**Association between perinatal methylation of the neuronal differentiation regulator *HES1* and later childhood neurocognitive function and behaviour**

*Karen A. Lillycrop*1,6*\*, Paula M. Costello*2*\*, Ai Ling Teh*3*, Robert J. Murray*2*, Rebecca Clarke-Harris*2*, Sheila J. Barton*4*, Emma S. Garratt*2 *, Sherry Ngo*5*, Allan M. Sheppard*5*, Johnny Wong*3*, Shaillay Dogra*3*, Graham C. Burdge*2*, Cyrus Cooper*4,6,7*, Hazel M. Inskip*4*, Catharine R. Gale*4,8*, Peter D. Gluckman*3,5*, Nicholas C. Harvey*4*, Yap-Seng Chong*3,9*, Fabian Yap*10,11,12*, Michael J. Meaney*3,13,*, Anne Rifkin-Graboi*3*, Joanna D. Holbrook*3*, The EpiGen consortium, Keith M. Godfrey*4,6 \*Joint first authors

1 Centre for Biological Sciences, University of Southampton, Southampton, UK

2 Academic Unit of Human Development and Health, University of Southampton, Southampton, UK

3 Singapore Institute for Clinical Sciences (SICS), Agency for Science Technology and Research (A\*STAR), Singapore

4 MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

5 Liggins Institute, University of Auckland, Auckland, New Zealand

6 NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust and University of Southampton, Southampton, UK

7 NIHR Musculoskeletal Biomedical Research Unit, University of Oxford, Oxford, UK

8 Centre for Cognitive Ageing and Cognitive Epidemiology, Dept. of Psychology, University of Edinburgh, Edinburgh, UK

9 Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National

University of Singapore, National University Health System, Singapore

10 Department of Paediatrics, KK Women’s and Children’s Hospital, Singapore

11 Duke NUS Graduate School of Medicine, Singapore

12 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

13 Ludmer Centre for Neuroinformatics and Mental Health, McGill University, Montréal, Canada

Corresponding author:- Professor Karen Lillycrop, Centre for Biological Sciences, University of Southampton, SO17 1BJ, UK. Telephone +44 (0)23 80798663; FAX +44 (0)23 8079 5255; E-mail: [kal@soton.ac.uk](mailto:kal@soton.ac.uk)

**Background**

Early life environments induce long-term changes in neurocognitive development and behaviour. In animal models, early environmental cues affect neuropsychological phenotypes via epigenetic processes but as yet there is little direct evidence for such mechanisms in humans.

**Method**

We examined the relation between DNA methylation at birth and child neuropsychological outcomes in two culturally diverse populations using a genome-wide methylation analysis and validation by pyrosequencing.

**Results**

Within the UK Southampton Women’s Survey (SWS) we first which identified 41 differentially methylated regions of interest (DMROI) at birth associated with child’s full-scale IQ at age 4-years. Associations between *HES1* DMROI methylation and later cognitive function were confirmed by pyrosequencing in 175 SWS children. Consistent with these findings, higher *HES1* methylation was associated with higher executive memory function in a second independent group of 200 SWS seven-year olds. Finally, we examined a pathway for this relationship within a Singaporean cohort (n=108). Here, *HES1* DMROI methylation predicted differences in early infant behavior, known to be associated with academic success.  In vitro, methylation of *HES1* inhibited ETS transcription factor binding, suggesting a functional role of this site.

**Conclusions**

Thus, our findings suggest that perinatal epigenetic processes mark later neuro-cognitive function and behavior, providing support for a role of epigenetic processes in mediating the long-term consequences of early life environment on cognitive development.

**Keywords**

*HES1*; methylation, neurocognitive development, epigenetic, perinatal

**Key messages:**

An association between umbilical cord methylation of CpG loci within the *HES1* gene, a key regulator of neuronal differentiation and brain patterning, with child’s full-scale IQ age 4 years and executive function at 6 years in two independent groups of UK children was found. Methylation of the identified CpG loci within *HES1* in vitro inhibited ETS transcription factor binding, suggesting a functional role of this site. Thus, our findings suggest that perinatal epigenetic processes mark later neuro-cognitive function and behavior, providing support for a role of epigenetic processes in mediating the long-term consequences of early life environment on cognitive development.

**Introduction**

There is now substantial evidence that the quality of the early life environment both before and after birth is important for later cognitive function. Birthweight ([1](#_ENREF_1), [2](#_ENREF_2)), maternal ([3](#_ENREF_3)) or childhood ([4](#_ENREF_4)) stress, and poor nutrition ([5](#_ENREF_5), [6](#_ENREF_6)) in early life have all been linked to poorer neuro-behavioural and cognitive function in later life, but to date the mechanisms mediating these affects are largely unknown.

Experimental studies suggest that the developmental environment can influence neuropsychological function through alterations in epigenetic gene regulation. Epigenetic processes such as DNA methylation can induce changes in gene expression without a change in DNA base sequence ([7](#_ENREF_7)). Such processes are involved in cell differentiation and genomic imprinting, as well as the phenomenon of developmental plasticity in response to environmental influences ([8](#_ENREF_8)). Through these mechanisms early life environmental factors can affect the developmental trajectory, with long-term effects on gene expression and phenotypic outcome ([9](#_ENREF_9)). For example, in rodents maternal behaviour induced stable changes in DNA methylation and histone modifications in the hippocampal glucocorticoid receptor (*GR*) gene promoter in the offspring, affecting stress responses throughout the lifecourse ([10](#_ENREF_10)). In humans, the evidence for such processes is necessarily indirect. Adult suicide victims abused as children had higher *GR* methylation in post-mortem hippocampal samples compared to suicide victims with no such history ([11](#_ENREF_11)). The hippocampus is essential to both stress regulation and learning, raising the possibility that methylation changes induced in early life may affect behavioural and cognitive functioning. However, to date there have been no longitudinal studies showing that prenatal epigenetic processes are associated with childhood neurocognitive development.

While many DNA methylation patterns are tissue specific, recent studies indicate that some epigenetic marks show both inter-individual variation and some equivalence between different tissue types ([12-15](#_ENREF_12)). For example, a relationship between childhood adversity and *GR* methylation has been reported in both the hippocampus and in peripheral blood cells ([13](#_ENREF_13)), suggesting that peripheral tissues could be used to study developmentally induced epigenetic marks associated with later neuropsychological function.

