscape in this region of chromosome 14 could impact chromatin architecture and gene activity with a bearing on growth and development outcomes. It is interesting to note that a study examining the effect of maternal periconceptional micronutrient supplementation of Gambian mothers found increased methylation of a *DLK1* associated CpG in cord blood DNA from offspring of mothers who had received the supplements (24).

A number of limitations should be noted. An accessible tissue such as blood as a proxy for methylation changes in other key target tissues will not capture all the relevant alterations in methylation status. However, there is sufficient concordance between tissues to yield a subset of potentially relevant loci (115, 116, 117). Analysis has been performed with males and females combined; hence sex differences in the methylation patterns have not been determined. The sample size is small, nevertheless, as outlined in the methods, BiSeq is designed for the analysis of targeted methyl sequence data and takes advantage of the conservation of methylation across short distances, co-assessing methylation changes at several individual cytosine residues within intervals of 100 base pairs. This reduces data dimensions and increases detection power by borrowing nearby CpG information and provides a more detailed and statistically significant evaluation of the methylation

status across any given genomic region; this has allowed identification of statistically valid differentially methylated CpGs from this small study. Greater coverage (3.7 million CpGs as compared to the Illumina HM450k and Epic 850k chips) of the SureSelect targeted sequencing approach of key gene regulatory elements (adjacent and distant, proximal or distal) to genes they control, offers the opportunity to identify additional methylation marks not necessarily scored by the array-based platforms.

Studies such as the one reported here provide associations and not cause and effect relationships between genes and phenotypes. Mutational evidence is helpful in establishing the likelihood that a gene contributes to a complex phenotype. Identification of methylation marks can be useful in that (a) they might act as biomarkers of early life adverse exposures that impact on early growth and may potentially indicate those individuals with higher future disease risks and (b) potentially flag genes that may be useful intervention targets to ameliorate the consequences of stunting. An integrated large scale analysis of inter-individual variation of methylation marks in relation to genotype (Trans and Cis-meQTLs), eQTLs (expression quantitative trait loci including Cis-eQTMs), disease susceptibility, developmental phenotypes, nutrition and environmental exposures provides a means of potentially unpicking causal relationships and the relevance of implicated genes. Clearly, the most effective approach to mitigate stunting and associated disease susceptibilities would be to ensure healthy nutrition, adequate sanitation and living conditions early in the life course.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

CRQ developed and managed experimental procedures and KMH performed them; MS and MG devised the initial analysis pipeline and JB and BS finalised the pipelines and provided bioinformatic support; CRQ performed the methylation data analysis; CAS and NAA provided supervisory support; AMD and SD managed the collection and extraction of DNA from samples; CRQ and NAA wrote the first draft of the manuscript; RMB, NA, DBD, KKO, AMP, and SEM conceived of and designed the HERO-G study. All authors contributed to manuscript revision, read and approved the submitted version.

Funding

This work was funded by the Bill and Melinda Gates Foundation (OPP1066932) and by core funding to the MRC Unit The Gambia at LSHTM (MC-A760-5QX00) by the UK MRC and the UK Department for the International Development (DFID) under the MRC/DFID Concordat agreement.

Acknowledgements

We thank the families of West Kiang who patiently participated in this study. We acknowledge the enthusiastic work of the whole HERO-G Working Group, especially the fieldworkers, village assistants, midwives, clinical staff, data office staff, and laboratory technicians who tirelessly collected the data and samples.

The raw data supporting the conclusions of this article are stored on the Open Science Framework (OSF), doi: 10.17605/OSF.IO/5ND3Y, and at the time of article submission are available on request and subject to review. These data will be made publicly available no later than July 1, 2021. Requests to access the datasets should be directed to the corresponding author.

References

1. Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z.A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R. and Uauy, R. Maternal and Child Nutrition Study Group. (2013) Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet*. **382,** 427-451. Erratum in (2013): *Lancet*. **382,** 396.
2. Moore, S. E. (2016) Early life nutritional programming of health and disease in The Gambia. *J Dev Orig Health Dis.* **7,** 123-31.
3. Olofin, I., McDonald, C. M., Ezzati M., Flaxman, S., Black, R. E., Fawzi, W. W., Caulfield, L. E, and Danaei, G. (2013) Associations of suboptimal growth with all-cause and cause-specific mortality in children under five years: a pooled analysis of ten prospective studies. *PloS one*, **8**, e64636.
4. Ong, K.K., Hardy, R., Shah, I., Kuh, D. (2013). Childhood stunting and mortality between 36 and 64 years: the British 1946 Birth Cohort Study. National Survey of Health and Development Scientific and Data Collection Teams. *J Clin Endocrinol Metab*. **98(5**), 2070-7.
5. UNICEF. WHO. World Bank Levels and trends in child malnutrition, UNICEF–WHO–World Bank joint child malnutrition estimates. (2015). www.who.int/entity/nutrition/publications/jointchildmalnutrition_2015_estimates/en/ (accessed February 16, 2020).
6. Christian, P., Lee, S. E., Angel, M. D., Adair, L. S., Arifeen, S. E., Ashorn, P., Barros, F. C., Fall, C. H. D., Fawzi, W. W., Hao, W., Hu, G., Humphrey, J. H., Huybregts, L., Joglekar, C. V., Kariuki, S. K., Kolsteren, P., Krishnaveni, G. V., Liu, E., Martorell, R., Osrin, D., Persson, L., Ramakrishnan, U., Richter, L., Roberfroid, D., Sania, A., Kuile, F. O. T., Tielsch, J., Victora, C. G., Yajnik, C. S., Yan, H., Zeng, L. and Black, R. E. (2013). Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *International Journal of Epidemiology*, **42**, 1340–1355.
7. Gluckman, P. D. (2011) Epigenetics and metabolism in 2011: Epigenetics, the life-course and metabolic disease. *Nat Rev Endocrinol*. **8,** 74–76.
8. Barker, D. J. P., (2004). Developmental origins of adult health and disease. *Journal of Epidemiology & Community Health*, 58, 114–115.

9. Gluckman, P. D., Hanson, M. A., Spencer, H. G., and Bateson, P. (2005) Environmental influences during development and their later consequences for health and disease: implications for the interpretation of empirical studies. *Proc. Biol. Sci.* **272**, 671–677.
10. Stein, A. D. et al., (2006). Exposure to a nutrition supplementation intervention in early childhood and risk factors for cardiovascular disease in adulthood: Evidence from Guatemala. *American Journal of Epidemiology*, **164,** 1160–1170.
11. Adu-Afarwuah, S., Young, R. T., Lartey, A., Okronipa, H., Ashorn, P., Ashorn, U., Zeilani, M. and Dewey, K. G. (2007). Randomized comparison of 3 types of micronutrient supplements for home fortification of complementary foods in Ghana: Effects on growth and motor development. *American Journal of Clinical Nutrition*, **86**, 412–420.
12. Bhutta, Z. A., Ahmed, T., Black, R. E. Cousens, S., Dewey, K., Giugliani, E., Haider, B. A., Kirkwood, B., Saul S. Morris, S. S., Sachdev, H.P.S. and Shekar, M. (2008) What works? Interventions for maternal and child undernutrition and survival. *Lancet*. **371**, 417–440.
13. Fleming, T.P., Watkins, A.J., Velazquez, M.A., Mathers, J.C., Prentice, A.M., Stephenson, J., Barker, M., Saffery, R., Yajnik, C.S., Eckert, J.J., Hanson, M.A., Forrester, T., Gluckman, P.D,, Godfrey, K.M. (2018) Origins of lifetime health around the time of conception: causes and consequences. *Lancet.* **391(10132),**1842-1852.
14. Lillycrop, K. A., and Burdge, G. C. (2012) Epigenetic mechanisms linking early nutrition to long term health. *Best Pract. Res. Clin. Endocrinol. Metab*. **26**, 667–676.
15. Küpers, L.K., Monnereau, C., Sharp, G. C., Yousefi, P., Salas, L. A., Ghantous, A., Page, C. M., Reese, S. E., Wilcox, A. J., Czamara, D., Starling, A. P., Novoloaca, A., Lent, S., Roy, R., Hoyo, C., Breton, C. V., Allard, C., Just, A. C., Bakulski, K. M., Holloway, J. W., Everson, T. M., Xu, C. J., Huang, R. C., van der Plaat, D. A., Wielscher, M., Merid, S.K., Ullemar, V., Rezwan, F. I., Lahti, J., van Dongen, J., Langie, S. A. S, Richardson, T. G., Magnus, M. C., Nohr, E. A., Xu, Z., Duijts, L., Zhao, S., Zhang, W., Plusquin, M., DeMeo, D. L., Solomon, O., Heimovaara, J.H., Jima, D. D., Gao, L., Bustamante, M., Perron, P., Wright, R. O., Hertz-Picciotto, I., Zhang, H., Karagas, M. R., Gehring U, Marsit CJ, Beilin LJ, Vonk JM, Jarvelin MR, Bergström A, Örtqvist AK, Ewart S, Villa PM, Moore SE, Willemsen G, Standaert ARL, Håberg SE, Sørensen TIA, Taylor JA, Räikkönen K, Yang IV, Kechris K, Nawrot TS, Silver MJ, Gong YY, Richiardi L, Kogevinas M, Litonjua AA, Eskenazi B, Huen K, Mbarek H, Maguire RL, Dwyer T, Vrijheid M, Bouchard L, Baccarelli AA, Croen LA, Karmaus W, Anderson D, de Vries M, Sebert S, Kere J, Karlsson R, Arshad SH, Hämäläinen E, Routledge MN, Boomsma DI, Feinberg AP, Newschaffer CJ, Govarts E, Moisse M, Fallin MD, Melén E, Prentice AM, Kajantie E, Almqvist C, Oken E, Dabelea D, Boezen HM, Melton PE, Wright R.J., Koppelman G.H., Trevisi L., Hivert M.F., Sunyer J., Munthe-Kaas M.C., Murphy S.K., Corpeleijn E., Wiemels J., Holland N., Herceg Z., Binder E.B., Davey Smith G., Jaddoe, VWV, Lie, RT, Nystad, W, London, SJ, Lawlor, DA, Relton, CL, Snieder, H, Felix JF. (2019) Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat Commun*. **10,** 1893.
16. Schoenbuchner SM, Dolan C, Mwangome M, Hall A, Richard SA, Wells JC, Khara T, Sonko B, Prentice AM, Moore SE (2019). The relationship between wasting and stunting: a retrospective cohort analysis of longitudinal data in Gambian children from 1976 to 2016. Am J Clin Nutr. 2019 Aug 1;110(2):498-507.
17. Nabwera, H. M., Bernstein, R. M., Agbla, S. C., Moore, S. E., Darboe, M. K., Mariama Colley, M., Jallow, A. T.,

Bradbury, R., Karafin, J., Fulford, A. J., and Prentice, A. M. (2018) Hormonal Correlates and Predictors of Nutritional Recovery in Malnourished African Children. *J Trop Pediatr*. **64,** 364–372.

18. Rayco-Solon P, Fulford AJ, Prentice AM. (2005) Differential effects of seasonality on preterm birth and intrauterine growth restriction in rural Africans. *Am J Clin Nutr*. **81(1):**134-9.
19. Waterland, R.A., Kellermayer, R., Laritsky, E., Rayco-Solon, P., Harris, R.A., Travisano, M., Zhang, W., Torskaya, M.S., Zhang, J., Shen, L., Manary, M.J., Prentice, A.M. (2010). Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* **6**:e1001252.
20. Dominguez-Salas, P., Moore, S.E., Cole, D., da Costa, K.A., Cox, S.E., Dyer, R.A., Fulford, A.J., Innis, S.M., Waterland, R.A., Zeisel, S.H., Prentice, A.M., Hennig, B.J. (2013). DNA methylation potential: dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. Am J Clin Nutr. 97,1217-1227.
21. Dominguez-Salas, P., Moore, S. E., Baker, M. S., et al. (2014) Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat Commun*. **5**:3746.
22. Silver M.J., Kessler N.J., Hennig B.J., Dominguez-Salas P., Laritsky E., Baker M.S., Coarfa C., Hernandez-Vargas H., Castelino J.M., Routledge M.N., Gong Y.Y., Herceg Z., Lee Y.S., Lee K., Moore S.E., Fulford A.J., Prentice A.M., Waterland R.A. (2015). Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biol.*;16 (1):118.
23. Kessler N.J., Waterland R.A., Prentice A.M., Silver M.J. (2018**)** Establishment of environmentally sensitive DNA methylation states in the very early human embryo. Sci Adv. 4(7):eaat2624.
24. Khulan, B., Cooper, W. N., Skinner, B. M., Bauer, J., Owens, S., Prentice, A. M., Belteki, G., Constancia, M., Dunger, D. and Affara, N. A. (2012). Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: A study of a unique resource in the Gambia. *Human Molecular Genetics* **21**, 2086–2101.
25. Cooper, W. N., Khulan, B., Owens, S., Elks, C.E., Seidel, V., Prentice, A. M., Belteki, G., Ong, K. K., Affara, N.A., Constância, M. and Dunger, D. B. (2012) DNA methylation profiling at imprinted loci after periconceptional micronutrient supplementation in humans: results of a pilot randomized controlled trial. *FASEB J*. **26,** 1782–1790.
26. Hernandez-Vargas H., Castelino J., Silver M.J., Dominguez-Salas P., Cros M.P., Durand G., Le Calvez-Kelm F., Prentice A.M., Wild C.P., Moore S.E., Hennig B.J., Herceg Z., Gong Y.Y., Routledge M.N. (2015). Exposure to aflatoxin B1 in utero is associated with DNA methylation in white blood cells of infants in The Gambia. *Int J Epidemiol.* 44(4):1238-48.
27. Moore, S.E., Doel, A.M, Ong, K.K., Dunger, D.B., Affara, N.A,, Prentice, A.M., Bernstein, R.M. HERO-G Working Group. (2020). Identification of nutritionally modifiable hormonal and epigenetic drivers of positive and negative growth deviance in rural African fetuses and infants: Project protocol and cohort description [version 1; peer review: awaiting peer review]. *Gates Open Res* 2020, **4**:25.
28. Andrews, S. (2010). FastQC A Quality Control Tool for High Throughput Sequence Data. Available at: http://www.bioinformatics,babraham.ac.uk/projects/fastqc/

29. Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* **9,** 145-51.
30. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2 *Nat Methods*. **9,** 357–359.
31. Krueger, F., and Andrews, S. R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. **27**, 1571–1572.
32. Klein, H. U., and Hebestreit, K. (2016) An evaluation of methods to test predefined genomic regions for differential methylation in bisulfite sequencing data. *Brief Bioinform*. **17**, 796–807.
33. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
34. Yu, X., and Sun, S. (2016) Comparing five statistical methods of differential methylation identification using bisulfite sequencing data. *Stat Appl Genet Mol Biol.* **15,** 173-191.
35. Benjamini, Y. and Heller, R. (2007) False discovery rates for spatial signals. *J. Am. Stat. Assoc.* **102**, 1271-1281.
36. Lövkvist, C., Dodd, I..B, Sneppen, K., Haerter, J.O. (2016). DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res.* **44**, 5123-32.
37. Affinito, O., Palumbo, D., Fierro, A., Cuomo, M., De Riso, G., Monticelli, A., Miele, G., Chiariotti, L., Cocozza, S.(2020). Nucleotide distance influences co-methylation between nearby CpG sites *Genomics* **112**, 144-150.
38. Salas, L.A., Koestler, D.C., Butler, R.A., Hansen M.H., Wiencke J.K., Kelsey K.T. and Christensen B.C. (2018). An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* **19,** 64. doi: 10.1186/s13059-018-1448-7.
39. Reinius L.E., Acevedo N., Joerink M., Pershagen G., Dahlén S-E., Greco D., Söderhäll C., Scheynius A., Kere J. 2012). Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLOS ONE 7:e41361.
40. Huang da W., Sherman, B.T., and Lempicki, R. A., (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* **4,** 44–57.
41. Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with PANTHER Classification System *Nat Protoc* **8,** 1551–1566.
42. Mi H., Muruganujan A., Ebert D., Huang X. and Thomas, (2019) P.D PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucl. Acids Res.* doi: 10.1093/nar/gky1038
43. Morin, A. M., Gatev, E., McEwen, L. M., MacIsaac, J. L., Lin, D.T.S., Koen, N., Czamara, D., Räikkönen, K., Zar, H. J., Koenen, K., Stein, D. J., Kobor, M.S., and Jones, M. J. (2017) Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. *Clin Epigenetics* **9**, 75.
44. Helbig, I., Lopez-Hernandez, T., Shor, O., Galer, P., Ganesan, S., Pendziwiat, M., Rademacher, A., Ellis, C. A., Hümpfer, N., Schwarz, N., Seiffert, S., Peeden, J., Shen, J., Štěrbová, K., Hammer, T. B., Møller, R. S., Shinde, D.N., Tang, S., Smith, L., Poduri, A., Krause, R., Benninger, F., Helbig, K. L., Haucke, V., and Weber Y. G.; EuroEPINOMICS-RES Consortium; GRIN Consortium. (2019) A Recurrent Missense Variant in AP2M1 Impairs Clathrin-Mediated Endocytosis and Causes Developmental and Epileptic Encephalopathy. *Am J Hum Genet*. **104,** 1060-1072.