To investigate whether developmentally induced epigenetic processes relate to later cognitive function, we employed an epigenome wide approach to identify methylation differences in umbilical cord genomic DNA that were associated with child’s cognitive performance at age four-years. We validated the association between perinatal methylation levels of *HES1*, a gene with a pivotal role in neuronal differentiation and the formation of organising centres within the brain ([16](#_ENREF_16), [17](#_ENREF_17)), and later cognitive function in two culturally diverse populations, demonstrating that epigenetics may mediate the long-term consequences of the early life environment on cognitive development.

**Methods**

***Southampton Women’s Survey***

The Southampton Women’s Survey (SWS) is a prospective mother-offspring cohort ([18](#_ENREF_18)). At age four years a sub-sample of participants had their full-scale IQ assessed (Wechsler Pre-School and Primary Scale of Intelligence (WPPSI-III, UK))([19](#_ENREF_19)). At seven years, a different SWS sub-sample participated in the Cambridge Neuropsychological Test Automated Battery (CANTAB® Delayed Matching to Subject (DMS), Intra-Extra Dimensional Set Shift (IED) and Spatial Span (SSP))([20](#_ENREF_20)) tests. Further details are in Supplementary Methods 1 and cohort characteristics are shown in Supplementary Table 1.

***Growing Up in Singapore Towards Healthy Outcomes (GUSTO)***

In the GUSTO prospective mother-offspring cohort study ([21](#_ENREF_21)), socio-emotional data were available for 124 one-year old infants for whom umbilical cord DNA had previously been collected. Socio-emotional behaviour was assessed via maternal report using the Infant Toddler Socio-Emotional Assessment (ITSEA)([22](#_ENREF_22)). The Externalising domain of this tool assesses early manifestations of socially disruptive behaviour such as aggression and defiance, linked with lower cognitive performance ([23](#_ENREF_23)). Further details are in Supplementary Methods 2 and cohort characteristics are shown in Supplementary Table 2.

***Whole genome methylation analysis***

Genomic DNA from SWS umbilical cord samples with later neurocognitive data at age 4-years (n=24, min and max IQ for each group: Group 1 81-99, Group 2 101-107, Group 3 113-18 and Group 4 121-122) was extracted, sonicated and methylated DNA isolated using a His-tagged MBD2b (methyl binding domain of MeCP2) protein according to the manufacturer’s instructions (MethylCollector kit, Active Motif). After methyl capture, the labelled methylated DNA and input DNA was hybridised to the Agilent Human Promoter Whole-Genome ChIP-on-chip array (G4489A; see Supplementary Methods 3) which contains probes spanning the promoter regions of 25,000 genes from -7.5 kb of the TSS to 2.5 kb downstream.

***Methylation array data analysis***

The log2 of Cy5/Cy3 values were obtained for each probe after background subtraction and processed by the Bayesian Tool for Methylation Analysis (BATMAN)([24](#_ENREF_24)). Log2 ratios of tiled probes and CpG densities in the probe and 100 nt of flanking genomic sequence are assessed to calculate likely %methylation value distributions. The mode of the distribution for each 100 nt region returned by BATMAN was used for further analysis. Examining the frequency distribution of the BATMAN output as well as the raw log2 ratios([25](#_ENREF_25)) revealed that most samples had a frequency distribution close to a beta distribution, both before and after BATMAN analysis. The peaks were mapped to the probes/genes using the Agilent identifiers.

***Identification of DMRs and DMROIs***

Differentially methylated regions (DMRs) and DMROIs were identified using WPPSI data. SWS subjects were grouped into four separate ordinal categories according to WPPSI score, with group one having the lowest scores and group four the highest. DMRs were defined as 100 nt. regions fulfilling the following criteria: i) robust regression analysis p≤0.02 (to correct for heteroscedasticity ([26](#_ENREF_26))); ii) Mann-Whitney test between WPSSI groups one vs. four giving p≤0.02 and p≤0.01 for Mann-Whitney test between WPSSI group one vs. two, or group two vs. three, or group three vs. four; iii) MethOR ≤0.667 or MethOR ≥1.5 for WPSSI group one vs. four; iv) absolute % methylation differences between WPSSI groups one and four ≥20%. The Fisher Exact test was applied to test if a Region of Interest (ROI) was enriched (i.e. number of DMRs within the ROI getting p-value<0.01) against the background. We calculated the proportion of DMRs with p-value less than 0.01 and compared it against the background. The probability was calculated as below:

Where a = # of DMR with p-value < 0.01 in the ROI

b = # of DMR with p-value < 0.01 in the entire data set

c = # of DMR with p-vale ≥ 0.01

d = # of DMR with p-value ≥ 0.01 in the entire data set

n = total number of DMRs

DMROIs were then defined as giving Fisher Exact tests p≤0.01 for 100 nucleotide regions and Mann Whitney p≤0.02 between group one vs. four and containing at least one DMR. The cut offs used to select DMRs/DMROIs were designed to be a stringent filter to prioritise genes for the pathway analysis. Pathway enrichment analysis used the MetaCoreTM network analysis suite (GeneGo Inc)([27](#_ENREF_27)), with the design of the array set as the background in the pathway analysis.

***Pyrosequencing***

Array methylation results were validated by sodium bisulphite pyrosequencing. Briefly, pyrosequencing assays were designed to sequence the individual CpG dinucleotides within the DMROI (primers are listed in Supplementary Table 3) using Pyromark Assay Design Software 2.0 (Qiagen, Hilden, Germany); assays were analysed (PSQ 96MA machine; Biotage, Uppsala, Sweden) and %methylation calculated using Pyro Q-CpG software (Biotage). The sequenced region for *HES1* encompassed only nine of fifteen CpGs in the 920bp BATMAN DMROI due to sequence design constraints.

***Electrophoretic mobility shift assays***

Electrophoretic mobility shift assays were carried out ([28](#_ENREF_28)) using 5ug of IMR32 nuclear extract (sc-2148, Santa Cruz Biotechnology, USA). Supplementary Table 3 shows oligonucleotide sequences.