45. Tarpey, P., Parnau, J., Blow, M., Woffendin, H., Bignell, G., Cox, C., Cox, J., Davies, H., Edkins, S., Holden, S., Korny, A., Mallya, U., Moon, J., O'Meara, S., Parker, A., Stephens, P., Stevens, C., Teague, J., Donnelly, A., Mangelsdorf, M., Mulley, J., Partington, M., Turner, G., Stevenson, R., Schwartz, C., Young, I., Easton, D., Bobrow, M., Futreal, P.A., Stratton, M.R., Gecz, J., Wooster, R., and Raymond, F.L. (2004) Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. *Am J Hum Genet.* **75,** 318-324.
46. Philips, A. K., Sirén, A., Avela, K., Somer, M., Peippo., M., Ahvenainen, M., Doagu, F., Arvio, M., Kääriäinen, H., Van Esch, H., Froyen, G., Haas, S.A., Hu, H., Kalscheuer, V. M., and Järvelä, I. (2014) X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. *Orphanet J Rare Dis*. **9,** 49.
47. L.ee K., Kunkeaw N., Jeon S.H., Lee I., Johnson B.H., Kang G-Y., Bang J.Y., Park H.S., Leelayuwat J., Lee Y.S (2011). Precursor miR-886, a Novel Noncoding RNA Repressed in Cancer, Associates With PKR and Modulates Its Activity. *RNA* 17, 1076-1089. 14.
48. Treppendahl, M.B., Qiu, X., Sogaard, A., Yang, X., Nandrup-Bus, C., Hother, C., Andersen, M.K., Kjeldsen, L., Möllgård, L., Hellström-Lindberg, E., Jendholm, J., Porse, B.T., Jones, P.A., Liang, G., Grønbæk, K. (2012). Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. *Blood* **119**, 206–16.
49. 15. Cao, J., Song Y., Bi, N., Shen, J., Liu, W., Fan, J., Sun, G., Tong, T., He, J., Shi, Y., Zhang, X., Lu, N., He, Y., Zhang, H., Ma, K., Luo, X., Lv, L., Deng, H., Cheng, J., Zhu, J., Wang, L., Zhan, Q. (2013). DNA methylation-mediated repression of miR-886-3p predicts poor outcome of human small cell lung cancer. *Cancer Res*. **73**, 3326–35.
50. 16. Lee, H.S., Lee, K., Jang, H.J., Lee, G.K., Park, J.L., Kim, S.Y., Kim, S.B., Johnson, B.H., Zo, J.I., Lee, J.S., Lee, Y.S. (2014). Epigenetic silencing of the non-coding RNA nc886 provokes oncogenes during human esophageal tumorigenesis. *Oncotarget.* **5**, 3472–81.
51. 17. Paliwal, A., Temkin, A.M., Kerke,l K., Yale, A., Yotova, I., Drost, N., Lax, S., Nhan-Chang, C.L., Powell, C., Borczuk, A., Aviv, A., Wapner, R., Chen, X., Nagy, P.L., Schork, N., Do, C., Torkamani, A., Tycko, B. (2013). Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet. 9 doi: 10.1371/journal.pgen.1003622.
52. Romanelli, V., Nakabayashi, K., Vizoso, M., Moran, S., Iglesias-Platas, I., Sugahara, N., Simón C., Hata K., Esteller M., Court F., Monk D. (2014). Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics* 9, 783–90.
53. Ptácek, R., Kuzelová, H., and Stefano, G. B. (2011) Dopamine D4 receptor gene DRD4 and its association with psychiatric disorders. *Med Sci Monit*. **17.** RA215-20.
54. Beecham G.W., Dickson D.W., Scott W.K., Martin E.R., Schellenberg G., Nuytemans K., Larson E.B., Buxbaum J.D., Trojanowski J.Q., Van Deerlin V.M., Hurtig H.I., Mash D.C., Beach T.G., Troncoso J.C., Pletnikova O., Frosch M.P., Ghetti B., Foroud T.M., Honig L.S., Marder K., Vonsattel J.P., Goldman S.M., Vinters H.V., Ross O.A., Wszolek Z.K., Wang L., Dykxhoorn D.M., Pericak-Vance M.A., Montine T..J, Leverenz J.B., Dawson T.M., Vance

J.M. (2015). PARK10 is a major locus for sporadic neuropathologically confirmed Parkinson disease. *Neurology*. 84, 972-80

55. Osada, Y., Hashimoto, T., Nishimura, A., Matsuo, Y., Wakabayashi, T., and Iwatsubo, T. (2005) CLAC binds to amyloid beta peptides through the positively charged amino acid cluster within collagenous domain 1 and inhibits formation of amyloid fibrils. *J. Biol. Chem.* **280**, 8596-8605. Note: Erratum: *J. Biol. Chem*. **280**, 15484.
56. Cruz-Garcia, D., Vazquez-Martinez, R., Peinado, J. R., Anouar, Y., Tonon, M. C., Vaudry, H., Castano, J. P. and Malagon, M. M. (2007) Identification and characterization of two novel (neuro)endocrine long coiled-coil proteins. *FEBS Lett*. **581,** 3149-3156.
57. Kowalski, E. J. A., and Li, L. (2017) Toll-Interacting Protein in Resolving and Non-Resolving Inflammation. *Front Immunol.* **8**, 511.
58. MacDonald, J. I., Kubu, C. J., and Meakin, S. O. (2004) Nesca, a novel adapter, translocates to the nuclear envelope and regulates neurotrophin-induced neurite outgrowth. *J Cell Biol.* **164,** 851-862. Ptácek, R., Kuzelová, H., and Stefano, G. B. (2011) Dopamine D4 receptor gene DRD4 and its association with psychiatric disorders. *Med Sci Monit*. **17.** RA215-20.
59. Anazi, S., Maddirevula, S., Salpietro, V., Asi, Y. T., Alsahli, S., Alhashem, A., Shamseldin, H. E., AlZahrani, F., Patel, N., Ibrahim, N., Abdulwahab, F. M., Hashem, M., and 31 others. (2017) Expanding the genetic heterogeneity of intellectual disability. *Hum. Genet*. **136**: 1419-1429,. Note: Erratum (2018): *Hum. Genet*. **137**, 105-109.
60. Smigiel, R., Sherman, D. L., Rydzanicz, M., Walczak, A., Mikolajkow, D., Krolak-Olejnik, B., Kosinska, J., Gasperowicz, P., Biernacka, A., Stawinski, P., Marciniak, M., Andrzejewski, W., Boczar, M., Krajewski, P., Sasiadek, M. M., Brophy, P. J., and Ploski, R. (2018) Homozygous mutation in the neurofascin gene affecting the glial form of neurofascin causes severe neurodevelopment disorder with hypotonia, amimia and areflexia. *Hum. Molec. Genet*. **27**, 3669-3674.
61. Monfrini, E., Straniero, L., Bonato, S., Monzio Compagnoni, G., Bordoni, A., Dilena, R., Rinchetti, P., Silipigni, R., Ronchi, D., Corti, S., Comi, G. P., Bresolin, N., Duga, S., and Di Fonzo, A. (2019) Neurofascin (NFASC) gene mutation causes autosomal recessive ataxia with demyelinating neuropathy. *Parkinsonism Relat. Disord*. **63**, 66-72.
62. Pillai, A. M., Thaxton, C., Pribisko, A. L., Cheng, Jr-G., Dupree, J. L., and Bhat, M. A. (2009) Spatiotemporal ablation of myelinating glia-specific neurofascin (Nfasc-NF155) in mice reveals gradual loss of paranodal axoglial junctions and concomitant disorganization of axonal domains. *J. Neurosci. Res.* **87**, 1773-1793.
63. Krupp, M., Weinmann, A., Galle, P.R., Teufel, A. (2006) Actin binding LIM protein 3 (abLIM3). *Int J Mol Med.* **17(1)**, 129-33.
64. Burch, G. H., Gong, Y., Liu, W., Dettman, R. W., Curry, C.J., Smith, L., Miller, W. L. and Bristow, J. (1997) Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet*. **17,** 104-108.
65. Gbadegesin, R. A., Brophy, P. D., Adeyemo, A., Hall, G., Gupta, I. R., Hains, D., Bartkowiak, B., Rabinovich, C. E., Chandrasekharappa, S., Homstad, A., Westreich, K., Wu, G., Liu, Y., Holanda, D., Clarke, J., Lavin, P., Selim, A., Miller, S., Wiener, J. S., Ross, S. S., Foreman ,J., Rotimi, C., and Winn, M. P. (2013) TNXB mutations can cause vesicoureteral reflux. *J Am Soc Nephrol.* **24,** 1313-22.

66. Saba R., Kato Y. and Saga Y. (2014). Nanos2 promotes male germ cell development independent of meiosis suppression. *Developmental Biology* **385**, 32-40.
67. Karaca, E., Harel, T., Pehlivan, D., Jhangiani, S. N., Gambin, T., Akdemir, Z. C., Gonzaga-Jauregui, C., Erdin, S., Bayram, Y., Campbell, I. M., Hunter, J. V., Atik, M. M., and 52 others. (2015) Genes that affect brain structure and function identified by rare variant analyses of mendelian neurologic disease. *Neuron* **88**, 499-513.
68. Redler, S., Strom, T. M., Wieland, T., Cremer, K., Engels, H., Distelmaier, F., Schaper, J., Kuchler, A., Lemke, J. R., Jeschke, S., Schreyer, N., Sticht, H., Koch, M., Ludecke, H.-J., and Wieczorek, D. (2017) Variants in CPLX1 in two families with autosomal-recessive severe infantile myoclonic epilepsy and ID. *Europ. J. Hum. Genet*. **25,** 889-893.
69. Ye Y.P., Jiao H.L., Wang S.Y., Xiao Z.Y., Zhang D., Qiu J.F., Zhang L.J., Zhao Y.L., Li T.T., Li-Liang, Liao W.T., Ding Y.Q. (2018). Hypermethylation of DMTN promotes the metastasis of colorectal cancer cells by regulating the actin cytoskeleton through Rac1 signaling activation. *J Exp Clin Cancer Res.* 37, 299.
70. Espinosa Diez, C., Wilson, R., Mukherjee, R., Feltham, M., Hudson, C., Ruhl, R. and Anand, S. (2018) DNA damage dependent hypomethylation regulates the pro-angiogenic LncRNA MEG9. BioRxiv, the preprint server for biology.
71. Roberts, J. D., Thapaliya, A., Martínez-Lumbreras, S., Krysztofinska, E. M., and Isaacson, R. L. (2015) Structural and Functional Insights into Small, Glutamine-Rich, Tetratricopeptide Repeat Protein Alpha. *Front Mol Biosci.* **2,** 71.
72. Paul, A., Garcia, Y. A., Zierer, B., Patwardhan, C., Gutierrez, O., Hildenbrand, Z., Harris, D. C., Balsiger, H. A., Sivils, J.C., Johnson, J. L., Buchner, J., Chadli, A., Cox, M. B. (2014) The cochaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing protein alpha) demonstrates regulatory specificity for the androgen, glucocorticoid, and progesterone receptors. *J Biol Chem*. **289,** 15297-15308.
73. Leznicki, P., and High, S. (2012) SGTA antagonizes BAG6-mediated protein triage.*Proc Natl Acad Sci U S A.* **109,** 19214-19219.
74. Deguchi, Y., Agus, D. and Kehrl, J. H. (1993) A human homeobox gene, HB24, inhibits development of CD4+ T cells and impairs thymic involution in transgenic mice. *J Biol Chem*. **268,** 3646-3653.
75. Hentsch, B., Lyons, I., Li, R., Hartley, L., Lints, T. J., Adams, J. M., and Harvey, R. P. (1996) Hlx homeo box gene is essential for an inductive tissue interaction that drives expansion of embryonic liver and gut. *Genes Dev.* **10,** 70-79.
76. Rajaraman, G., Murthi, P., Quinn, L., Brennecke, S.P., Kalionis, B. (2008) Homeodomain protein HLX is expressed primarily in cytotrophoblast cell types in the early pregnancy human placenta *Reprod Fertil Dev.* **20(3),** 357-67.
77. Rajaraman, G., Murthi, P., Leo, B., Brennecke, SP., Kalionis, B. (2007) Homeobox gene HLX1 is a regulator of colony stimulating factor-1 dependent trophoblast cell proliferation. *Placenta* **28(10),** 991-8.
78. Murthi, P., Doherty, V., Said, J., Donath, S., Brennecke, S.P., Kalionis, B. (2006) Homeobox gene HLX1 expression is decreased in idiopathic human fetal growth restriction..*Am J Pathol.* **168(2),** 511-8.
79. Sakata, N., Kaneko, S., Ikeno, S., Miura, Y., Nakabayashi, H., Dong, X. Y., Dong, J. T., Tamaoki, T., Nakano, N., and Itoh, S. (2014) TGF- β Signaling Cooperates with AT Motif-Binding Factor-1 for Repression of the α -Fetoprotein Promoter. *J SignalTransduct.* **2014**, 970346.

80. Mori, Y., Kataoka, H., Miura, Y., Kawaguchi, M., Kubota, E., Ogasawara, N., Oshima, T., Tanida, S., Sasaki, M., Ohara, H., Mizoshita, T., Tatematsu, M., Asai, K., Joh, T. (2007) Subcellular localization of ATBF1 regulates MUC5AC transcription in gastric cancer. *Int J Cancer.* **121,** 241-247.
81. Berry, F. B., Miura, Y., Mihara, K., Kaspar, P., Sakata, N., Hashimoto-Tamaoki, T. and Tamaoki, T. (2001) Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. *J Biol Chem*. **276**, 25057-25065.
82. Dong, X. Y., Sun, X., Guo, P., Li, Q., Sasahara, M., Ishii, Y. and Dong, J. T. (2010) ATBF1 inhibits estrogen receptor (ER) function by selectively competing with AIB1 for binding to the ER in ER-positive breast cancer cells. *J Biol Chem.* **285,** 32801-32809.
83. Sun, X., Frierson, H. F., Chen, C., Li, C., Ran, Q., Otto, K. B., Cantarel, B. L., Vessella, R. L., Gao, A. C., Petros, J., Miura, Y., Simons, J. W., Dong, J. T. (2005) Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet.* **37,** 407-412. Erratum in: *Nat Genet.* **37,** 652. Cantarel, Brandi M [corrected to Cantarel, Brandi L].
84. Yu C-L., Xu X-L. and Yuan F. (2019). LINC00511 is associated with the malignant status and promotes cell proliferation and motility in cervical cancer. *Bioscience Reports* **39**, BSR20190903.
85. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, Slieker RC, Jhamai PM, Verbiest M, Suchiman HE, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Arindrarto W, Kielbasa SM, Jonkers I, van 't Hof P, Nooren I, Beekman M, Deelen J, van Heemst D, Zhernakova A, Tigchelaar EF, Swertz MA, Hofman A, Uitterlinden AG, Pool R, van Dongen J, Hottenga JJ, Stehouwer CD, van der Kallen CJ, Schalkwijk CG, van den Berg LH, van Zwet EW, Mei H, Li Y, Lemire M, Hudson TJ; BIOS Consortium, Slagboom PE, Wijmenga C, Veldink JH, van Greevenbroek MM, van Duijn CM, Boomsma DI, Isaacs A, Jansen R, van Meurs JB, 't Hoen PA, Franke L, Heijmans BT. (2017). Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* **49,** 131-138.
86. Merid SK, Novoloaca A, Sharp GC, Küpers LK, Kho AT, Roy R, Gao L, Annesi-Maesano I, Jain P, Plusquin M, Kogevinas M, Allard C, Vehmeijer FO, Kazmi N, Salas LA, Rezwan FI, Zhang H, Sebert S, Czamara D, Rifas-Shiman SL, Melton PE, Lawlor DA, Pershagen G, Breton CV, Huen K, Baiz N, Gagliardi L, Nawrot TS, Corpeleijn E, Perron P, Duijts L, Nohr EA, Bustamante M, Ewart SL, Karmaus W, Zhao S, Page CM, Herceg Z, Jarvelin MR, Lahti J, Baccarelli AA, Anderson D, Kachroo P, Relton CL, Bergström A, Eskenazi B, Soomro MH, Vineis P, Snieder H, Bouchard L, Jaddoe VW, Sørensen TIA, Vrijheid M, Arshad SH, Holloway JW, Håberg SE, Magnus P, Dwyer T, Binder EB, DeMeo DL, Vonk JM, Newnham J, Tantisira KG, Kull I, Wiemels JL, Heude B, Sunyer J, Nystad W, Munthe-Kaas MC, Räikkönen K, Oken E, Huang RC, Weiss ST, Antó JM, Bousquet J, Kumar A, Söderhäll C, Almqvist C, Cardenas A, Gruzieva O, Xu CJ, Reese SE, Kere J, Brodin P, Solomon O, Wielscher M, Holland N, Ghantous A, Hivert MF, Felix JF, Koppelman GH, London SJ, Melén E. (2020). Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age. *Genome Med.* **12**, 25.

87. Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Hougaard DM, Bresnahan M, Mors O, Hollegaard MV, Bækvad-Hansen M, Hornig M, Mortensen PB, Børglum AD, Werge T, Pedersen MG, Nordentoft M; iPSYCH-Broad ASD Group, Buxbaum JD, Daniele Fallin M, Bybjerg-Grauholm J, Reichenberg A, Mill J. (2019). Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Phios Trans R Soc Lond B Bilo Sci* **374**, (1770) 20180120374(1770):20180120
88. Quilter CR, Cooper WN, Cliffe KM, Skinner BM, Prentice PM, Nelson L, Bauer J, Ong KK, Constância M, Lowe WL, Affara NA, Dunger DB. (2014) Impact on offspring methylation patterns of maternal gestational diabetes mellitus and intrauterine growth restraint suggest common genes and pathways linked to subsequent type 2 diabetes risk. *FASEB J.* **28(11),** 4868-79.
89. Bakulski, K.M., Feinberg, J.I., Andrews, S.V., Yang, J., Brown, S.L., McKenney, S., Witter, F., Walston, J,, Feinberg, A.P., Fallin, M.D. (2016). DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* **11**, 354–362.
90. Karlberg, J., Fryer, J. G., Engström, I., and Karlberg, P. (1987) Analysis of linear growth using a mathematical model. II. From 3 to 21 years of age. *Acta Paediatr Scand Suppl*. **337**, 12–29.
91. Bernstein, R.M. O'Connor, G.K., Vance, E.A., Affara, N., Drammeh, S., Dunger, D.B., Faal, A., Ong, K.K., Prentice, A.M., Sosseh, F., Moore, S.E. (2020). Timing of the infancy-childhood transition in rural Gambia. *Frontiers in Endocrinology.* doi: 10.3389/fendo.2020.00142.
92. Holwerda S.J.B. and de Laat, W. (2013). CTCF:the protein, the binding partners, the binding sites and their chromatin loops. *Philos Tran R Soc Lond B Biol Sci* **368**, 20120369.
93. Bell, A.C. and Felsenfeld G. (2000) Methylationof a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482-485.
94. Wang, H., Maurano M.T., Qu H., Varley K.E., Gertz J., Pauli F., Lee K., Canfield T., Weaver M., Sandstrom R., Thurman R.E., Kaul K., Myers R.M., Stamatoyannopoulos J.A. (2012). Widespread Plasticity in CTCF Occupancy Linked to DNA Methylation *Genome Res* **22**, 1680-1688.
95. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
96. Cui, C., Lu, Z., Yang, L., Gao, Y., Liu, W., Gu, L., Yang, C., Wilson, J., Zhang, Z., Xing, B., Deng, D. and Sun, Z.S. (2016) Genome-wide identification of differential methylation between primary and recurrent hepatocellular carcinomas. *Mol Carcinog.* **55,** 1163-1174.
97. Trost, S., Diekhof, E. K., Mohr, H., Vieker, H., Krämer, B., Wolf, C., Keil, M., Dechent, P., Binder, E. B., and Gruber, O. (2016) Investigating the Impact of a Genome-Wide Supported Bipolar Risk Variant of MAD1L1 on the Human Reward System. *Neuropsychopharmacology* **41,** 2679-2687.
98. Ge J, Dong H, Yang Y, Liu B, Zheng M, Cheng Q, Peng L, Li J. (2018) NFIX downregulation independently predicts poor prognosis in lung adenocarcinoma, but not in squamous cell carcinoma. *Future Oncol*. **14(30)**,3135-3144.
99. Piper M, Gronostajski R, Messina G. (2019) Nuclear Factor One X in Development and Disease. Trends Cell Biol. **29(1)**, 20-30.

100. Trimouille, A., Houcinat, N., Vuillaume, M. L., Fergelot, P., Boucher, C., Toutain, J., Caignec, C. L., Vincent, M., Nizon ,M., Andrieux, J., Vanlerberghe, C., Delobel, B., Duban, B., Mansour, S., Baple, E., McKeown, C., Poke, G., Robertshaw, K., Fifield, E., Fabretto, A., Pecile, V., Gasparini, P., Carrozzi, M., Lacombe, D., Arveiler, B., Rooryck, C., and Moutton, S. (2018) 19p13 microduplications encompassing NFIX are responsible for intellectual disability, short stature and small head circumference. *Eur J Hum Genet*. **26,** 85-93.
101. Vergano, S. A., van der Sluijs, P. J., and Santen, G. (2019) ARID1B-Related Disorder. In: Adam, M. P., Ardinger, H. H., Pagon, R. A., Wallace, S. E., Bean L. J .H., Stephens K., Amemiya A., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2020. Available from http://www.ncbi.nlm.nih.gov/books/NBK541502/
102. Özcan, C., Özdamar, Ö., Gökbayrak, M. E., Doğer, E., Çakıroğlu, Y., and Çine,N. (2019) HOXA-10 gene expression in ectopic and eutopic endometrium tissues: Does it differ between fertile and infertile women with endometriosis? *Eur J Obstet Gynecol Reprod Biol*. **233**, 43-48.
103. Kagami, M., Sekita, Y., Nishimura, G., Irie, M., Kato, F, Okada, M., Yamamor,,i S., Kishimoto, H., Nakayama, M., Tanaka, Y., Matsuoka, K., Takahashi, T., Noguchi, M., Tanaka, Y., Masumoto, K., Utsunomiya, T., Kouzan, H., Komatsu, Y., Ohashi, H., Kurosawa, K., Kosak,i K., Ferguson-Smith, A.C., Ishino, F., Ogata, T. (2008).Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nat Genet.* **40(2)**, 237-42.
104. Royo, H., and Cavaillé, J. (2008) Non-coding RNAs in imprinted gene clusters. *Biol Cell.* **100,** 149-166.
105. Cleaton, M. A. M., Dent, C. L., Howard, M., Corish, J. A., Gutteridge, I., Sovio, U., Gaccioli, F., Takahashi, N., Bauer, S. R., Charnock-Jones, D. S., Powell, T. L., Smith, G. C. S., Ferguson-Smith, A. C. and Charalambous, M. (2016) Fetus-derived DLK1 is required for maternal metabolic adaptations to pregnancy and is associated with fetal growth restriction. *Nature Genet*. **48**, 1473-1480.
106. Mei, B., Zhao, L., Chen, L., and Sul, H. S. (2002) Only the large soluble form of preadipocyte factor-1 (Pref-1), but not the small soluble and membrane forms, inhibits adipocyte differentiation: role of alternative splicing. *Biochem. J.* **364**, 137-144.
107. Perry, J. R. B., Day, F., Elks, C. E., Sulem, P., Thompson, D. J., Ferreira, T., He, C., Chasman, D. I., Esko, T., Thorleifsson, G., Albrecht, E., Ang, W. Q., and 192 others. (2014) Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**: 92-97.
108. Day, F. R., Thompson, D. J., Helgason, H., Chasman, D. I., Finucane, H., Sulem P, Ruth K. S., Whalen, S., Sarkar, A. K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C. M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J.J., Karlsson, R., Kolcic, I., Loh, P.R., Lunetta, K.L., Mangino, M., Marco, B., McMahon, G., Medland, S.E., Nolte, I.M., Noordam, R., Nutile, T., Paternoster, L., Perjakova, N., Porcu, E., Rose, L.M., Schraut, K.E., Segrè, A.V., Smith, A.V., Stolk, L., Teumer, A., Andrulis IL, Bandinelli S, Beckmann MW, Benitez J, Bergmann S, Bochud M, Boerwinkle E, Bojesen SE, Bolla MK, Brand JS, Brauch H, Brenner H, Broer L, Brüning T, Buring JE, Campbell H, Catamo E, Chanock S, Chenevix-Trench G, Corre T, Couch FJ, Cousminer DL, Cox A, Crisponi L, Czene K, Davey Smith G, de Geus EJCN, de Mutsert R, De Vivo I, Dennis J, Devilee P, Dos-Santos-Silva I,

Dunning AM, Eriksson JG, Fasching PA, Fernández-Rhodes L, Ferrucci L, Flesch-Janys D, Franke L, Gabrielson M, Gandin I, Giles GG, Grallert H, Gudbjartsson DF, Guénel P, Hall P, Hallberg E, Hamann U, Harris TB, Hartman CA, Heiss G, Hooning MJ, Hopper JL, Hu F, Hunter DJ, Ikram MA, Im HK, Järvelin MR, Joshi PK, Karasik D, Kellis M, Kutalik Z, LaChance G, Lambrechts D, Langenberg C, Launer LJ, Laven JSE, Lenarduzzi S, Li J, Lind PA, Lindstrom S, Liu Y, Luan J, Mägi R, Mannermaa A, Mbarek H, McCarthy MI, Meisinger C, Meitinger T, Menni C, Metspalu A, Michailidou K, Milani L, Milne RL, Montgomery GW, Mulligan AM, Nalls MA, Navarro P, Nevanlinna H, Nyholt DR, Oldehinkel AJ, O'Mara TA, Padmanabhan S, Palotie A, Pedersen N, Peters A, Peto J, Pharoah PDP, Pouta A, Radice P, Rahman I, Ring SM, Robino A, Rosendaal FR, Rudan I, Rueedi R, Ruggiero D, Sala CF, Schmidt MK, Scott RA, Shah M, Sorice R, Southey MC, Sovio U, Stampfer M, Steri M, Strauch K, Tanaka T, Tikkanen E, Timpson NJ, Traglia M, Truong T, Tyrer JP, Uitterlinden AG, Edwards DRV, Vitart V, Völker U, Vollenweider P, Wang Q, Widen E, van Dijk KW, Willemsen G, Winqvist R, Wolffenbuttel BHR, Zhao JH, Zoledziewska M, Zygmunt M, Alizadeh BZ, Boomsma DI, Ciullo M, Cucca F, Esko T, Franceschini N, Gieger C, Gudnason V, Hayward C, Kraft P, Lawlor DA, Magnusson PKE, Martin NG, Mook-Kanamori DO, Nohr EA, Polasek O, Porteous D, Price AL, Ridker PM, Snieder H, Spector TD, Stöckl D, Toniolo D, Ulivi S, Visser JA, Völzke H, Wareham NJ, Wilson JF; LifeLines Cohort Study; InterAct Consortium; kConFab/AOCS Investigators; Endometrial Cancer Association Consortium; Ovarian Cancer Association Consortium; PRACTICAL consortium, Spurdle AB, Thorsteindottir U, Pollard KS, Easton DF, Tung JY, Chang-Claude J, Hinds D, Murray A, Murabito JM, Stefansson K, Ong KK, Perry JRB. (2017) Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet.* **49,** 834-841.

109. Dauber, A., Cunha-Silva, M., Macedo, D. B., Brito, V. N., Abreu, A. P., Roberts, S. A., Montenegro, L. R., Andrew, M., Kirby, A., Weirauch, M. T., Labilloy, G., Bessa, D. S., Carroll, R. S., Jacobs, D. C., Chappell, P. E., Mendonca, B. B., Haig, D., Kaiser, U. B. and Latronico, A. C. (2017) Paternally inherited DLK1 deletion associated with familial central precocious puberty. *J. Clin. Endocr. Metab.* **102**, 1557-1567.

110. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, Bacelis J, Peng S, Hao K, Feenstra B, Wood AR, Mahajan A, Tyrrell J, Robertson NR, Rayner NW, Qiao Z, Moen GH, Vaudel M, Marsit CJ, Chen J, Nodzenski M, Schnurr TM, Zafarmand MH, Bradfield JP, Grarup N, Kooijman MN, Li-Gao R, Geller F, Ahluwalia TS, Paternoster L, Rueedi R, Huikari V, Hottenga JJ, Lyytikäinen LP, Cavadino A, Metrustry S, Cousminer DL, Wu Y, Thiering E, Wang CA, Have CT, Vilor-Tejedor N, Joshi PK, Painter JN, Ntalla I, Myhre R, Pitkänen N, van Leeuwen EM, Joro R, Lagou V, Richmond RC, Espinosa A, Barton SJ, Inskip HM, Holloway JW, Santa-Marina L, Estivill X, Ang W, Marsh JA, Reichetzeder C, Marullo L, Hocher B, Lunetta KL, Murabito JM, Relton CL, Kogevinas M, Chatzi L, Allard C, Bouchard L, Hivert MF, Zhang G, Muglia LJ, Heikkinen J; EGG Consortium, Morgen CS, van Kampen AHC, van Schaik BDC, Mentch FD, Langenberg C, Luan J, Scott RA, Zhao JH, Hemani G, Ring SM, Bennett AJ, Gaulton KJ, Fernandez-Tajes J, van Zuydam NR, Medina-Gomez C, de Haan HG, Rosendaal FR, Kutalik Z, Marques-Vidal P, Das S, Willemsen G, Mbarek H, Müller-Nurasyid M, Standl M, Appel EVR, Fonvig CE, Trier C, van Beijsterveldt CEM, Murcia M, Bustamante M, Bonas-Guarch S, Hougaard DM, Mercader JM, Linneberg A, Schraut KE, Lind PA, Medland SE, Shields BM, Knight BA, Chai JF, Panoutsopoulou K, Bartels M, Sánchez F, Stokholm J,

Torrents D, Vinding RK, Willems SM, Atalay M, Chawes BL, Kovacs P, Prokopenko I, Tuke MA, Yaghootkar H, Ruth KS, Jones SE, Loh PR, Murray A, Weedon MN, Tönjes A, Stumvoll M, Michaelsen KF, Eloranta AM, Lakka TA, van Duijn CM, Kiess W, Körner A, Niinikoski H, Pahkala K, Raitakari OT, Jacobsson B, Zeggini E, Dedoussis GV, Teo YY, Saw SM, Montgomery GW, Campbell H, Wilson JF, Vrijkotte TGM, Vrijheid M, de Geus EJCN, Hayes MG, Kadarmideen HN, Holm JC, Beilin LJ, Pennell CE, Heinrich J, Adair LS, Borja JB, Mohlke KL, Eriksson JG, Widén EE, Hattersley AT, Spector TD, Kähönen M, Viikari JS, Lehtimäki T, Boomsma DI, Sebert S, Vollenweider P, Sørensen TIA, Bisgaard H, Bønnelykke K, Murray JC, Melbye M, Nohr EA, Mook-Kanamori DO, Rivadeneira F, Hofman A, Felix JF, Jaddoe VWV, Hansen T, Pisinger C, Vaag AA, Pedersen O, Uitterlinden AG, Järvelin MR, Power C, Hyppönen E, Scholtens DM, Lowe WL Jr, Davey Smith G, Timpson NJ, Morris AP, Wareham NJ, Hakonarson H, Grant SFA, Frayling TM, Lawlor DA, Njølstad PR, Johansson S, Ong KK, McCarthy MI, Perry JRB, Evans DM, Freathy RM. (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet.* **51**, 804-814.

111. Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., Ogata, T., Yokoyama, M., Kaneko-Ishino, T., and Ishino, F. (2008) Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nature Genet.* **40,** 243-248.
112. Hernandez, A., Martinez, M. E., Croteau, W., and St. Germain, D. L. (2004) Complex organization and structure of sense and antisense transcripts expressed from the DIO3 gene imprinted locus. *Genomics* **83**, 413-424.
113. Takahashi, N., Okamoto, A., Kobayashi, R., Shirai, M., Obata, Y., Ogawa, H., Sotomaru, Y., and Kono, T. (2009) Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum. Molec. Genet*. **18**, 1879-1888.
114. Sheng, F., Sun, N., Ji, Y., Ma, Y., Ding, H., Zhang, Q., Yang, F., and Li, W. (2019) Aberrant expression of imprinted lncRNA MEG8 causes trophoblast dysfunction and abortion. *J Cell Biochem.* **120,** 17378-17390.
115. Huang, Y.T., Chu, S., Loucks, E.B., Lin, C.L., Eaton, C.B., Buka, S.L., and Kelsey, K.T. (2016) Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. *Epigenetics*. **11,** 227-236.
116. Walton, E., Hass, J., Liu, J., Roffman, J. L., Bernardoni, F., Roessner, V., Kirsch, M., Schackert, G., Calhoun, V., and Ehrlich, S. (2016) Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. *Schizophr Bull.* **42,** 406-414.
117. Ma, B., Wilker, E. H., Willis-Owen, S. A., Byun, H. M., Wong, K. C., Motta, V., Baccarelli, A. A., Schwartz, J., Cookson, W. O., Khabbaz, K., Mittleman, M.A., Moffatt, M.F., and Liang L. (2014) Predicting DNA methylation level across human tissues. *Nucleic Acids Res*. **42,** 3515-28.

Legends to Figures

Figure 1 Gene Ontology Analysis of Genes Implicate by Associated Methylation Marks

Panther version 14 was used to provide an overview of the gene ontology characterising genes implicated by methylation marks. Panther 14.0 identified 161 hits from the uploaded list of 173 genes. A. GO terms for Molecular Function found 93 molecular function hits. B. GO terms for Biological Process found 276 process hits. C.GO terms for Protein Class found 94 class hits. D. GO terms for Cellular Component found 300 cellular component hits. Colour coding has been assigned starting at 12 o'clock and working clockwise on the pie chart.

Figure 2 Implicated Genes Associated with Methylation Changes of 5% or Greater in Regulatory Elements

This figure documents those genes where a methylation change of 5% or greater has occurred within a defined regulatory feature. It also provides a summary of function and any disease associations resulting from Genome-Wide Association Studies (GWAS – culled from the GAD and EMBL genetic association catalogue databases), mutation analysis and functional investigations. The methylation change is expressed relative to the high birthweight and high length for age groups. All mapping of DMRs is based on human genome build hg38 version GRCh38 of the human genome. The positions of CpGs is given in relation to the Transcription Start Site (TSS) of the implicated gene. Also shown highlighted in red in the first column is whether the gene is associated with Trans and/or Cis-meQTLs and/or Cis-eQTMs. Where plural is shown, this indicates 2 or more Trans-meQTLs,Cis-metQTLs or Cis-eQTMs associated with the gene (information obtained from the BIOS QTL Browser at www.genenetwork.nl/biosqtlbrowser). NI= No Information. Colour key of gene disease and functional associations: brown=neurological; purple=fertility; light blue=growth and development; dark blue=oncological; light green=immunological. The asterisks mark the implicated genes found associated with methylation marks in related studies (15, 86, 87).

Figure 3 Implicated Genes Associated with Methylation Changes of 5% or Greater in Gene Bodies and Intergenic Regions

This figure documents those genes where a methylation change of 5% or greater has occurred within a gene body or intergenic region. It also provides a summary of function and any disease associations resulting from Genome-Wide Association Studies (GWAS – culled from the GAD and EMBL genetic association catalogue databases), mutation analysis and functional investigations. The methylation change is expressed relative to the high birthweight and high length for age groups. All mapping of DMRs is based on human genome build hg38 version GRCh38 of the human genome. The positions of CpGs is given in relation to the Transcription Start Site (TSS) of the implicated gene. Also shown highlighted in red in the first column is whether the gene is associated with Trans and/or Cis-meQTLs and/or Cis-eQTMs. Where plural is shown, this indicates 2 or more Trans-meQTLs,Cis-metQTLs or Cis-eQTMs associated with the gene (information obtained from the BIOS QTL Browser at www.genenetwork.nl/biosqtlbrowser). NI= No Information. Colour key of gene disease and functional associations: brown=neurological; purple=fertility; light blue=growth and development; dark blue=oncological; light green=immunological; dark green=connective tissue; orange= metabolic; vermillion=cardiovascular; grey=hearing. The asterisks mark the implicated genes found associated with methylation marks in related studies (15, 86, 87).

Figure 4 Number of Implicated Genes from Figures 2 and 3 Associated with Different Disease Categories

The colour key allows cross-reference to the gene lists in figure 2 and 3. Neur=Neuorlogical; Repro=Reproductive; Growth and Dev=Growth and Development; Onco=Oncological; Imm=Immunological; Conn Tiss=Connective Tissue; Cardio-vasc=Cardi-vascular

Legends to Tables

Table 1 Summary of Individual Specific Data of Those Included in the Study

Summary of individual specific data for subjects contributing to study. Key: Mat Age = Maternal age, GA = Gestational age, S_MOC = Season of month of conception, S_MOB = Season of month of birth, BW = Birthweight, LAZ = Length for age Z score change birth to 12 months, BL = Birth length, D = Dry season, W = Wet season, pl = Placenta, CB = Cord blood, 12mB = 12 month blood sample selected on birthweight, 12mH = 12 month blood sample selected on LAZ score. The samples categorized as high or low for both birthweight and length for age were those used in the analysis. Table 2 shows the numbers and sex for each tissue.

Table 2 Summary of DNA Samples Analysed in the Study

Numbers of DNA samples analysed for each tissue according to sex and test group and the number of of subjects in common between tissues and test groups. Key: BW = Birthweight, LAZ = Length for age Z score.

Table 3 Summary of Numbers of DMRs, CpG Loci and Implicated Genes

Summary of total number of DMRs, CpG loci, implicated genes (in brackets) and direction of median methylation change identified from the comparisons made at each time-point between groupings. Key: BW = Birthweight, LAZ = Length for age Z score. Median methylation change is expressed relative to the high birthweight and high length for age groupings.

Table 4 The distribution of methylation marks between regulatory features, gene bodies and intergenic regions.

Table 5 Methylation Analysis of the DMR Associated with the *ZFHX3* Gene by Pyrosequencing

This table compares the methylation levels determine for the DMR associated with the ZFHX3 gene by pyrosequencing and methylseq analysis. Individuals from the high (n=5) and low (n=6) 12 month length for age comparison groups were selected for analysis. The table shows the mean % methylation values and standard deviation for the two types of analysis. The

quantum and direction of change is close to that observed for median methylation change from the comparison of the high and low groups determined by Biseq analysis of methylseq data.

Table 1

Mat ID	Mat Age	GA (wks)	Parity (Cat)	S_MOC	S_MOB	BW (kg)	BW Cat	BL (cm)	LAZ Change V1-12m	LAZ Cat	Tissue
MALES											
	20.27	36	Primiparous	W	W	1.7	low	45.00	-0.45	low	Pl, CB, 12mB, 12mH
	35.19	38.1	Multiparous	D	D	2.14	low	44.43	1.55	high	Pl, 12mB, 12mH,
	33.32	38.1	Multiparous	D	D	2.32	low	47.00	0.53	mid	Pl, CB, 12mB
	38.61	38.6	Multiparous	W	D	2.38	low	45.50	-0.39	low	Pl, CB, 12mB, 12mH
	23.6	37	Primiparous	W	D	2.44	low	47.80	0.42	mid	Pl, CB, 12mB
	23.32	37.8	Multiparous	W	D	2.48	low	45.90	0	mid	Pl
	26.33	40.7	Multiparous	D	D	2.51	low	49.00	2.14	high	CB,12mB, 12mH
	27	39.9	Multiparous	W	W	2.53	low	49.00	1.7	high	Pl, CB, 12mB, 12mH
	24.39	38.9	Multiparous	D	D	2.56	low	47.40	0.99	high	Pl, CB, 12mB
	18.65	39.4	Primiparous	D	W	2.59	low	50.40	0.5	mid	Pl, CB, 12mB
	31.49	41.2	Multiparous	W	D	2.59	low	42.30	-0.49	low	Pl, 12mB, 12mH
	20.67	38.9	Primiparous	D	D	2.67	low	46.70	1.8	high	12mB, 12mH
	31.16	40.7	Multiparous	D	D	2.7	low	48.40	1.21	high	12mB, 12mH
		45		W	D	2.91	mid		0.97	high	12mH

	38.5	41	Multiparous	D	W	2.92	mid	49.00	-0.91	low	12mH
	40.82	39.7	Multiparous	D	W	2.96	mid	50.10	1.96	high	12mH
	28.69	40.7	Multiparous	D	D	3.06	mid	49.33	-0.43	low	12mH
	41.27	40.7	Multiparous	D	D	3.07	mid	52.00	-0.75	low	12mH
	32.82	39.9	Multiparous	W	D	3.13	mid	53.00	-0.66	low	12mH
	27.11	41.2	Multiparous	D	D	3.26	high	51.50	-0.49	low	12mH
	23.04	39.4	Multiparous	W	D	3.26	high	51.23	-0.91	low	12mH
	29.15	38.6	Multiparous	D	D	3.26	high	48.00	1.52	high	12mB, 12mH
	37	38.1	Multiparous	D	D	3.27	high	48.10	1.37	high	12mB, 12mH
	37.53	38.1	Multiparous	W	D	3.28	high	49.30	1.08	high	Pl, CB, 12mB, 12mH
	25.64	40.2	Multiparous	D	W	3.34	high	51.00	0.95	mid	Pl, CB, 12mB
	34.46	40.4	Multiparous	W	W	3.34	high	48.30	0.19	mid	Pl, CB, 12mB
	20.29	40.2	Multiparous	W	W	3.36	high	53.47	-1.01	low	Pl, CB, 12mB, 12mH
	37.18	39.1	Multiparous	D	D	3.36	high	50.50	1.82	high	Pl, CB, 12mB, 12mH
	22.07	41	Multiparous	D	W	3.37	high	48.47	0	mid	Pl, CB
	28.91	40.2	Multiparous	D	D	3.39	high	50.50	0	mid	Pl, CB
	23.51	38.9	Multiparous	D	W	3.45	high	50.47	1.02	high	Pl
	31.48	40.7	Multiparous	D	D	3.5	high	50.00	0.08	mid	12mB
	37.96	41.2	Multiparous	D	D	3.52	high	50.40	1.7	high	12mB, 12mH
	40.36	39.1	Multiparous	W	D	3.55	high	49.50	0.6	mid	Pl
	41.88	41	Multiparous	D	W	3.59	high	51.00	-0.33	mid	Pl, CB, 12mB

	39.33	40.7	Multiparous	D	D	3.72	high	52.50	-0.93	low	Pl, CB
	31.19	41.8	Multiparous	D	D	3.79	high	50.77	-0.28	mid	Pl, CB, 12mB
	39.69	40.2	Multiparous	D	W	3.8	high	53.00	-1.35	low	Pl, CB, 12mB, 12mH
	39.65	41	Multiparous	D	W	3.9	high	50.00	0.82	mid	Pl, CB, 12mB
	36.61		Multiparous	D	W				1.31	high	12mH
FEMALES											
	18.72	39.4	Primiparous	D	W	2.42	low	46.00	1.12	high	Pl, 12mB, 12mH,
	19.04	39.9	Primiparous	W	W	2.46	low	46.00	0.34	mid	Pl, CB, 12mB
	29.89	39.1	Multiparous	D	D	2.48	low	50.00	0.54	mid	Pl, CB
	21.16	38.9	Multiparous	D	D	2.48	low	45.50	1.86	high	12mB, 12mH
	23.01	38.7	Multiparous	D	D	2.5	low	47.00	0.11	mid	12mB
	39.72	38.6	Multiparous	D	W	2.53	low	46.40	-0.28	mid	Pl, 12mB
	22.98	38.3	Multiparous	D	D	2.54	low	49.97	1.14	high	Pl, CB, 12mB, 12mH
	38.38	39.7	Multiparous	W	D	2.56	low	47.37	0.05	mid	Pl, CB, 12mB
	29.65	40.2	Multiparous	W	D	2.58	low	45.00	0	mid	Pl, CB
	26.92	38.1	Multiparous	W	D	2.61	low	46.50	0.65	mid	Pl, CB, 12mB
	41.69	38.1	Multiparous	D	D	2.63	low	48.50	0.9	mid	Pl, 12mB
	40.49	37	Multiparous	W	W	2.64	low	47.20	1.15	high	Pl, CB, 12mB, 12mH
	19.19	38.1	Primiparous	W	W	2.67	low	46.50	-4.44	low	Pl, CB, 12mB
	38.56	38.9	Multiparous	D	W	2.69	low	45.67	1.44	high	12mH
	32.99	40.7	Multiparous	D	D	2.73	low	46.27	-0.39	low	12mH

	23.95	39.9	Multiparous	W	W	2.96	mid	50.27	-0.47	low	12mH
	22.03	42	Multiparous	D	D	2.96	mid	49.40	1.67	high	12mH
	37.22	39.4	Multiparous	W	D	2.99	mid	49.50	-0.46	low	12mH
	26.08	38.3	Multiparous	D	D	3	mid	46.00	-0.46	low	12mH
	36.2	41.2	Multiparous	D	D	3.07	mid	50.07	-0.89	low	12mH
	38.07	40.4	Multiparous	D	D	3.21	mid	50.27	-0.76	low	12mH
	34.8	38.9	Multiparous	W	D	3.25	high	53.20	1.1	high	CB
	35.4	40.4	Multiparous	D	D	3.29	high	50.50	0.78	mid	CB
	36.78	40.2	Multiparous	D	W	3.31	high	49.00	0.07	mid	Pl
	32.43	39.7	Multiparous	W	D	3.33	high	51.00	0.81	mid	Pl, CB
	24.1	39.9	Multiparous	W	D	3.33	high	50.20	0	mid	CB
	34.3	39.4	Multiparous	W	D	3.33	high	50.00	-0.67	low	CB
	34.25	41	Multiparous	W	D	3.37	high	45.17	-1.39	low	Pl, CB, 12mB, 12mH
	39.54	40.2	Multiparous	W	D	3.42	high	49.80	-1.11	low	12mB, 12mH
	27.27	39.9	Multiparous	D	D	3.75	high	53.00	-0.17	mid	Pl, CB
	38.38	40.7	Multiparous	D	D	3.84	high	47.50	0.97	high	Pl, CB, 12mB
	27.12	41	Multiparous	D	D	3.97	high	52.00	-0.87	low	Pl, 12mB, 12mH
	27.07		Multiparous	D	W				1.1	high	12mH

Placenta	DNA extraction					
	Low BW	High BW	Total			
Male	10	14	24			
Female	11	6	17			
Both	21	20	**41**			
Cordbloods	**DNA extraction**					
	Low BW	**High BW**	**Total**			
Male	8	12	20			
Female	8	8	16			
Both	16	20	**36**			
Infant blood (12m)	**DNA extraction**					
	Low BW	**High BW**	**Total**	**Low LAZ**	**High LAZ**	**Total**
Male	12	13	25	11	13	24
Female	11	4	15	9	7	16
Both	23	17	**40**	20	20	**40**

Subjects in Common	
Placenta BW and cord blood BW	30
Cordblood BW and Infant blood (12m) BW	26
Cordblood BW and Infant blood (12m) LAZ	11
Infant blood (12m) BW & Infant blood (12m) LAZ	22

Table 2

Cohorts				
	Total Number of DMRs	Direction of Median Methylation Change for DMRs Relative to High Groupings for Birthweight and LAZ		Total Number of CpG sites in DMRs and implicated genes
		+ve	-ve	
Placenta BW	4	2	2	4 (4)
Cord blood BW	68	25	43	88 (78)
Infant blood (12m) BW	54	29	25	71 (65)
Infant blood (12m) LAZ	25	13	12	31 (26)

Table 3

Genomic Feature	Number of CpGs in Feature	Number of CpGs in Feature	% of Total
	>5% median methylation change	<5% median methylation change	
Promoter	30	44	37.9
Promoter and CTCF binding site	9	3	6.2
CTCF binding site	11	3	7.2
Transcription Factor binding site	4	1	2.5
Enhancer	0	1	0.5
Exon	16	5	10.8
Intron	28	9	19.5
Intergenic	17	13	15.4