***Statistical analysis of pyrosequencing data***

Statistical analysis used Stata (Statacorp) versions 11.2/12.1. Pyrosequencer methylation measurements did not approximate a Normal distribution and were transformed using Fisher-Yates Normal scores with mean of zero and standard deviation (SD) of one. Regression models were built using the child’s neuropsychological measure (at one (GUSTO), four or seven (SWS) years) as the outcome and methylation of the nine CpG’s measured as the predictor, adjusted for sex and then further adjusted for sex and either mother’s IQ (four-year WPPSI) or mother’s highest educational attainment (seven-year CANTAB, one-year ITSEA) as available; our previous studies found little additional influence of socio-economic status after controlling for mother’s IQ([29](#_ENREF_29)). Subsequently, age at assessment, birthweight, maternal smoking and parity were included as covariates. Results presented are regression coefficients (β), representing the change in neurodevelopmental outcome per SD change in % methylation, and associated P-values.

**Results**

***Characteristics of the cohorts***

The SWS cohort subjects (n=175) with four year cognitive measurements (median 4.4 years) had a median birth weight of 3.5kg (Supplementary Table 1). The 200 children from the SWS cohort with seven year cognitive measurements (median 7.0 years) had a similar birth weight distribution with a median birth weight of 3.4kg (Supplementary Table 1). The median maternal age at birth of the child and pre-pregnancy body mass index was similar in the SWS four and seven year subjects (30.4 vs 32.2 years and 24.5 vs 24.3 kg/m2, respectively). The 124 children from the GUSTO cohort had a median age of 0.99 years and a birth weight of 3.09kg (Supplementary Table 2). The median maternal age at birth and pre-pregnancy body mass index was also similar in the GUSTO cohort (31.7 years and 25.1 kg/m2).

***Identification of differentially methylated regions of interest at birth associated with later cognitive performance***

Genomic umbilical cord DNA from twenty-four SWS children was screened using the MBD array for differences in DNA methylation at birth associated with WPPSI IQ age 4 years. The subjects selected were representative of the range of WPPSI IQ measurements within the whole cohort. Statistical analysis of the data identified 41 DMROIs associated with IQ at age four-years (Supplementary Table 4; Figure 2a).

***The Diencephalon development process was significantly enriched for DMROIs***

The top pathway enriched for DMROIs in the GO process category was diencephalon development (4/71 genes, p=0.000044; Supplementary Table 5), which is important for the integration of cognitive function ([30](#_ENREF_30)). Figure 2b shows a sub-network created by direct interactions between genes contained within the diencephalon development GO process and includes four genes, *HES1*, *NR4A2* (also known as *NURR1*), *ETS1* and *TCF4* which contained DMROIs. Methylation at birth within the *HES1* DMROI, as estimated by BATMAN, was positively associated with WPPSI IQ at age four-years (Figure 2c). The associations for *NR4A2* and *TCF4* were also positive whilst the *ETS1* association was negative.

***Validation of* HES1 *DMROI***

We chose to validate *HES1* since it has been shown to play an essential role in the generation of organizing centres within the brain of the appropriate size, shape and specification by controlling the timing of cell differentiation within the CNS ([16](#_ENREF_16)). Moreover, the region of *HES1* identified as a DMROI was located 4.8kb upstream of the transcription start site (TSS), in a region of *HES1* that is evolutionary conserved, suggesting that altered methylation of this region of *HES1* may have important functional consequences for neuronal differentiation and function. The region of *HES1* identified as a DMROI was located 4.8kb upstream of the TSS. Methylation levels of nine CpGs within this region were analysed by pyrosequencing in an extended sample of 175 SWS subjects (including the 24 samples used for the MBD-array) for which four-year WPPSI data was available. The concordance of methylation values with WPPSI scores for the 100bp region within the *HES1* ROI selected for pyrosequencing validation can be seen in Figure 2d. Consistent with these findings from the MBD-array, associations were seen between the cord DNA methylation status of individual CpGs within the DMROI of *HES1* and the child’s four-year WPPSI IQ (Figure 3a). Higher %methylation of CpG2 associated with higher four-year WPPSI IQ (β=2.693, p=0.009) with a trend for CpG5 (β=1.951, p=0.072, Table 1). Adjusting for the mother’s IQ strengthened associations between the %methylation of CpGs 2 and 5 and child’s IQ (β=3.192, p=0.002; β=2.140, p=0.045, respectively; Table 1); omitting the original 24 discovery samples from these analyses had little effect on the associations (e.g. for CpG 2 revised β=3.179, p=0.005). Likewise, further adjustment for maternal smoking, BMI and parity and the child’s birthweight and age at WPPSI measurement had little effect on the magnitude and statistical significance of the association with CpG2 but for CpG5 there was an attenuation (Supplementary Table 6). The multivariate model combining *HES1* CpG2, child’s IQ and maternal IQ explained 15.7% of the WPPSI variability; similar variability was explained by models replacing CpG2 with CpG5 or CpG7 methylation. The presence of SNPs at the identified CpG sites in *HES1* were excluded by direct sequencing of this region.

***Association of cord* HES1 *methylation with executive function at age seven-years***

To determine whether the methylation status of *HES1* at birth was also associated with cognitive function at seven-years of age, methylation of the CpGs within the DMROI of *HES1* were also measured in a further subset of SWS children assessed for executive function using CANTAB®. The range and average methylation of CpGs within the DMROI of *HES1* were similar in the two groups of children assessed at ages four and seven-years (Supplementary Table 7) and the CpGS were highly correlated for both age groups with correlation coefficients between 0.31 and 0.85 (Supplementary Table 8). Higher *HES1* CpG 1, 5, 6 and 7 %methylation were associated with enhanced executive function, indicated by greater SSP span length (CpG5, β=0.168, p=0.007; CpG6, β=0.135, p=0.026; and CpG7, β=0.144, p=0.017; Table 2, Figure 3b) and greater DMS 12s delay total correct (CpG1, β=0.196, p=0.028; CpG 5, β=0.285, p=0.002; CpG6 β=0.174, 0.043 and CpG7 β=0.167, p=0.052; Table 2, Figure 3c). *HES1* CpG8 methylation was associated with IED total errors (Stage 1). The positive association seen between *HES1* methylation and CANTAB measurements was in the same direction as found between *HES1* methylation and child’s IQ at 4 years of age.