Table 4

Table 5

***ZFHX3* (% methylation)**

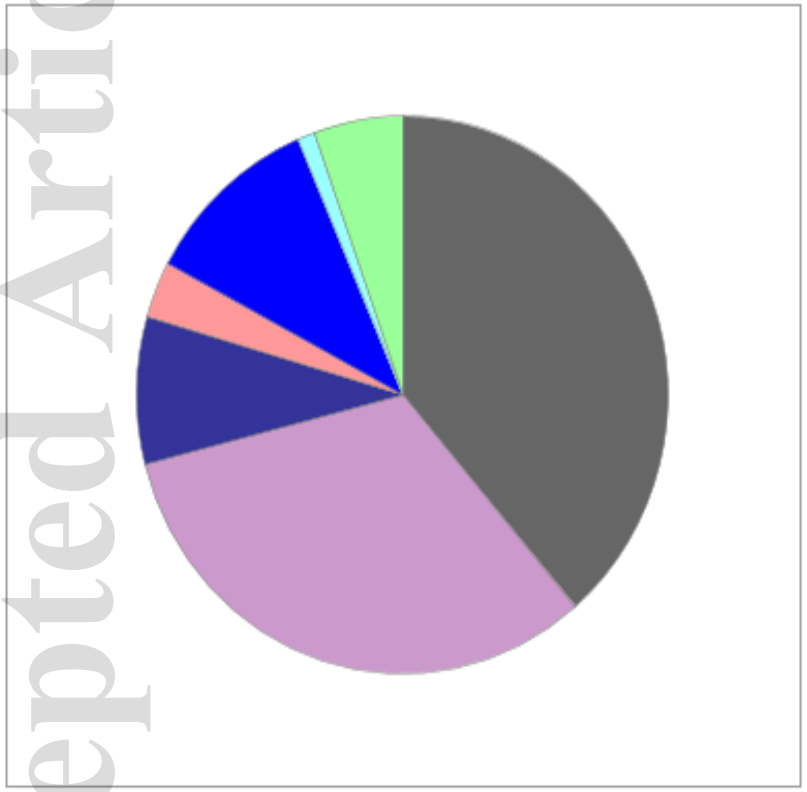
Pyrosequencing				Methylseq			
Sample	**High**	**Sample**	**Low**	**Sample**	**High**	**Sample**	**Low**
1	20	1	31	1	22	1	40
2	22	2	31	2	22	2	42
3	23	3	31	3	23	3	42
4	23	4	33	4	23	4	41
5	30	5	35	5	24	5	40
		6	35			6	46
Mean:	**23.6**	**sd+/-3.38**	**32.6**	**sd+/-1.79**	**22.8**	**sd+/- 0.74**	**41.3**	**sd+/- 1.1**

Individual samples sourced from the 12 month high and low LAZ (length for age) comparison groups

Median methylseq determined methylation change between groups = -14.8 referenced to high LAZ value

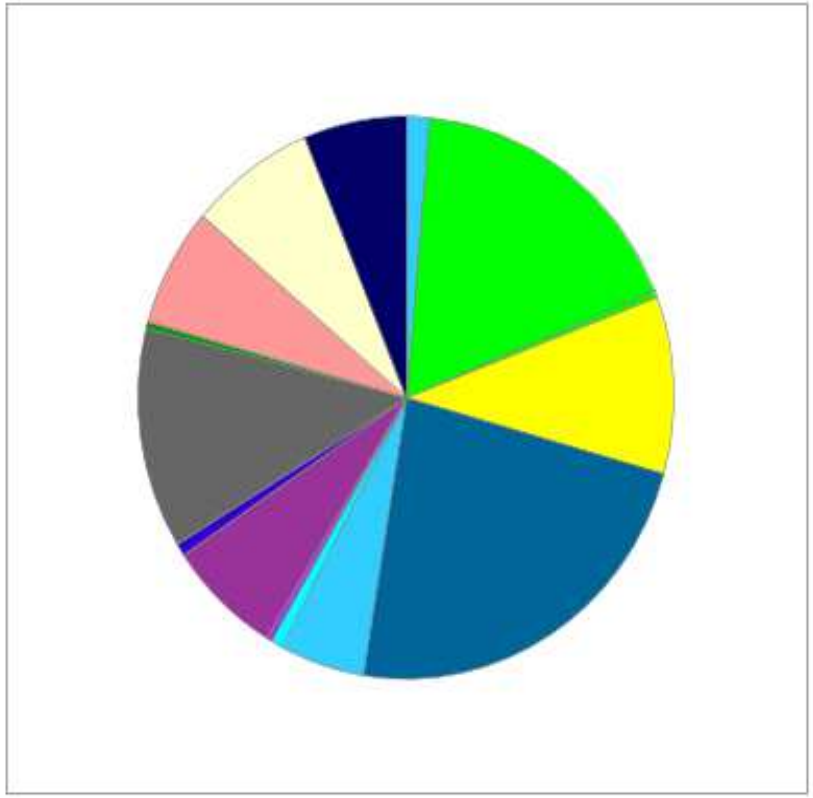
Figure 1

A. Panther GO-Slim Molecular Function



- Binding (GO:0005488)
- Catalytic Activity (GO:0003824)
- Molecular Function Regulator (GO:0098772)
- Molecular Transducer Activity (GO:0060089)
- Transcription Regulator Activity (GO:00140110)
- Translation Regulator Activity (GO:0045182)
- Transporter Activity (GO:0005215)

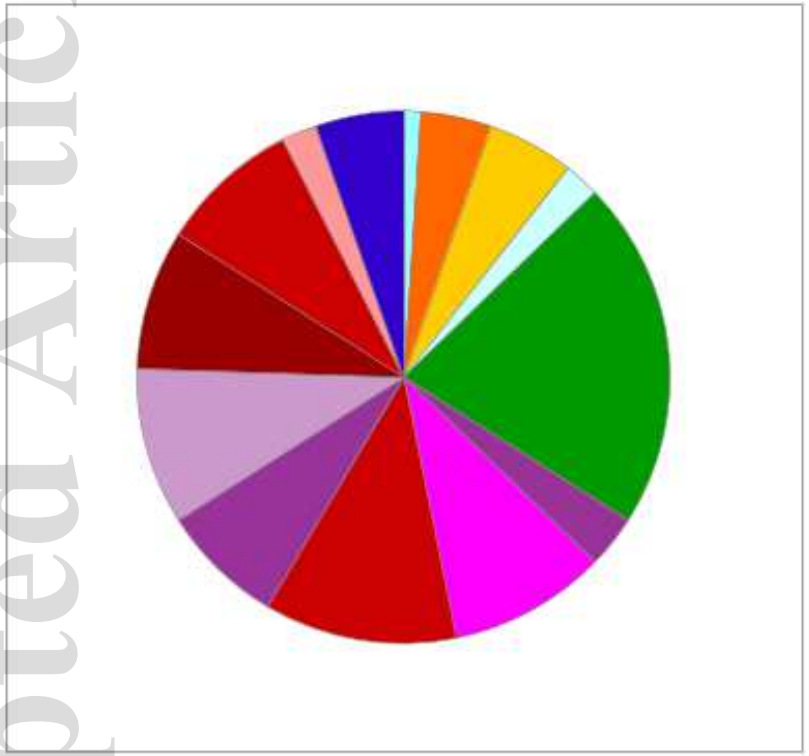
B. Panther GO-Slim Biological Process



- Biological Adhesion (GO:0022610)
- Biological Regulation (GO:0065007)
- Cell Population Proliferation (GO:0008283)
- Cellular Component Organisation or Biogenesis(GO:0071840)
- Cellular Process (GO:0009987)
- Developmental Process (GO:0032502)
- Growth (GO:0040007)
- Immune System (GO:0002376)
- Localization (GO:0051179)
- Locomotion (GO:0040011)
- Metabolic Process (GO:0008152)
- Multi-organism Process (GO:0051704)
- Multicellular Organismal Process (GO:0032501)
- Response to Stimulus (GO:0050896)
- Signalling (GO:0023052)

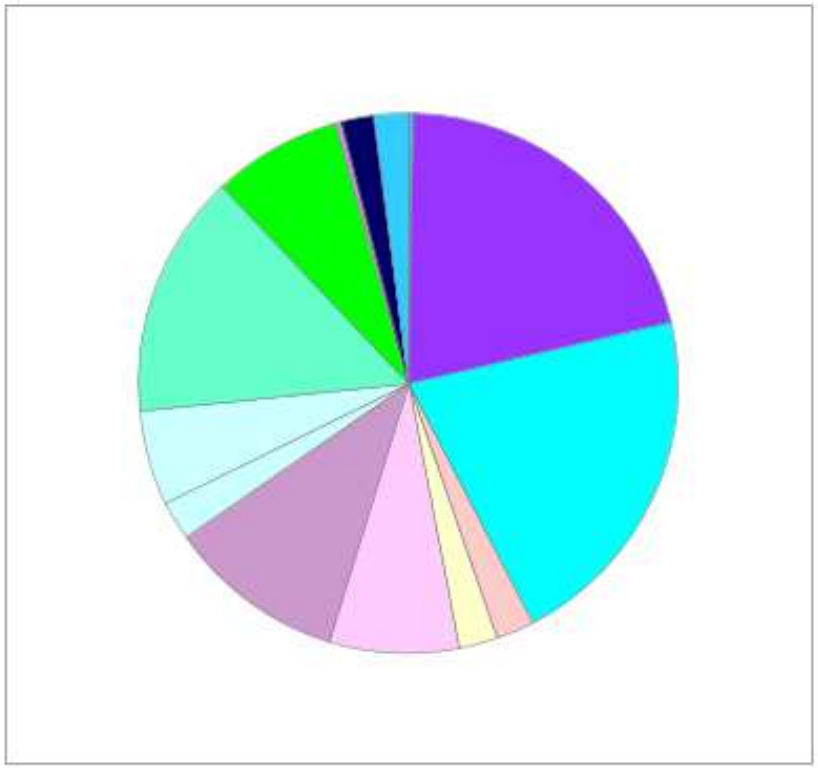
Figure 1

C. Panther GO-Slim Protein Class



- Cell Junction Protein (PC00070)
- Chromatin/Chromatin-Binding or Regulatory Protein (PC00077)
- Cytoskeletal Protein (PC00085)
- Extracellular Matrix (PC00102)
- Gene-specific Transcriptional Regulator (PC00264)
- Intracellular Signal Protein (PC00207)
- Membrane Traffic Protein (PC00150)
- Metabolite Interconversion Enzyme (PC00262)
- Nucleic Acid Binding Protein (PC00171)
- Protein Modifying Enzyme (PC00260)
- Protein-binding Activity Modulator (PC00095)
- Scaffold Adapter Protein (PC00226)
- Transmembrane Signal Receptor (PC00197)
- Transporter (PC00227)

D. Panther GO-Slim Cellular Component



- Synapse (GO:0045202)
- Cell Part (GO:0044464)
- Cell (GO:0005623)
- Extracellular Region Part (GO:004421)
- Extracellular Region (GO:0005576)
- Membrane Part (GO:0044425)
- Membrane (GO:0016020)
- Membrane-enclosed Lumen (GO:0031974)
- Organelle Part (GO:0044422)
- Organelle (GO:0043226)
- Protein-containing Complex (GO:0065007)
- Supramolecular Complex (GO:0099080)
- Synapse Part (GO:0044456)

Figure 2

Implicated Gene Name Closest to CpG	Location of Methylation Change	Median % Methylation Change and p-value	Disease Associations from GAWAS Studies	Entrez Summary; Phenotype from Functional Studies and Mutation Analysis
BiSeq Analysis >5% Methylation Change				
Trophoblast				
AL109910.1: **Long intergenic noncoding RNA** *CpG intergenic 22.152 kb of TSS*	6q27: position 170240417	-10 p=4.2E-17	NI	NI
IKBKB: **inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta*** *CpG within intron of gene 44.847 kb downstream from TSS*	8p11.2: position 42316149	+5.8 p=3.45E-08	Arthritis, Rheumatoid |,Arthritis, Rheumatoid |Rheumatoid Arthritis| Anti-TNF Response, Asthma| Bronchiolitis, Viral| Respiratory Syncytial Virus Infections, benzene haematotoxicity ,Bone Mineral Density, Bronchiolitis, Viral| Respiratory Syncytial Virus Infections, Colorectal Cancer, diabetes, type 2,Hepatitis C| Remission, Spontaneous, HIV, Hodgkin Disease| Leukemia, Lymphocytic, Chronic, B-Cell| Lymphoproliferative Disorders| Waldenstrom Macroglobulinemia, Lymphoma, Non-Hodgkin, multiple myeloma, respiratory syncytial virus bronchiolitis, hyroid cancer	The protein encoded by this gene phosphorylates the inhibitor in the inhibitor/NF-kappa-B complex, causing dissociation of the inhibitor and activation of NF-kappa-B. The encoded protein itself is found in a complex of proteins. Mutations in this gene have been found in Immunodeficiency 15A (IMD15A) and 15B (IMD15B). See OMIM entries 603258, 618204, 615592
Cord Blood Birth Weight				
LGR6: **Leucine rich repeat containing G protein-coupled receptor*** *CpG within 3' exon of gene 124.641 kb downstream of TSS* *Cis-meQTLs* *Cis-emQTMs*	1q32.1: position 202318542	-12.4 P=5.65E-07	Erythrocyte Count	This gene encodes a member of the leucine-rich repeat-containing subgroup of the G protein-coupled 7-transmembrane protein superfamily. The encoded protein is a glycoprotein hormone receptor with a large N-terminal extracellular domain that contains leucine-rich repeats important for the formation of a horseshoe-shaped interaction motif for ligand binding.
CRACD: **Cancer-Related Regulator of Actin Dynamics** *CpG within intron of gene 120.614 downstream of TSS*	4q12: position 56169687	-14.8 p=5.77E-08	Erythrocyte Count, Prostatic Neoplasms, Respiratory Function Tests	Cytoskeletal regulator that stabilizes cadherin-catenin-actin complex. Mutated in colorectal cancer. See OMIM entry 618327
SEMA3B: **Semaphorin 3B*** *CpGs in CpG island in 3' exon of gene* *Cis-eQTM*	3p21.3: positions 50276465 50276472 50276474	+5.1 p=9.76E-07	Lung cancer prostate cancer, Schizophrenia	Belongs to the class-3 semaphorin/collapsin family, whose members function in growth cone guidance during neuronal development. This family member inhibits axonal extension and has been shown to act as a tumor suppressor by inducing apoptosis. See OMIM entry 601281
PHYKPL: **5-phosphohydroxy-L-lysine phospho-lyase** *CpG within exon of gene 10.339 kb downstream of TSS*	5q35.3: position 178222463	-5 p=1.47E-08	Acquired Immunodeficiency Syndrome| Disease Progression	This is a nuclear gene encoding a mitochondrial enzyme that catalyzes the conversion of 5-phosphonooxy-L-lysine to ammonia, inorganic phosphate, and 2-aminoadipate semialdehyde. Mutations in this gene may cause phosphohydroxylysinuria. See OMIM entries 614683,