***Association of cord HES1 methylation and infant externalising behaviour in the GUSTO cohort***

Because of the strong associations found between methylation of specific CpGs within the promoter region of *HES1* and later cognitive function*,* we examined a mediating pathway for this relationship within the Singaporean GUSTO cohort and investigated whether methylation was related to socio-emotional difficulties at an earlier developmental stage, as socially disruptive behaviours have been linked with inattention and a reduced ability to learn. We specifically wanted to test the a priori hypothesis that the DMROI CpGs significantly associated with measures of cognition in the SWS cohort were also associated with measures of emotional regulation. We therefore examined the methylation status of HES1 CpGs 2, 5 and 7 in relation to externalising behaviour in the GUSTO cohort. Adjusting for the child’s sex and maternal educational attainment, higher cord DNA *HES1* CpG7 methylation was associated with a lower infant externalising score at age one-year (β=-0.068, p=0.02; Table 3, Figure 3d); CpG5 methylation had no association with externalising score, while higher CpG2 methylation had a borderline association with higher externalising score (β=0.053, p=0.05; Table 3).

***Functional significance of altered CpG methylation***

To determine whether methylation of these CpG loci had functional consequences by influencing transcription factor binding to the *HES1* promoter, electrophoretic mobility shift assays (EMSAs) were used. In the human neuroblastoma cell line IMR32, one specific protein complex bound to the *HES1* promoter -4706 to -4740 region, containing CpG sites 2-5 (Figure 4a). *In silico* analysis of this region of the *HES1* promoter using the *Predict transcription factor binding sites* (PROMO) software ([31](#_ENREF_31)) predicted that CpG5 was located within an ELK1 (part of the ETS family) binding site. Multiplexed consensus competitor EMSAs ([32](#_ENREF_32)) identified the transcription factor bound at this site as part of the ETS transcription factor family. This was confirmed by specific competitive binding with an ETS consensus sequence but not an oligonucleotide containing a mutated core ‘GGAA’ ETS binding sequence (Figure 4b). Moreover, while binding was substantially reduced in the presence of 100-fold excess of unmethylated specific competitor, it was unaffected in the presence of 100-fold excess of a specific competitor sequence containing methylated CpG5 (Figure 4c). This suggests that ETS binds preferentially to the unmethylated sequence upstream of *HES1* and methylation of CpG5 inhibits ETS binding to this locus.

**Discussion**

There has been much debate regarding the contribution of fixed genetic sequences to variation in IQ in the population, and genome-wide association studies have consistently failed to detect specific SNPs which are associated with a substantial effect ([33](#_ENREF_33)). Here we examined whether perinatal epigenetic processes contribute to cognitive development and function. We show for the first time that methylation of CpG loci in umbilical cord DNA at birth are associated with later neuropsychological outcomes. This provides novel evidence for the importance of developmental epigenetic processes in influencing later cognitive function. Using a MBD array we identified 41 DMROIs at birth associated with WPPSI IQ at age four-years. These DMROIs were associated with genes which have been previously linked to cognitive development or function such as *TCF4* (Transcription factor 4), a bHLH transcription factor deleted in Pitt-Hopkins syndrome ([34](#_ENREF_34)) where individuals exhibit severe motor dysfunction and mental retardation; *IL1RN* (interleukin 1 receptor antagonist) and *MMP3* (matrix metallopeptidase 3), where SNPs associated with these genes have been linked to cognitive decline ([35](#_ENREF_35), [36](#_ENREF_36)); and *NFE2L2* (nuclear factor erythroid 2-like 2) which is known to be decreased in the brain during oxidative stress ([37](#_ENREF_37)). Other genes containing DMROIs such as *FANK1* (fibronectin type III and ankyrin repeat doman 1), *FAM83F* (family with sequence similarity 83 member F) and *SERPINH1* (serpin peptidase inhibitor clade H) have not previously been linked to cognitive function.

Gene ontology analysis of the DMROIs revealed that the top GO process enriched amongst the DMROIs was diencephalon development. Four genes (*ETS1*, *HES1*, *TCF4*, and *NR4A2*) in the diencephalon network contained DMROIs and were included in a direct interactors sub-network. The diencephalon is a region of the brain that functions as a crucial relay and integration centre and modulates sensory, motor, and cognitive functions.

Consistent with the findings from the MBD array, sodium bisulphite pyrosequencing in a larger number of SWS subjects confirmed that higher perinatal methylation of CpGs within the *HES1* DMROI correlated with higher IQ at four-years; adjusting for maternal IQ strengthened the associations between *HES1* methylation and child’s IQ. Higher methylation of *HES1* CpGs 1, 5, 6 and 7 was also associated with higher executive function in an independent group of SWS children at seven-years, including measures of a better visual working memory, greater working memory capacity and increased proficiency in retaining selective attention (assessed by CANTAB DMS and SSP); CANTAB IED outcomes, assessing the ability to engage in deliberate, goal-directed thought and action were associated with *HES1* CpG8 methylation and there were non significant trends observed between IED outcomes and methylation of *HES1* CpGs 4,5,6 and 9. However, there were differences in the associations found within the WPPSI and CANTAB measurements, for example the methylation of CpGs 5 and 7 were associated with both child’s WPPSI IQ at four-years and executive function at seven-years, while the methylation of CpG2 was only associated with child’s IQ at four-years and not replicated in relation to executive function at seven-years of age. These differences may reflect different CpG loci within the DMORI having particular effects at different times during development (the cognitive function tests were carried out at different ages), or that the WPPSI and CANTAB tests measure related but different aspects of cognitive function. The DMROI of *HES1* does span a region of over 200bp and it will be interesting to determine the precise role that these different CpG sites play in the temporal and spatial regulation of HES1 expression.

To explore the pathway linking *HES1* methylation to later cognitive function we also examined whether methylation of *HES1* was related to socio-emotional difficulties at an earlier developmental stage, as socially disruptive behaviours have been linked with inattention and a reduced ability to learn ([38](#_ENREF_38)). Interestingly, higher *HES1* CpG7 methylation was also associated with lower externalising scores at age one-year in the independent GUSTO cohort, suggesting a possible mediating pathway between *HES1* methylation, emotional regulation and eventual cognitive ability, with the lower externalising scores reflecting a decrease in socially disruptive behaviours, consistent with the associations seen between higher *HES1* methylation and increased cognitive function at later ages. Alternatively, *HES1* methylation may impact neural function to result in both poor emotion regulation and accompanying externalising behaviour as well as cognitive difficulties. In contrast, we observed a borderline association between higher CpG2 methylation and greater externalising. Further replication of the association between *HES1* methylation and socio-emotional behaviour will be required to confirm the direction of the association and whether there is a differential effect of methylation at the different CpG loci within this region on behaviour. It would also be beneficial to examine both externalising and cognitive function outcomes in the same population of children, which may be possible in the GUSTO cohort as the children get older, in order to clarify the potential mediation by emotional regulation on *HES1* methylation and cognitive function.

HES1 is an effector of the NOTCH signalling pathway that is essential for neural development and function ([39](#_ENREF_39)). Disruption of *Notch1* signalling in *Drosophila* blocks memory consolidation ([40](#_ENREF_40), [41](#_ENREF_41)). Moreover, mice with antisense-reduced hippocampal *Notch1* mRNA and protein levels fail to sustain long-term neural potentiation([42](#_ENREF_42)). HES1,which was originally isolated as a mammalian homolog of *hairy* and *Enhancer of Split*, is an essential mediator of Notch function ([43-45](#_ENREF_43)). Loss and gain of function studies in mice show that *Hes1* is crucial for generating the correct numbers and full diversity of neurons and glial cells by maintaining neural stem cells until later stages through repression of proneural bHLH differentiation factors such as *Mash1* and *Ngn2* ([46](#_ENREF_46), [47](#_ENREF_47)).

The DMROI region in *HES1* associated with later cognitive function lies 4.8kb upstream of the TSS in the *HES1* gene; this is a region highly conserved between species ([48](#_ENREF_48)). DNaseI hypersensitive sites and H3K27ac have also been localised to this region in both embryonic stem cells and neuronal cell lines (<http://www.genome.ucsc.edu/ENCODE/>), marks associated with active enhancer elements ([49](#_ENREF_49)). Methylation of CpGs within the promoter or regulatory regions of genes is generally thought to block transcription factor binding and/or lead to the recruitment of methyl binding proteins that in turn recruit histone deacetylases to the DNA, silencing gene expression ([50](#_ENREF_50), [51](#_ENREF_51)). We found that methylation of CpG5, one of the CpGs most strongly associated with later neuropsychological function, blocked binding of an ETS transcription factor to this region. The ETS family of transcription factors comprise 30 different members, which play key roles in ontogenic processes ([52](#_ENREF_52)) including the development of the diencephalon, as does HES1. Interestingly, the MBD array also identified a DMROI within the ETS1 promoter that was associated with later cognitive function, suggesting that the interplay between these two factors may be important for cognitive development. The reciprocal relationship between *HES1*/*ETS* methylation is consistent with the results of the molecular studies showing that *HES1* methylation blocks ETS binding at the *HES1* promoter. ETS proteins initially contact DNA as a monomeric factor but they can also form homo- or hetero-dimers with other ETS proteins and/or interact with accessory proteins; dependent upon these interactions they can act as activators or repressors of gene expression ([52](#_ENREF_52)). Thus the effect of inhibiting ETS binding by methylation of specific CpGs within the promoter of *HES1* is likely to be both cell-type and developmental stage specific. This demonstration that altered methylation can affect transcription factor binding *in vitro* does suggest that methylation at these CpG loci may have functional consequences and potential implications for neuronal development and function. Although a prenatal exposure may affect both *HES1* methylation and neurocognitive outcomes through independent pathways and the methylation change observed may not directly lead to altered neurocognitive function. Further work is required to establish whether altered methylation of this region of *HES1* is causally involved in neurocognitive development.

To date, genome-wide association studies have identified mutations in *HMGA2* as having the largest impact on IQ; sequence variation within *HMGA2*, however, only alters IQ by 1.29 points ([53](#_ENREF_53)). Here we find that a one SD change in *HES1* methylation is associated with a difference in IQ score of 3.2 points at age 4 years, after controlling for the influences of gender and maternal IQ. These findings suggest that the early life environment operating through epigenetic mechanisms also makes an important contribution to subsequent variation in IQ. It has been shown that the peak enrichment for the distance between CpG and SNPs that are part of *cis*-acting methylation quantitative trait loci (mQTLs) is 45 bp from the CpG site in question ([54](#_ENREF_54)). In our subjects the presence of SNPs at the identified CpG sites in *HES1* were excluded by direct sequencing of this region but without genome-wide analysis it is not possible to exclude a genetic effect of distant SNPs which could influence the DNA methylation of a particular sequence.

There are a number of limitations to this study. Firstly in terms of the methytlome approach we used. We measured methylation differences at birth using an MBD array, while this has some advantages over the Illumina HumanMethylation450K BeadChip array in terms of greater coverage of the CpG sites within the genome. MBD capture is biased towards heavily methylated CpG rich regions. Moreover as the methylated DNA was hybridised to a Human Agilent promoter array, this limits the analysis to CpG sites located within regions relatively close to the TSS of a gene. Thus changes in DNA methylation outside this region will be missed. However, studies from both animal and humans have shown that many environmentally modifiable CpGs sites are located within the promoter regions of genes ([55-58](#_ENREF_55)). We also used a region centric approach to identify DMROIs, while this increases the likelihood of functionally relevant findings, it comes at the expense of minimising information from the smaller regions of differential methylation. Stringent cut offs were used to select DMROIs in order to prioritise genes for the pathway analysis, but nonetheless it is likely that the list will include false positives. The pathway analysis returned a network at a significance level which survived correction for multiple testing, suggesting it includes true positives; the functionally linked candidate *HES1* picked from this pathway went on to independently replicate. We also measured methylation in cord tissue at birth and it is possible that differences in *HES1* methylation may reflect differences in cellular heterogeneity within the cord tissue but even if this were the case, these studies still show that altered methylation of *HES1* in cord at birth is an effective marker of later neurocognitive function. Although recent data has shown that for some genomic regions methylation appears largely independent of tissue of origin, whereas for others there is a clear tissue-specific dependence ([59](#_ENREF_59)). For instance differential GR methylation in relation to childhood adversity was observed both in peripheral blood and the hippocampus ([11](#_ENREF_11)). It would be interesting to determine whether *HES1* methylation is associated with neurocognitive function in other perinatal tissues or in peripheral blood at later ages and whether the same assocation between *HES1* methylaition and cognitive function is observed also in brain tissue. A further limitation arises from the challenges in assessing neuropsychological function at different ages; IQ cannot be measured in infants and executive function is widely recognised to be the most important measure of neuropsychological function but cannot easily be assessed in infants and very young children. As a consequence we used different tests at different ages.