Figure 2

				615011
ANKS1A: **ankyrin repeat and sterile alpha motif domain containing 1A*** *CpGs within last exon of gene 198.866, 198.869 downstream of TSS*	6q21.31: positions 35088121 35088124	-8 p=1.03E-07	Alcoholism, Coronary Artery Disease, height, Lupus Erythematosus, Systemic| Systemic lupus erythematosus, Tobacco Use Disorder, Type 2 Diabetes| edema | rosiglitazone	May have a negative role in growth factor receptor signalling pathways. See OMIM entry 608994
TNXB: ***tenascin XB**** *CpG intergenic 35.762 kb upstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	6q21.3: position 32096415	-18.2 p=1.08E-5	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid |,Chronic renal failure |Kidney Failure, Chronic, Diabetes Mellitus, Type 1|,diabetes, type 1 ,Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes| edema | rosiglitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux –VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMIM entries 600985, 606408, 615963
AL078602.1: LncRNA *CpG within intron of transcript 163.270 kb downstream of TSS*	6q14.1: position 163834847	-8.1 p=3.06E-06	NI	NI
C7orf50: **Chromosome 7 open reading frame 50*** *CpGs in CpG island within intron* *Cis-meQTLs* *Cis-eQTMs*	7p22.3: positions 1006309 1006312	-10.6 p=1.44E-07	Longevity	Associated with Longevity.
MAD1L1: **MAD1 mitotic arrest deficient like 1*** *CpG in last intron of gene 411.326 downstream of TSS* *Trans-meQTL* *Cis-meQTLs* *Cis-eQTMs*	7p22: position 1821917	-6.9 p=7.64E-06	Bipolar Disorder, Carcinoma, Hepatocellular| Liver Neoplasms| Neoplasm Recurrence, Local, Chronic renal failure| Kidney Failure, Chronic, Iron, longevity, Lung Cancer, Mental Disorders, Myocardial Infarction, Narcolepsy, Neutrophils, Schizophrenia, Tobacco Use Disorder	MAD1L1 is a component of the mitotic spindle-assembly checkpoint that prevents the onset of anaphase until all chromosome are properly aligned at the metaphase plate. MAD1L1 functions as a homodimer and interacts with MAD2L1. MAD1L1 may play a role in cell cycle control and tumor suppression. See OMIM entry 602686
LMTK2: **Lemur tyrosine kinase 2** *CpG in CpG island 311 base pairs 3' to the gene* *Cis-meQTL* *Cis-eQTM*	7q21.3: position 98210819	+5 p=1.06E-06	Body Fat Distribution, Body Weights and Measures, Chronic renal failure| Kidney Failure, Chronic, Echocardiography, null, prostate cancer, Prostatic Neoplasms	Belongs to the protein tyrosine kinase family and interacts with several other proteins, such as Inhibitor-2 (Inh2), protein phosphatase-1 (PP1C), p35, and myosin VI. It phosporylates other proteins, and is itself also phosporylated when interacting with cyclin-dependent kinase 5 (cdk5)/p35 complex. This protein is involved in nerve growth factor (NGF)-TrkA signalling, and also plays a critical role in endosomal membrane trafficking. Mouse studies suggested an essential role of this protein in spermatogenesis. See OMIM entry 610989
ZNF783: **Zinc finger family member 783**	7: positions 149294971	+7.1 p=5.6E-07	NI	NI

Figure 2

CpGs in CpG island 32.8 and 32.817 kb downstream of TSS within intron of gene *Cis-meQTLs* *Cis-eQTMs*	149294988			
ZNF395:* Zinc finger protein 395 *CpG within intron of gene 49.401 kb downstream of TSS*	8p21.1: position 28353300	-15.5 p=7.44E-09		8p21.1: Binding factor for the papilloma virus promoter and the Huntington gene promoter. See OMIM entry 606494
NACC2*: NACC family member 2 *CpG 5' 96 bp upstream from CTCF binding site in 3'exon of gene* *Cis-meQTLs* *Cis-eQTMs*	9q34.3: position 136009696	+7.2 p=4.88E-08	Body Fat Distribution, Insulin, Stroke, Tobacco Use Disorder	Acts as a transcriptional repressor, has histone deacetylase activity and coprecipitates with subunits of the nucleosome remodelling and deacetylase (NURD) complex. See OMIM entry 615786
XPNPEP1:* X-prolyl aminopeptidase 1 *CpG within intron of gene 30.583 kb down stream of TSS* *Trans-meQTL* *Cis-eQTM*	10q25.3: position 109892970	-10.9 p=1.85E-08	Alzheimer's disease, biliary atresia, Tobacco Use Disorder	Cytosolic form of a metalloaminopeptidase that catalyzes the cleavage of the N-terminal amino acid adjacent to a proline residue and may play a role in degradation and maturation of tachykinins, neuropeptides, and peptide hormones. See OMIM entry 602443
***OR52I1:* Olfactory receptor family 52 subfamily I member 1** *CpG within exon 600 bp downstream of TSS*	11p15.4: position 4594214	-5.4 p=7.63E-07	NI	The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals.
EPS8L2*: EPS8 like 2 *CpG in intron 4.532 kb downstream of TSS* *Cis-meQTL* *Cis-eQTM* ***TMEM80*: Trans-membrane protein 80** *CpG in intron 3.542 kb downstream of TSS* *Cis-meQTL*	11p15.5: position 698970	-10.6 p=4.84E-07	EPS8L2 :Body Weight,Inflammatory Bowel Diseases	EPS8L2 protein, like other members of the family, is thought to link growth factor stimulation to actin organization, generating functional redundancy in the pathways that regulate actin cytoskeletal remodelling. Mutations in humans (DFNB106 –Deafness, autosomal recessive) and a mouse knock-out model lead to hearing loss. See OMIM entries 614988, 617637

Figure 2

Cis-eQTM				
B4GALNT4:* beta-1,4-N-acetyl-galactosaminyltransferase 4 *CpG in CpG island within exon of gene 10.697 kb downstream of TSS*	11p15.5: position 380196	-5 p=5.38E-06	NI	B4GALNT4 modifies N- or O-linked oligosaccharide structures by transferring beta-1,4-linked GalNAc to the terminal acceptors. It is most highly expressed in ovary, followed by fetal brain and various adult brain regions. B4GALNT4 is also highly expressed in foetal kidney and lung. See OMIM entry 618560
AL512484.1: LncRNA *CpG intergenic142.139 kb upstream from TSS*	13q: position 22736010	-16.6 p=6.7E-6	NI	NI
ADPRHL1:* ADP-ribosylhydrolase like 1 *CpG within last exon of gene 47.628 kb downstream of TSS* *Cis-meQTLs* *Cis-eQTMs*	13q34: position 113405896	-6.8 p=3.79E-07	NI	ADP-ribosyltransferases (see ART1; OMIM 601625) transfer ADP-ribose from NAD+ to the target protein, and ADP-ribosylhydrolases, such as ADPRHL1, reverse the reaction.
RASA3*: RAS p21 protein activator 3 *CpG within intron of gene 73.375 kb downstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	13q34: position 114055314	-13.38 p=1.07E-06	NI	Negative regulator of the RAS signalling pathway.
***SERPINA4:* Serpin family A member 4** *CpG within intron of gene 8.289 kb downstream of TSS*	14q31-q32: position 94569731	-14.4 p=3.25E-07		Inhibitors of serine proteases and play and play a role in haemostasis and thrombosis. See OMIM entry 147935
***AL355102.2:* novel protein** *CpG within intron 47.695 kb downstream of TSS*	14: position 96252539	-7 P=1.66E-6	NI	
DLK1:* Delta like non-canonical Notch ligand *CpG within 3'UTR 11.855 kb downstream of TSS*	14q32: position 100737560	+5.8 p=1.59E-06	Bone Mineral Density, diabetes, type 1 ,Disease Models, Animal| Obesity, null,Optic Disk, type 1 diabetes	This gene encodes a transmembrane protein that contains multiple epidermal growth factor repeats that functions as a regulator of cell growth. The encoded protein is involved in the differentiation of several cell types including adipocytes. This gene is located in a region of chromosome 14 frequently showing uniparental disomy, and is imprinted and expressed from the paternal allele. This imprinted region is important for growth and development and leads to Temple Syndrome when uniparental disomy is present. A single nucleotide variant in this gene is associated with child and adolescent obesity and shows polar overdominance, where heterozygotes carrying an active paternal allele express the phenotype, while mutant homozygotes are

Figure 2

				normal. See OMIM entry 176290
***GABRG3*: gamma-aminobutyric acid type A receptor gamma3 subunit** *CpGs within intron of gene 527.896, 527.887 kb downstream of TSS*	15q12: positions 27499077 27499086	+7.1 p=1.22E-07	alcohol consumption, alcohol dependence, autism, Bipolar Disorder, bipolar schizoaffective disorder, Blood Pressure Determination, Bulimia, Cholesterol, LDL, Dyskinesia, Drug-Induced|,Interleukin-6,null,Parkinson Disease, Psychiatric Disorders, schizophrenia | autism, several psychiatric disorders, Tobacco Use Disorder	GABA is the major inhibitory neurotransmitter in the mammalian brain where it acts at GABA-A receptors, which are ligand-gated chloride channels. Chloride conductance of these channels can be modulated by agents such as benzodiazepines that bind to the GABA-A receptor. GABA-A receptors are pentameric, consisting of proteins from several subunit classes: alpha, beta, gamma, delta and rho. The protein encoded by this gene is a gamma subunit, which contains the benzodiazepine binding site.
***MARVELD3:* MARVEL domain containing 3** *CpGs in intron 14.366 kb downstream of TSS*	16q22.2: positions 71640523 71640527	-9.9 p=2.04E-07	NI	Protein found colocalised with occludin at junctional complexes. Possible association with resistance to malaria. See OMIM entries 6140494, 611162.
***MBTPS1:* Membrane bound transcription factor peptidase, site 1** *CpG within intron 57.788 kb downstream of TSS*	16q23.3-q24.1: position 84059154	-5 p=6.64E-07	plasma HDL cholesterol (HDL-C) levels, Type 2 Diabetes| edema | rosiglitazone	This gene encodes a type 1 membrane bound protease which is ubiquitously expressed and regulates cholesterol or lipid homeostasis via cleavage of substrates at non-basic residues. Mutations in this gene may be associated with lysosomal dysfunction. See OMIM entries 603355, 618392
***PLIN4*: Perilipin 4** *CpG within 3'UTR of gene 15.910 kb downstream of TSS*	19p13.3: position 4502564	-15.8 p=7.54E-09	Obesity	19p13.3: Coats intracellular lipid storage droplets. Associated with obesity.
NFIX:* Nuclear factor I X *CpG within intron of gene 92.356 kb downstream of TSS* *Cis-eQTM*	19p13.3: position 13087964	-9.18 p=4.61E-07	NI	The protein encoded by this gene is a transcription factor that binds the palindromic sequence 5-TTGGCNNNNNGCCAA-3 in viral and cellular promoters. The encoded protein can also stimulate adenovirus replication in vitro. Mutations in this gene have been found in Sotos Syndrome 2 and Marshall-Smith Syndrome, both associated with over-growth, skeletal malformation and impaired neural function. See OMIM entries 164005, 614753, 602535
***NANOS2:* Nanos C2HC-type zinc finger** *CpG within exon of gene 204 bp downstream of TSS*	19q13.32: position 45914574	-19.8 p=1.87E-05	NI	Zinc finger protein expressed in testis. Inactivation of the genes in mouse model causes ablation of germ line and infertility. (66) See OMIM entry 608228
***BCOR:* BCL6 corepressor** *CpG in CpG island in first intron 74.386 kb downstream of TSS*	Xp21-p11.4: position 40102943	+11.6 p=1.25E-05	NI	Interacting corepressor of BCL6, a POZ/zinc finger transcription repressor that is required for germinal center formation and may influence apoptosis. This protein selectively interacts with the POZ domain of BCL6, but not with eight other POZ proteins. Specific class I and II histone deacetylases (HDACs) have been shown to interact with this protein, which suggests a possible link between the two classes of HDACs. Mutations in this gene have been shown to cause developmental syndromic microphthalmia-2 (MCOPS2) and oculofaciocardiodental syndrome (OFCD). See OMIM entries 300485 and 300166
Infant Blood Birth Weight				

Figure 2

PYGO2*: Pygopus family PHD finger 2 *CpG in 3' UTR of gene 5.964 kb downstream of TSS*	1q21.3: position 154957889	-6.2 p=8.65E-08	NI	The gene product is required for WNT signal transduction at the level of nuclear beta-catenin in Drosophila. It is suggested that the recruitment of human PYGO permits beta-catenin to transcriptionally activate WNT target genes and that deregulation of this interaction may play a causal role in the development of B-cell malignancies. See OMIM entry 606903
***CFAP74*: Cilia and flagella associated protein 74** *CpG within intron 18.37, 18.360 kb downstream of TSS*	1p36.33: positions 1985466 1985477	+10 p=4.2E-07	NI	Cilia and flagella associated protein with biased expression in the testis.
***KLHL23*: Kelch like family member 23** *CpG within intron of gene 21.502 kb downstream of TSS*	2q31.3: position 169715990	+5.1 p=2.94E-07	NI	NI
CPLX1*: Complexin 1 *CpG intergenic 41.512 kb downstream of TSS*	4p16.3: position 784617	+12.7 p=2.9E-07	Alcoholism, Behcet Syndrome, Schizophrenia	4p16.3: Cytosolic protein that functions in synaptic vesicle exocytosis. Binds to the SNAP receptor complex and disrupts it, allowing transmitter release.. Mutations cause Early Infantile Epileptic Encephalopathy – EIEE63. See OMIM entry 605032 (67, 68)
TNXB*: Tenascin XB *All three CpGs are located within intron of the gene 7.159, 7.146, 9.024 kb downstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	6p21.3 positions 32053494 32053507 position 32069677	+10.1 p=1.07E-08 +5.8 p=3.33E-06	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid |,Chronic renal failure |Kidney Failure, Chronic, Diabetes Mellitus, Type 1|,diabetes, type 1 ,Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes| edema | rosiglitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux –VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMIM entries 600985, 606408, 615963
DMTN*: Dematin actin binding protein *CpG within first intron of gene 6.539 kb downstream of TSS*	8p22.1: position 22055354	-5.9 p=1.39E-07	NI	The protein encoded by this gene is an actin binding and bundling protein that plays a structural role in erythrocytes, by stabilizing and attaching the spectrin/actin cytoskeleton to the erythrocyte membrane in a phosphorylation-dependent manner. Disruption of this gene in a mouse model causes severe instability of the erythrocyte membrane and hypermethylation metastasis of colorectal cancer (69)
***AC016816.1*: novel transcript LncRNA** *CpG intergenic 306.103 kb downstream of TSS and between AC016816.1 and MIR378C*	10: positions 130834812, 1300834815	-9.5 p=4.2E-9	NI	NI
MEG9*: maternally expressed 9 (*LINC00584*) *CpG within last exon of gene 4.468 kb downstream of TSS*	14Q32: position 101077251	+22.6	Body mass and age of menarche (ref)	14q32: Maternally expressed imprinted non-coding RNA (LINC00584). Functional studies indicate role in megakaryocyte differentiation and angiogenesis (70). Located in the imprinted 14q32 region involved in Temple Syndrome

Figure 2

KIF26A: **Kinesin family member 26A*** *CpGs intergenic 32.529 and 32.545 kb 3' to gene*	14q32.33: positions 104213423 104213437	+16.7 p=5.77E-08	Pancreatic Neoplasms, Stroke, Waist Circumference	Microtubule associated kinesin. Mouse KO model shows growth retardation with defective development of the bowel. See OMIM entry 613231
SGTA: **Small glutamine rich tetratricopeptide repeat containing alpha** *CpG within intron of gene 11.165 kb downstream of TSS* *Cis-meQTLs* *Cis-eQTM*	19p13.3: position 2772117	+5.7 p=9.42E-07	Insulin Resistance |Polycystic Ovary Syndrome, POLYCYSTIC OVARIAN SYNDROME |Polycystic Ovary Syndrome	*SGTA* protein interacts with the growth hormone (69) and steroid hormone (both androgen and progesterone) receptor signalling pathways (72, 73), potentially regulating growth and development of polycystic ovary syndrome, prostate and breast cancers (73). *SGTA* also promotes the proteasomal degradation of mislocalized proteins and protects Amyloid Precursor Protein from such a fate, possibly implicating it in Alzheimer's Disease (73)
MYT1: **myelin transcription factor 1.** *CpG in CpG island within first intron of gene 31.563 kb downstream of TSS*	20q13.33: position 64186999	+5 p=4.47E-06	NI	The protein encoded by this gene is a member of a family of neural specific, zinc finger-containing DNA-binding proteins. The protein binds to the promoter regions of proteolipid proteins of the central nervous system and plays a role in the developing nervous system. Inactivation of the gene in a mouse model is embryonic lethal with poor innervation of the diaphragm. Conditional pancreatic inactivation is associated with a glucose intolerance phenotype. Mutation of MYT1 has been found in a patient with Hemifacial macrosomia (HFM). See OMIM entries 600379, 164210
TAB1: **TGF-beta activated kinase 1 (MAP3K7) binding protein 1*** *CpGs in CpG island 10.653.10.646, 10.642, 10.636 kb upstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTM*	22q13.1: position 39389125 39389131 39389134 39389141	-35.6 p=3.39E-06	Arthritis, Rheumatoid |Rheumatoid Arthritis, Arthritis, Rheumatoid| Rheumatoid Arthritis| Anti-TNF Response, Crohn Disease| Crohn's disease	This gene was identified as a regulator of the MAP kinase kinase kinase MAP3K7/TAK1, which is known to mediate various intracellular signaling pathways, such as those induced by TGF beta, interleukin 1, and WNT-1. This protein interacts and thus activates TAK1 kinase. It has been shown that the C-terminal portion of this protein is sufficient for binding and activation of TAK1, while a portion of the N-terminus acts as a dominant-negative inhibitor of TGF beta, suggesting that this protein may function as a mediator between TGF beta receptors and TAK1. This protein can also interact with and activate the mitogen-activated protein kinase 14 (MAPK14/p38alpha), and thus represents an alternative activation pathway, in addition to the MAPKK pathways, which contributes to the biological responses of MAPK14 to various stimuli. See OMIM entry 602615
Infant Blood LAZ				
HLX: **H2.0 like homeobox*** *CpG intergenic 10.8 kb 5' to gene*	1q41: position 220890231	+21.3 p=2.27E-07	Aorta, Asthma, Asthma|, Blood Pressure, Body Height, Cholesterol, Cholesterol, LDL, Dupuytren Contracture, Echocardiography, Hip, Leukemia, Myeloid, Mortality	Over expression and insertional ablation in transgenic mice indicate a role in CD4+ T lymphocyte development (74) and liver and gut organogenesis (75) respectively. It is interesting to note that the gene has also been found to be expressed in cytotrophoblast cell types in early pregnancy human placentas (76), to be a regulator of trophoblast proliferation (77) and shows reduced expression associated with foetal growth restriction (78).