**Conclusions**

The associations between CpG methylation and neuropsychological function were found in children whose birthweight lay within the normal range and in two culturally diverse populations. The finding of a consistent association between *HES1* methylation at birth and later measures of neuropsychological function suggest that epigenetic processes are important in the regulation of genes and pathways involved in neuropsychological development. Although our data is only correlative and can only imply an association between *HES1* DMROI methylation at birth and later cognitive function. Nevertheless even if it is a noncausal association, the differential methylation of *HES1* provides an objective marker of an altered developmental trajectory at birth. This has important implications for policymakers and health professionals and strongly supports the growing emphasis on the quality of early life environment not only for optimal short-term health outcomes but also for longer health and wellbeing

**Acknowledgments**

This work was supported by grants from the UK Medical Research Council, British Heart Foundation, Arthritis Research UK, National Osteoporosis Society, International Osteoporosis Foundation, Cohen Trust, Gravida-National Research Centre for Growth and Development, Abbott Nutrition, National Institute for Health Research Musculoskeletal Biomedical Research Unit, University of Oxford and National Institute for Health Research Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Singapore National Research Foundation and Agency for Science Technology and Research (NMRC/TCR/004-NUS/2008). The funders had no role in study design, data collection and analysis, decision to publish or preparation of manuscript.

**Competing interests**

Keith Godfrey, Peter Gluckman, Cyrus Cooper and Yap Seng Chong have received travel reimbursement for speaking at conferences sponsored by companies selling nutritional and pharmaceutical products. The research groups involved in this work are part of an academic consortium that has received funding from Abbott Nutrition, Nestec and Danone.

1. Shenkin SD, Starr JM, Deary IJ. Birth weight and cognitive ability in childhood: a systematic review. Psychol Bull. 2004;130(6):989-1013.

2. Pharoah PO, Stevenson CJ, Cooke RW, Stevenson RC. Prevalence of behaviour disorders in low birthweight infants. ArchDisChild. 1994;70(4):271-4.

3. Charil A, Laplante DP, Vaillancourt C, King S. Prenatal stress and brain development. Brain Research Reviews. 2010;65(1):56-79.

4. Beers SR, De Bellis MD. Neuropsychological function in children with maltreatment-related posttraumatic stress disorder. AmJPsychiatry. 2002;159(3):483-6.

5. Galler JR, Bryce CP, Zichlin ML, Fitzmaurice G, Eaglesfield GD, Waber DP. Infant malnutrition is associated with persisting attention deficits in middle adulthood. Journal of Nutrition. 2012;142(4):788-94.

6. Grantham-McGregor S, Baker-Henningham H. Review of the evidence linking protein and energy to mental development. Public Health Nutr. 2005;8(7A):1191-201.

7. Bird A, Macleod D. Reading the DNA methylation signal. Cold Spring HarbSympQuantBiol. 2004;69:113-8.

8. Gluckman PD, Hanson MA. Developmental plasticity and human disease: research directions. JInternMed. 2007;261(5):461-71.

9. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. Br J Nutr. 2007;97(6):1064-73.

10. Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, et al. Epigenetic programming by maternal behavior. NatNeurosci. 2004;7(8):847-54.

11. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. NatNeurosci. 2009;12(3):342-8.

12. Murphy SK, Huang Z, Hoyo C. Differentially methylated regions of imprinted genes in prenatal, perinatal and postnatal human tissues. PLoS One. 2012;7(7):e40924.

13. Tyrka AR, Price LH, Marsit C, Walters OC, Carpenter LL. Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. PLoS One. 2012;7(1):e30148.

14. Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, et al. Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Res. 2012;72(9):2304-13.

15. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, et al. Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. The FASEB Journal. 2010;24(9):3135-44.

16. Baek JH, Hatakeyama J, Sakamoto S, Ohtsuka T, Kageyama R. Persistent and high levels of Hes1 expression regulate boundary formation in the developing central nervous system. Development. 2006;133(13):2467-76.

17. Kageyama R, Ohtsuka T, Kobayashi T. Roles of Hes genes in neural development. DevGrowth Differ. 2008;50 Suppl 1:S97-103.

18. Inskip HM, Godfrey KM, Robinson SM, Law CM, Barker DJ, Cooper C, et al. Cohort profile: The Southampton Women's Survey. Int J Epidemiol. 2006;35(1):42-8.

19. Gale CR, Martyn CN, Marriott LD, Limond J, Crozier S, Inskip HM, et al. Dietary patterns in infancy and cognitive and neuropsychological function in childhood. J Child Psychol Psychiatry. 2009;50(7):816-23.

20. Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, et al. A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. J Int Neuropsychol Soc. 1998;4(5):474-90.

21. Soh SE, Lee SS, Hoon SW, Tan MY, Goh A, Lee BW, et al. The methodology of the GUSTO cohort study: a novel approach in studying pediatric allergy. Asia Pac Allergy. 2012;2(2):144-8.

22. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD. The Infant-Toddler Social and Emotional Assessment (ITSEA): factor structure, reliability, and validity. J Abnorm Child Psychol. 2003;31(5):495-514.

23. Brennan LM, Shaw DS, Dishion TJ, Wilson M. Longitudinal predictors of school-age academic achievement: unique contributions of toddler-age aggression, oppositionality, inattention, and hyperactivity. J Abnorm Child Psychol. 2012;40(8):1289-300.

24. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol. 2008;26(7):779-85.

25. Palmke N, Santacruz D, Walter J. Comprehensive analysis of DNA-methylation in mammalian tissues using MeDIP-chip. Methods. 2011;53(2):175-84.

26. Barton SJ, Crozier SR, Lillycrop KA, Godfrey KM, Inskip HM. Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. BMC Genomics. 2013;14:161.

27. Nikolsky Y, Kirillov E, Zuev R, Rakhmatulin E, Nikolskaya T. Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. Methods Mol Biol. 2009;563:177-96.

28. Dent CL, Lillycrop KA, Estridge JK, Thomas NS, Latchman DS. The B-cell and neuronal forms of the octamer-binding protein Oct-2 differ in DNA-binding specificity and functional activity. MolCell Biol. 1991;11(8):3925-30.