Figure 2

***GAR1:* ribonucleoprotein homolog (yeast) pseudogene** *CpG 781 bp upstream of TSS*	4q28: position 184560564	-7.3 p=2.89E-10	NI	NI
***IRF4:* Interferon regulatory factor 4** *CpG intergenic 10.49 kb 5' to gene* *Trans-meQTL* *Cis-meQTLs* *Cis-eQTM*	6p25.3: position 390703	-15.3 p=2.7E-07	Abortion, Spontaneous, benzene haematotoxicity, Black vs blond hair color, Black vs red hair color, Carcinoma, Basal Cell, Celiac disease Chromosome Aberrations| Chromosome abnormality| Chronic Lymphocytic Leukemia Leukemia, Lymphocytic, Chronic, B-Cell,Chromosome Aberrations| Chromosome abnormality |Lymphocytosis, Chronic lymphocytic leukemia, Chronic Lymphocytic Leukemia |Hodgkin Disease |Leukemia, Lymphocytic, Chronic, B-Cell| Multiple Myeloma, Chronic Lymphocytic Leukemia| Leukemia, Lymphocytic, Chronic, B-Cell, Chronic Lymphocytic Leukemia| Leukemia, Lymphocytic, Chronic, B-Cell |Lymphoma| Syndrome, Chronic renal failure| Kidney Failure, Chronic, Eye Color, freckles, Hair Color, HIV, Hodgkin Disease |Leukemia, Lymphocytic, Chronic, B-Cell| Lymphoproliferative Disorders| Waldenstrom Macroglobulinemia, leukemia, Leukemia, Lymphocytic, Chronic, B-Cell ,Leukemia, Lymphoid, lymphoma,melanoma ,melanoma| Nevus| Skin Neoplasms, melanoma| Nevus| Skin Neoplasms| Sunburn, Melanosis, multiple myeloma ,Multiple Sclerosis, Neuroblastoma, Skin Neoplasms, Suntan, Supranuclear Palsy, Progressive, tanning phenotype	Belongs to the IRF (interferon regulatory factor) family of transcription factors, characterized by an unique tryptophan pentad repeat DNA-binding domain. The IRFs are important in the regulation of interferons in response to infection by virus, and in the regulation of interferon-inducible genes. This family member is lymphocyte specific and negatively regulates Toll-like-receptor (TLR) signaling that is central to the activation of innate and adaptive immune systems. A chromosomal translocation involving this gene and the IgH locus, t(6;14)(p25;q32), may be a cause of multiple myeloma. See OMIM entries 601900, 612558
***ZFHX3:* Zinc finger* homeobox 3** *CpG within intron of gene 858.287 kb downstream of TSS*	16q22.2-q22.3 position 73033584	-14.8 p=6.9E-08	Alcoholism, Atrial Fibrillation, Atrial Fibrillation| Atrial Fibrillation| Brain Ischemia| Stroke, Body Mass Index, Cardiovascular Diseases, Coronary Disease, Kawasaki disease ,Mucocutaneous Lymph Node Syndrome, Myocardial Infarction| Ventricular Fibrillation, prostate cancer, Tobacco Use Disorder, Waist Circumference	*ZFHX3* (a zinc finger homeobox transcription factor - also known as *ATFB1*, Atrial fibrillation, family 1) can both repress and activate genes in cooperation with TGF beta signalling (79, 80, 81). Genetic evidence implicates the gene in a number of disease phenotypes; breast cancer through mutation and interaction with the oestrogen receptor (82, 83), gastric cancer (80) and prostate cancer (82). The important role the gene plays in myogenic differentiation (83) correlates with the genetic association to atrial fibrillation and other cardiac phenotypes. See OMIM entries 104155, 613055
***LINC00511*: long intergenic non-protein coding RNA 511.** *CpG in intron of gene 260.980 kb downstream of TSS*	17q24.3: position 72379492	-11 p=3.3E-09	NI	Long intergenic non-coding RNA involved in promoting several cancers through regulation of a variety of microRNAs. Acts as an oncogenic LINCRNA (84)

Figure 3

Implicated Gene Name Closest to CpG	Location of Methylation Change	Median % Methylation Change and p-value	Disease Associations from GAWAS Studies	Entrez Summary; Phenotype from Functional Studies and Mutation Analysis
BiSeq Analysis >5% Methylation Change				
Trophoblast				
AL109910.1: **Long intergenic noncoding RNA** *CpG intergenic 22.152 kb of TSS*	6q27: position 170240417	-10 p=4.2E-17	NI	NI
IKBKB: **inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta*** *CpG within intron of gene 44.847 kb downstream from TSS*	8p11.2: position 42316149	+5.8 p=3.45E-08	Arthritis, Rheumatoid |,Arthritis, Rheumatoid |Rheumatoid Arthritis| Anti-TNF Response, Asthma| Bronchiolitis, Viral| Respiratory Syncytial Virus Infections, benzene haematotoxicity ,Bone Mineral Density, Bronchiolitis, Viral| Respiratory Syncytial Virus Infections, Colorectal Cancer, diabetes, type 2,Hepatitis C| Remission, Spontaneous, HIV, Hodgkin Disease| Leukemia, Lymphocytic, Chronic, B-Cell| Lymphoproliferative Disorders| Waldenstrom Macroglobulinemia, Lymphoma, Non-Hodgkin, multiple myeloma, respiratory syncytial virus bronchiolitis, hyroid cancer	The protein encoded by this gene phosphorylates the inhibitor in the inhibitor/NF-kappa-B complex, causing dissociation of the inhibitor and activation of NF-kappa-B. The encoded protein itself is found in a complex of proteins. Mutations in this gene have been found in Immunodeficiency 15A (IMD15A) and 15B (IMD15B). See OMIM entries 603258, 618204, 615592
Cord Blood Birth Weight				
LGR6: **Leucine rich repeat containing G protein-coupled receptor*** *CpG within 3' exon of gene 124.641 kb downstream of TSS* *Cis-meQTLs* *Cis-emQTMs*	1q32.1: position 202318542	-12.4 P=5.65E-07	Erythrocyte Count	This gene encodes a member of the leucine-rich repeat-containing subgroup of the G protein-coupled 7-transmembrane protein superfamily. The encoded protein is a glycoprotein hormone receptor with a large N-terminal extracellular domain that contains leucine-rich repeats important for the formation of a horseshoe-shaped interaction motif for ligand binding.
CRACD: **Cancer-Related Regulator of Actin Dynamics** *CpG within intron of gene 120.614 downstream of TSS*	4q12: position 56169687	-14.8 p=5.77E-08	Erythrocyte Count, Prostatic Neoplasms, Respiratory Function Tests	Cytoskeletal regulator that stabilizes cadherin-catenin-actin complex. Mutated in colorectal cancer. See OMIM entry 618327
SEMA3B: **Semaphorin 3B*** *CpGs in CpG island in 3' exon of gene* *Cis-eQTM*	3p21.3: positions 50276465 50276472 50276474	+5.1 p=9.76E-07	Lung cancer prostate cancer, Schizophrenia	Belongs to the class-3 semaphorin/collapsin family, whose members function in growth cone guidance during neuronal development. This family member inhibits axonal extension and has been shown to act as a tumor suppressor by inducing apoptosis. See OMIM entry 601281
PHYKPL: **5-phosphohydroxy-L-lysine phospho-lyase** *CpG within exon of gene 10.339 kb downstream of TSS*	5q35.3: position 178222463	-5 p=1.47E-08	Acquired Immunodeficiency Syndrome| Disease Progression	This is a nuclear gene encoding a mitochondrial enzyme that catalyzes the conversion of 5-phosphonooxy-L-lysine to ammonia, inorganic phosphate, and 2-aminoadipate semialdehyde. Mutations in this gene may cause phosphohydroxylysinuria. See OMIM entries 614683,

Figure 3

				615011
ANKS1A: **ankyrin repeat and sterile alpha motif domain containing 1A*** *CpGs within last exon of gene 198.866, 198.869 downstream of TSS*	6q21.31: positions 35088121 35088124	-8 p=1.03E-07	Alcoholism, Coronary Artery Disease, height, Lupus Erythematosus, Systemic| Systemic lupus erythematosus, Tobacco Use Disorder, Type 2 Diabetes| edema | rosiglitazone	May have a negative role in growth factor receptor signalling pathways. See OMIM entry 608994
TNXB: tenascin XB* *CpG intergenic 35.762 kb upstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	6q21.3: position 32096415	-18.2 p=1.08E-5	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid |,Chronic renal failure |Kidney Failure, Chronic, Diabetes Mellitus, Type 1|,diabetes, type 1 ,Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes| edema | rosiglitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux –VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMIM entries 600985, 606408, 615963
AL078602.1: LncRNA *CpG within intron of transcript 163.270 kb downstream of TSS*	6q14.1: position 163834847	-8.1 p=3.06E-06	NI	NI
C7orf50: **Chromosome 7 open reading frame 50*** *CpGs in CpG island within intron* *Cis-meQTLs* *Cis-eQTMs*	7p22.3: positions 1006309 1006312	-10.6 p=1.44E-07	Longevity	Associated with Longevity.
MAD1L1: **MAD1 mitotic arrest deficient like 1*** *CpG in last intron of gene 411.326 downstream of TSS* *Trans-meQTL* *Cis-meQTLs* *Cis-eQTMs*	7p22: position 1821917	-6.9 p=7.64E-06	Bipolar Disorder, Carcinoma, Hepatocellular| Liver Neoplasms| Neoplasm Recurrence, Local, Chronic renal failure| Kidney Failure, Chronic, Iron, longevity, Lung Cancer, Mental Disorders, Myocardial Infarction, Narcolepsy, Neutrophils, Schizophrenia, Tobacco Use Disorder	MAD1L1 is a component of the mitotic spindle-assembly checkpoint that prevents the onset of anaphase until all chromosome are properly aligned at the metaphase plate. MAD1L1 functions as a homodimer and interacts with MAD2L1. MAD1L1 may play a role in cell cycle control and tumor suppression. See OMIM entry 602686
LMTK2: **Lemur tyrosine kinase 2** *CpG in CpG island 311 base pairs 3' to the gene* *Cis-meQTL* *Cis-eQTM*	7q21.3: position 98210819	+5 p=1.06E-06	Body Fat Distribution, Body Weights and Measures, Chronic renal failure| Kidney Failure, Chronic, Echocardiography, null, prostate cancer, Prostatic Neoplasms	Belongs to the protein tyrosine kinase family and interacts with several other proteins, such as Inhibitor-2 (Inh2), protein phosphatase-1 (PP1C), p35, and myosin VI. It phosporylates other proteins, and is itself also phosporylated when interacting with cyclin-dependent kinase 5 (cdk5)/p35 complex. This protein is involved in nerve growth factor (NGF)-TrkA signalling, and also plays a critical role in endosomal membrane trafficking. Mouse studies suggested an essential role of this protein in spermatogenesis. See OMIM entry 610989
ZNF783: **Zinc finger family member 783**	7: positions 149294971	+7.1 p=5.6E-07	NI	NI

Figure 3

CpGs in CpG island 32.8 and 32.817 kb downstream of TSS within intron of gene *Cis-meQTLs* *Cis-eQTMs*	149294988			
ZNF395:* Zinc finger protein 395 *CpG within intron of gene 49.401 kb downstream of TSS*	8p21.1: position 28353300	-15.5 p=7.44E-09		8p21.1: Binding factor for the papilloma virus promoter and the Huntington gene promoter. See OMIM entry 606494
NACC2*: NACC family member 2 *CpG 5' 96 bp upstream from CTCF binding site in 3'exon of gene* *Cis-meQTLs* *Cis-eQTMs*	9q34.3: position 136009696	+7.2 p=4.88E-08	Body Fat Distribution, Insulin, Stroke, Tobacco Use Disorder	Acts as a transcriptional repressor, has histone deacetylase activity and coprecipitates with subunits of the nucleosome remodelling and deacetylase (NURD) complex. See OMIM entry 615786
XPNPEP1:* X-prolyl aminopeptidase 1 *CpG within intron of gene 30.583 kb down stream of TSS* *Trans-meQTL* *Cis-eQTM*	10q25.3: position 109892970	-10.9 p=1.85E-08	Alzheimer's disease, biliary atresia, Tobacco Use Disorder	Cytosolic form of a metalloaminopeptidase that catalyzes the cleavage of the N-terminal amino acid adjacent to a proline residue and may play a role in degradation and maturation of tachykinins, neuropeptides, and peptide hormones. See OMIM entry 602443
***OR52I1:* Olfactory receptor family 52 subfamily I member 1** *CpG within exon 600 bp downstream of TSS*	11p15.4: position 4594214	-5.4 p=7.63E-07	NI	The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals.
EPS8L2*: EPS8 like 2 *CpG in intron 4.532 kb downstream of TSS* *Cis-meQTL* *Cis-eQTM* ***TMEM80*: Trans-membrane protein 80** *CpG in intron 3.542 kb downstream of TSS* *Cis-meQTL*	11p15.5: position 698970	-10.6 p=4.84E-07	EPS8L2 :Body Weight,Inflammatory Bowel Diseases	EPS8L2 protein, like other members of the family, is thought to link growth factor stimulation to actin organization, generating functional redundancy in the pathways that regulate actin cytoskeletal remodelling. Mutations in humans (DFNB106 –Deafness, autosomal recessive) and a mouse knock-out model lead to hearing loss. See OMIM entries 614988, 617637

Figure 3

Cis-eQTM				
B4GALNT4:* beta-1,4-N-acetyl-galactosaminyltransferase 4 *CpG in CpG island within exon of gene 10.697 kb downstream of TSS*	11p15.5: position 380196	-5 p=5.38E-06	NI	B4GALNT4 modifies N- or O-linked oligosaccharide structures by transferring beta-1,4-linked GalNAc to the terminal acceptors. It is most highly expressed in ovary, followed by fetal brain and various adult brain regions. B4GALNT4 is also highly expressed in foetal kidney and lung. See OMIM entry 618560
AL512484.1: LncRNA *CpG intergenic142.139 kb upstream from TSS*	13q: position 22736010	-16.6 p=6.7E-6	NI	NI
ADPRHL1:* ADP-ribosylhydrolase like 1 *CpG within last exon of gene 47.628 kb downstream of TSS* *Cis-meQTLs* *Cis-eQTMs*	13q34: position 113405896	-6.8 p=3.79E-07	NI	ADP-ribosyltransferases (see ART1; OMIM 601625) transfer ADP-ribose from NAD+ to the target protein, and ADP-ribosylhydrolases, such as ADPRHL1, reverse the reaction.
RASA3*: RAS p21 protein activator 3 *CpG within intron of gene 73.375 kb downstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	13q34: position 114055314	-13.38 p=1.07E-06	NI	Negative regulator of the RAS signalling pathway.
***SERPINA4:* Serpin family A member 4** *CpG within intron of gene 8.289 kb downstream of TSS*	14q31-q32: position 94569731	-14.4 p=3.25E-07		Inhibitors of serine proteases and play and play a role in haemostasis and thrombosis. See OMIM entry 147935
***AL355102.2:* novel protein** *CpG within intron 47.695 kb downstream of TSS*	14: position 96252539	-7 P=1.66E-6	NI	
DLK1:* Delta like non-canonical Notch ligand *CpG within 3'UTR 11.855 kb downstream of TSS*	14q32: position 100737560	+5.8 p=1.59E-06	Bone Mineral Density, diabetes, type 1 ,Disease Models, Animal| Obesity, null,Optic Disk, type 1 diabetes	This gene encodes a transmembrane protein that contains multiple epidermal growth factor repeats that functions as a regulator of cell growth. The encoded protein is involved in the differentiation of several cell types including adipocytes. This gene is located in a region of chromosome 14 frequently showing uniparental disomy, and is imprinted and expressed from the paternal allele. This imprinted region is important for growth and development and leads to Temple Syndrome when uniparental disomy is present. A single nucleotide variant in this gene is associated with child and adolescent obesity and shows polar overdominance, where heterozygotes carrying an active paternal allele express the phenotype, while mutant homozygotes are