29. Gale CR, O'Callaghan FJ, Godfrey KM, Law CM, Martyn CN. Critical periods of brain growth and cognitive function in children. Brain. 2004;127(Pt 2):321-9.

30. Critchley HD. Neural mechanisms of autonomic, affective, and cognitive integration. J Comp Neurol. 2005;493(1):154-66.

31. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics. 2002;18(2):333-4.

32. Smith AJ, Humphries SE. Characterization of DNA-binding proteins using multiplexed competitor EMSA. J Mol Biol. 2009;385(3):714-7.

33. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Molecular psychiatry. 2011;16(10):996-1005.

34. Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, et al. Severe mental retardation with breathing abnormalities (Pitt–Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. Human Molecular Genetics. 2007;16(12):1488-94.

35. Benke KS, Carlson MC, Doan BQ, Walston JD, Xue QL, Reiner AP, et al. The association of genetic variants in interleukin-1 genes with cognition: Findings from the cardiovascular health study. Experimental gerontology. 2011;46(12):1010-9.

36. Helbecque N, Cottel D, Hermant X, Amouyel P. Impact of the matrix metalloproteinase MMP-3 on dementia. Neurobiology of Aging.28(8):1215-20.

37. Morrison CD, Pistell PJ, Ingram DK, Johnson WD, Liu Y, Fernandez-Kim SO, et al. High fat diet increases hippocampal oxidative stress and cognitive impairment in aged mice: implications for decreased Nrf2 signaling. Journal of Neurochemistry. 2010;114(6):1581-9.

38. Treyvaud K, Doyle LW, Lee KJ, Roberts G, Lim J, Inder TE, et al. Social-emotional difficulties in very preterm and term 2 year olds predict specific social-emotional problems at the age of 5 years. J Pediatr Psychol. 2012;37(7):779-85.

39. Andersson ER, Sandberg R, Lendahl U. Notch signaling: simplicity in design, versatility in function. Development. 2011;138(17):3593-612.

40. Ge X, Hannan F, Xie Z, Feng C, Tully T, Zhou H, et al. Notch signaling in Drosophila long-term memory formation. Proc Natl Acad Sci U S A. 2004;101(27):10172-6.

41. Presente A, Boyles RS, Serway CN, de Belle JS, Andres AJ. Notch is required for long-term memory in Drosophila. Proc Natl Acad Sci U S A. 2004;101(6):1764-8.

42. Wang Y, Chan SL, Miele L, Yao PJ, Mackes J, Ingram DK, et al. Involvement of Notch signaling in hippocampal synaptic plasticity. Proc Natl Acad Sci U S A. 2004;101(25):9458-62.

43. Akazawa C, Sasai Y, Nakanishi S, Kageyama R. Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. J Biol Chem. 1992;267(30):21879-85.

44. Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. Genes Dev. 1992;6(12B):2620-34.

45. Nakao K, Campos-Ortega JA. Persistent expression of genes of the enhancer of split complex suppresses neural development in Drosophila. Neuron. 1996;16(2):275-86.

46. Hatakeyama J, Sakamoto S, Kageyama R. Hes1 and Hes5 regulate the development of the cranial and spinal nerve systems. Dev Neurosci. 2006;28(1-2):92-101.

47. Hatakeyama J, Kageyama R. Notch1 expression is spatiotemporally correlated with neurogenesis and negatively regulated by Notch1-independent Hes genes in the developing nervous system. Cereb Cortex. 2006;16 Suppl 1:i132-7.

48. Takebayashi K, Sasai Y, Sakai Y, Watanabe T, Nakanishi S, Kageyama R. Structure, chromosomal locus, and promoter analysis of the gene encoding the mouse helix-loop-helix factor HES-1. Negative autoregulation through the multiple N box elements. J Biol Chem. 1994;269(7):5150-6.

49. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011;470(7333):279-83.

50. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. Journal of Biological Chemistry. 2003;278(6):4035-40.

51. Bird A. The essentials of DNA methylation. Cell. 1992;70(1):5-8.

52. Sharrocks AD. The ETS-domain transcription factor family. Nat Rev Mol Cell Biol. 2001;2(11):827-37.

53. Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, et al. Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet. 2012;44(5):552-61.

54. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai S-L, et al. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. PLoS Genet. 2010;6(5):e1000952.

55. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. ProcNatlAcadSciUS A. 2008;105(44):17046-9.

56. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Human Molecular Genetics. 2009;18(21):4046-53.

57. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. BrJNutr. 2007;97(6):1064-73.

58. Plagemann A, Harder T, Brunn M, Harder A, Roepke K, Wittrock-Staar M, et al. Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. JPhysiol. 2009;587(Pt 20):4963-76.

59. Ollikainen M, Smith KR, Joo EJ-H, Ng HK, Andronikos R, Novakovic B, et al. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Human Molecular Genetics. 2010;19(21):4176-88.

**Table 1:** Association of umbilical cord *HES1* CpG methylation within the identified DMROI with 4 year WPPSI outcomes in 175 children, \*p-value<0.1 \*\*p-value≤0.05. WPPSI = Wechsler Pre-School and Primary Scale of Intelligence (full-scale IQ), WASI = Wechsler Abbreviated Scale of Intelligence (full-scale IQ), LCL = lower 95% confidence limit, UCL = upper 95% confidence limit.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **4 year WPPSI IQ, adjusted for child’s sex** | | | | | **4 year WPPSI IQ, adjusted for child’s sex and mother’s WASI** | | | | |
|  | n | β | LCL | UCL | p-value | n | β | LCL | UCL | p-value |
| ***HES1***  **CpG1** | 168 | 1.068 | -0.853 | 2.989 | *0.277* | 168 | 1.341 | -0.563 | 3.244 | 0.169 |
| **CpG2** | 157 | 2.693 | 0.694 | 4.692 | *0.009\*\** | 157 | 3.192 | *1.212* | *5.173* | *0.002\*\** |
| **CpG3** | 154 | 0.260 | -1.801 | 2.321 | *0.805* | 154 | 0.540 | *-1.510* | *2.589* | *0.606* |
| **CpG4** | 146 | 1.457 | -0.582 | 3.495 | *0.164* | 146 | 1.571 | *-0.437* | *3.579* | *0.127* |
| **CpG5** | 139 | 1.951 | -0.161 4.062 | | *0.072\** | 139 | 2.140 | *0.066* | *4.215* | *0.045\*\** |
| **CpG6** | 170 | 0.936 | -0.990 2.861 | | *0.342* | 170 | 1.041 | *-0.863* | *2.945* | *0.285* |
| **CpG7** | 155 | 1.399 | -0.623 3.421 | | *0.177* | 155 | 1.719 | *-0.292* | *3.729* | *0.096\** |
| **CpG8** | 146 | 1.014 | -1.027 3.054 | | *0.332* | 146 | 1.501 | *-0.543* | *3.546* | *0.152* |
| **CpG9** | 138 | 0.928 | -1.212 3.067 | | *0.397* | 138 | 1.512 | *-0.661* | *3.684* | *0.175* |