Figure 3

				normal. See OMIM entry 176290
***GABRG3*: gamma-aminobutyric acid type A receptor gamma3 subunit** *CpGs within intron of gene 527.896, 527.887 kb downstream of TSS*	15q12: positions 27499077 27499086	+7.1 p=1.22E-07	alcohol consumption, alcohol dependence, autism, Bipolar Disorder, bipolar schizoaffective disorder, Blood Pressure Determination, Bulimia, Cholesterol, LDL, Dyskinesia, Drug-Induced|,Interleukin-6,null,Parkinson Disease, Psychiatric Disorders, schizophrenia | autism, several psychiatric disorders, Tobacco Use Disorder	GABA is the major inhibitory neurotransmitter in the mammalian brain where it acts at GABA-A receptors, which are ligand-gated chloride channels. Chloride conductance of these channels can be modulated by agents such as benzodiazepines that bind to the GABA-A receptor. GABA-A receptors are pentameric, consisting of proteins from several subunit classes: alpha, beta, gamma, delta and rho. The protein encoded by this gene is a gamma subunit, which contains the benzodiazepine binding site.
***MARVELD3*: MARVEL domain containing 3** *CpGs in intron 14.366 kb downstream of TSS*	16q22.2: positions 71640523 71640527	-9.9 p=2.04E-07	NI	Protein found colocalised with occludin at junctional complexes. Possible association with resistance to malaria. See OMIM entries 6140494, 611162.
***MBTPS1*: Membrane bound transcription factor peptidase, site 1** *CpG within intron 57.788 kb downstream of TSS*	16q23.3-q24.1: position 84059154	-5 p=6.64E-07	plasma HDL cholesterol (HDL-C) levels, Type 2 Diabetes| edema | rosiglitazone	This gene encodes a type 1 membrane bound protease which is ubiquitously expressed and regulates cholesterol or lipid homeostasis via cleavage of substrates at non-basic residues. Mutations in this gene may be associated with lysosomal dysfunction. See OMIM entries 603355, 618392
***PLIN4*: Perilipin 4** *CpG within 3'UTR of gene 15.910 kb downstream of TSS*	19p13.3: position 4502564	-15.8 p=7.54E-09	Obesity	19p13.3: Coats intracellular lipid storage droplets. Associated with obesity.
NFIX*: Nuclear factor I X *CpG within intron of gene 92.356 kb downstream of TSS* *Cis-eQTM*	19p13.3: position 13087964	-9.18 p=4.61E-07	NI	The protein encoded by this gene is a transcription factor that binds the palindromic sequence 5-TTGGCNNNNNGCCAA-3 in viral and cellular promoters. The encoded protein can also stimulate adenovirus replication in vitro. Mutations in this gene have been found in Sotos Syndrome 2 and Marshall-Smith Syndrome, both associated with over-growth, skeletal malformation and impaired neural function. See OMIM entries 164005, 614753, 602535
***NANOS2*: Nanos C2HC-type zinc finger** *CpG within exon of gene 204 bp downstream of TSS*	19q13.32: position 45914574	-19.8 p=1.87E-05	NI	Zinc finger protein expressed in testis. Inactivation of the genes in mouse model causes ablation of germ line and infertility. (66) See OMIM entry 608228
***BCOR*: BCL6 corepressor** *CpG in CpG island in first intron 74.386 kb downstream of TSS*	Xp21-p11.4: position 40102943	+11.6 p=1.25E-05	NI	Interacting corepressor of BCL6, a POZ/zinc finger transcription repressor that is required for germinal center formation and may influence apoptosis. This protein selectively interacts with the POZ domain of BCL6, but not with eight other POZ proteins. Specific class I and II histone deacetylases (HDACs) have been shown to interact with this protein, which suggests a possible link between the two classes of HDACs. Mutations in this gene have been shown to cause developmental syndromic microphthalmia-2 (MCOPS2) and oculofaciocardiodental syndrome (OFCD). See OMIM entries 300485 and 300166
Infant Blood Birth Weight				

Figure 3

PYGO2*: Pygopus family PHD finger 2 *CpG in 3' UTR of gene 5.964 kb downstream of TSS*	1q21.3: position 154957889	-6.2 p=8.65E-08	NI	The gene product is required for WNT signal transduction at the level of nuclear beta-catenin in Drosophila. It is suggested that the recruitment of human PYGO permits beta-catenin to transcriptionally activate WNT target genes and that deregulation of this interaction may play a causal role in the development of B-cell malignancies. See OMIM entry 606903
***CFAP74*: Cilia and flagella associated protein 74** *CpG within intron 18.37, 18.360 kb downstream of TSS*	1p36.33: positions 1985466 1985477	+10 p=4.2E-07	NI	Cilia and flagella associated protein with biased expression in the testis.
***KLHL23*: Kelch like family member 23** *CpG within intron of gene 21.502 kb downstream of TSS*	2q31.3: position 169715990	+5.1 p=2.94E-07	NI	NI
CPLX1*: Complexin 1 *CpG intergenic 41.512 kb downstream of TSS*	4p16.3: position 784617	+12.7 p=2.9E-07	Alcoholism, Behcet Syndrome, Schizophrenia	4p16.3: Cytosolic protein that functions in synaptic vesicle exocytosis. Binds to the SNAP receptor complex and disrupts it, allowing transmitter release.. Mutations cause Early Infantile Epileptic Encephalopathy – EIEE63. See OMIM entry 605032 (67, 68)
TNXB*: Tenascin XB *All three CpGs are located within intron of the gene 7.159, 7.146, 9.024 kb downstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTMs*	6p21.3 positions 32053494 32053507 position 32069677	+10.1 p=1.07E-08 +5.8 p=3.33E-06	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid |,Chronic renal failure |Kidney Failure, Chronic, Diabetes Mellitus, Type 1|,diabetes, type 1 ,Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes| edema | rosiglitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux –VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMIM entries 600985, 606408, 615963
DMTN*: Dematin actin binding protein *CpG within first intron of gene 6.539 kb downstream of TSS*	8p22.1: position 22055354	-5.9 p=1.39E-07	NI	The protein encoded by this gene is an actin binding and bundling protein that plays a structural role in erythrocytes, by stabilizing and attaching the spectrin/actin cytoskeleton to the erythrocyte membrane in a phosphorylation-dependent manner. Disruption of this gene in a mouse model causes severe instability of the erythrocyte membrane and hypermethylation metastasis of colorectal cancer (69)
***AC016816.1*: novel transcript LncRNA** *CpG intergenic 306.103 kb downstream of TSS and between AC016816.1 and MIR378C*	10: positions 130834812, 1300834815	-9.5 p=4.2E-9	NI	NI
MEG9*: maternally expressed 9 (*LINC00584*) *CpG within last exon of gene 4.468 kb downstream of TSS*	14Q32: position 101077251	+22.6	Body mass and age of menarche (ref)	14q32: Maternally expressed imprinted non-coding RNA (LINC00584). Functional studies indicate role in megakaryocyte differentiation and angiogenesis (70). Located in the imprinted 14q32 region involved in Temple Syndrome

Figure 3

KIF26A: **Kinesin family member 26A*** *CpGs intergenic 32.529 and 32.545 kb 3' to gene*	14q32.33: positions 104213423 104213437	+16.7 p=5.77E-08	Pancreatic Neoplasms, Stroke, Waist Circumference	Microtubule associated kinesin. Mouse KO model shows growth retardation with defective development of the bowel. See OMIM entry 613231
SGTA: **Small glutamine rich tetratricopeptide repeat containing alpha** *CpG within intron of gene 11.165 kb downstream of TSS* *Cis-meQTLs* *Cis-eQTM*	19p13.3: position 2772117	+5.7 p=9.42E-07	Insulin Resistance |Polycystic Ovary Syndrome, POLYCYSTIC OVARIAN SYNDROME |Polycystic Ovary Syndrome	*SGTA* protein interacts with the growth hormone (69) and steroid hormone (both androgen and progesterone) receptor signalling pathways (72, 73), potentially regulating growth and development of polycystic ovary syndrome, prostate and breast cancers (73). *SGTA* also promotes the proteasomal degradation of mislocalized proteins and protects Amyloid Precursor Protein from such a fate, possibly implicating it in Alzheimer's Disease (73)
MYT1: **myelin transcription factor 1.** *CpG in CpG island within first intron of gene 31.563 kb downstream of TSS*	20q13.33: position 64186999	+5 p=4.47E-06	NI	The protein encoded by this gene is a member of a family of neural specific, zinc finger-containing DNA-binding proteins. The protein binds to the promoter regions of proteolipid proteins of the central nervous system and plays a role in the developing nervous system. Inactivation of the gene in a mouse model is embryonic lethal with poor innervation of the diaphragm. Conditional pancreatic inactivation is associated with a glucose intolerance phenotype. Mutation of MYT1 has been found in a patient with Hemifacial macrosomia (HFM). See OMIM entries 600379, 164210
TAB1: **TGF-beta activated kinase 1 (MAP3K7) binding protein 1*** *CpGs in CpG island 10.653.10.646, 10.642, 10.636 kb upstream of TSS* *Trans-meQTLs* *Cis-meQTLs* *Cis-eQTM*	22q13.1: position 39389125 39389131 39389134 39389141	-35.6 p=3.39E-06	Arthritis, Rheumatoid |Rheumatoid Arthritis, Arthritis, Rheumatoid| Rheumatoid Arthritis| Anti-TNF Response, Crohn Disease| Crohn's disease	This gene was identified as a regulator of the MAP kinase kinase kinase MAP3K7/TAK1, which is known to mediate various intracellular signaling pathways, such as those induced by TGF beta, interleukin 1, and WNT-1. This protein interacts and thus activates TAK1 kinase. It has been shown that the C-terminal portion of this protein is sufficient for binding and activation of TAK1, while a portion of the N-terminus acts as a dominant-negative inhibitor of TGF beta, suggesting that this protein may function as a mediator between TGF beta receptors and TAK1. This protein can also interact with and activate the mitogen-activated protein kinase 14 (MAPK14/p38alpha), and thus represents an alternative activation pathway, in addition to the MAPKK pathways, which contributes to the biological responses of MAPK14 to various stimuli. See OMIM entry 602615
Infant Blood LAZ				
HLX: **H2.0 like homeobox*** *CpG intergenic 10.8 kb 5' to gene*	1q41: position 220890231	+21.3 p=2.27E-07	Aorta, Asthma, Asthma|, Blood Pressure, Body Height, Cholesterol, Cholesterol, LDL, Dupuytren Contracture, Echocardiography, Hip, Leukemia, Myeloid, Mortality	Over expression and insertional ablation in transgenic mice indicate a role in CD4+ T lymphocyte development (74) and liver and gut organogenesis (75) respectively. It is interesting to note that the gene has also been found to be expressed in cytotrophoblast cell types in early pregnancy human placentas (76), to be a regulator of trophoblast proliferation (77) and shows reduced expression associated with foetal growth restriction (78).

Figure 3

***GAR1:* ribonucleoprotein homolog (yeast) pseudogene** *CpG 781 bp upstream of TSS*	4q28: position 184560564	-7.3 p=2.89E-10	NI	NI
***IRF4:* Interferon regulatory factor 4** *CpG intergenic 10.49 kb 5' to gene* *Trans-meQTL* *Cis-meQTLs* *Cis-eQTM*	6p25.3: position 390703	-15.3 p=2.7E-07	Abortion, Spontaneous, benzene haematotoxicity, Black vs blond hair color, Black vs red hair color, Carcinoma, Basal Cell, Celiac disease Chromosome Aberrations| Chromosome abnormality| Chronic Lymphocytic Leukemia Leukemia, Lymphocytic, Chronic, B-Cell,Chromosome Aberrations| Chromosome abnormality |Lymphocytosis, Chronic lymphocytic leukemia, Chronic Lymphocytic Leukemia |Hodgkin Disease |Leukemia, Lymphocytic, Chronic, B-Cell| Multiple Myeloma, Chronic Lymphocytic Leukemia| Leukemia, Lymphocytic, Chronic, B-Cell, Chronic Lymphocytic Leukemia| Leukemia, Lymphocytic, Chronic, B-Cell |Lymphoma| Syndrome, Chronic renal failure| Kidney Failure, Chronic, Eye Color, freckles, Hair Color, HIV, Hodgkin Disease |Leukemia, Lymphocytic, Chronic, B-Cell| Lymphoproliferative Disorders| Waldenstrom Macroglobulinemia, leukemia, Leukemia, Lymphocytic, Chronic, B-Cell ,Leukemia, Lymphoid, lymphoma,melanoma ,melanoma| Nevus| Skin Neoplasms, melanoma| Nevus| Skin Neoplasms| Sunburn, Melanosis, multiple myeloma ,Multiple Sclerosis, Neuroblastoma, Skin Neoplasms, Suntan, Supranuclear Palsy, Progressive, tanning phenotype	Belongs to the IRF (interferon regulatory factor) family of transcription factors, characterized by an unique tryptophan pentad repeat DNA-binding domain. The IRFs are important in the regulation of interferons in response to infection by virus, and in the regulation of interferon-inducible genes. This family member is lymphocyte specific and negatively regulates Toll-like-receptor (TLR) signaling that is central to the activation of innate and adaptive immune systems. A chromosomal translocation involving this gene and the IgH locus, t(6;14)(p25;q32), may be a cause of multiple myeloma. See OMIM entries 601900, 612558
***ZFHX3:* Zinc finger* homeobox 3** *CpG within intron of gene 858.287 kb downstream of TSS*	16q22.2-q22.3 position 73033584	-14.8 p=6.9E-08	Alcoholism, Atrial Fibrillation, Atrial Fibrillation| Atrial Fibrillation| Brain Ischemia| Stroke, Body Mass Index, Cardiovascular Diseases, Coronary Disease, Kawasaki disease ,Mucocutaneous Lymph Node Syndrome, Myocardial Infarction| Ventricular Fibrillation, prostate cancer, Tobacco Use Disorder, Waist Circumference	*ZFHX3* (a zinc finger homeobox transcription factor - also known as *ATFB1*, Atrial fibrillation, family 1) can both repress and activate genes in cooperation with TGF beta signalling (79, 80, 81). Genetic evidence implicates the gene in a number of disease phenotypes; breast cancer through mutation and interaction with the oestrogen receptor (82, 83), gastric cancer (80) and prostate cancer (82). The important role the gene plays in myogenic differentiation (83) correlates with the genetic association to atrial fibrillation and other cardiac phenotypes. See OMIM entries 104155, 613055
***LINC00511*: long intergenic non-protein coding RNA 511.** *CpG in intron of gene 260.980 kb downstream of TSS*	17q24.3: position 72379492	-11 p=3.3E-09	NI	Long intergenic non-coding RNA involved in promoting several cancers through regulation of a variety of microRNAs. Acts as an oncogenic LINCRNA (84)

DMR Associated Genes	Neur	Repro	Growth and Dev	Onco	Imm	ConnTiss	Cardio-vasc	Vision	Hearing	Metabolic
Trophoblast	2				1					
Cord Blood Birthweight	8	3	4	4	1	1			1	3
Infant Blood Birthweight	7		4	3	2			1		
Infant Blood Height for Age	9		3	3	1		1			
Total	20	3	11	10	5	1	1	1	1	3

Figure 4