**Table 2:** Association of umbilical cord *HES1* CpG methylation within the identified DMROI with 7 year CANTAB outcomes in 200 children,\*p-value<0.1 \*\*p-value≤0.05. CANTAB = Cambridge Neuropsychological Test Automated Battery, LCL= lower 95% confidence limit, UCL = upper 95% confidence limit.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CANTAB outcomes adjusted for sex** | | | | | **CANTAB outcomes adjusted for sex and mother’s educational attainment** | | | | |
| **DMS total correct** | **n** | **beta** | **LCL** | **UCL** | **p-value** | **n** | **beta** | **LCL** | **UCL** | **p-value** |
| HES1 CpG1 | 187 | 0.196 | 0.022 | 0.369 | 0.028\*\* | 186 | 0.205 | 0.028 | 0.382 | 0.024\*\* |
| HES1 CpG2 | 180 | 0.121 | -0.060 | 0.302 | 0.192 | 179 | 0.122 | -0.063 | 0.306 | 0.198 |
| HES1 CpG3 | 180 | 0.175 | -0.007 | 0.357 | 0.061\* | 179 | 0.171 | -0.012 | 0.354 | 0.069\* |
| HES1 CpG4 | 179 | 0.087 | -0.098 | 0.271 | 0.359 | 178 | 0.095 | -0.094 | 0.283 | 0.327 |
| HES1 CpG5 | 179 | 0.285 | 0.111 | 0.459 | 0.002\*\* | 178 | 0.289 | 0.111 | 0.467 | 0.002\*\* |
| HES1 CpG6 | 200 | 0.174 | 0.007 | 0.342 | 0.043\*\* | 199 | 0.175 | 0.005 | 0.345 | 0.045\*\* |
| HES1 CpG7 | 199 | 0.167 | -0.001 | 0.334 | 0.052\* | 198 | 0.164 | -0.007 | 0.334 | 0.061\* |
| HES1 CpG8 | 197 | 0.151 | -0.017 | 0.319 | 0.079\* | 196 | 0.152 | -0.020 | 0.324 | 0.085\* |
| HES1 CpG9 | 197 | 0.097 | -0.072 | 0.266 | 0.261 | 196 | 0.097 | -0.075 | 0.268 | 0.272 |
| **IED total errors (Stage 1)** | **n** | **beta** | **LCL** | **UCL** | **p-value** | **n** | **beta** | **LCL** | **UCL** | **p-value** |
| HES1 CpG1 | 185 | 0.023 | -0.150 | 0.196 | 0.797 | 184 | 0.024 | -0.153 | 0.201 | 0.793 |
| HES1 CpG2 | 178 | 0.143 | -0.034 | 0.321 | 0.115 | 177 | 0.142 | -0.039 | 0.322 | 0.125 |
| HES1 CpG3 | 178 | 0.128 | -0.051 | 0.308 | 0.163 | 177 | 0.123 | -0.058 | 0.303 | 0.185 |
| HES1 CpG4 | 177 | 0.155 | -0.025 | 0.335 | 0.093\* | 176 | 0.165 | -0.020 | 0.349 | 0.082\* |
| HES1 CpG5 | 177 | 0.070 | -0.106 | 0.245 | 0.437 | 176 | 0.063 | -0.116 | 0.243 | 0.492 |
| HES1 CpG6 | 198 | 0.161 | -0.007 | 0.330 | 0.062\* | 197 | 0.161 | -0.010 | 0.332 | 0.066\* |
| HES1 CpG7 | 197 | 0.095 | -0.073 | 0.263 | 0.269 | 196 | 0.091 | -0.080 | 0.263 | 0.298 |
| HES1 CpG8 | 195 | 0.195 | 0.028 | 0.363 | 0.023\*\* | 194 | 0.198 | 0.026 | 0.369 | 0.025\*\* |
| HES1 CpG9 | 195 | 0.149 | -0.020 | 0.318 | 0.086\* | 194 | 0.148 | -0.024 | 0.320 | 0.092\* |
| **IED total errors (Stage 8)** | **n** | **beta** | **LCL** | **UCL** | **p-value** | **n** | **beta** | **LCL** | **UCL** | **p-value** |
| HES1 CpG1 | 185 | -0.007 | -0.213 | 0.199 | 0.948 | 184 | -0.013 | -0.224 | 0.197 | 0.901 |
| HES1 CpG2 | 178 | -0.089 | -0.296 | 0.118 | 0.402 | 177 | -0.088 | -0.299 | 0.122 | 0.412 |
| HES1 CpG3 | 178 | -0.092 | -0.302 | 0.117 | 0.388 | 177 | -0.087 | -0.297 | 0.124 | 0.421 |
| HES1 CpG4 | 177 | 0.134 | -0.076 | 0.344 | 0.213 | 176 | 0.131 | -0.084 | 0.346 | 0.234 |
| HES1 CpG5 | 177 | -0.106 | -0.310 | 0.098 | 0.308 | 176 | -0.103 | -0.311 | 0.106 | 0.335 |
| HES1 CpG6 | 198 | -0.098 | -0.294 | 0.097 | 0.325 | 197 | -0.098 | -0.296 | 0.099 | 0.330 |
| HES1 CpG7 | 197 | -0.050 | -0.244 | 0.143 | 0.611 | 196 | -0.040 | -0.237 | 0.156 | 0.687 |
| HES1 CpG8 | 195 | -0.050 | -0.244 | 0.144 | 0.615 | 194 | -0.045 | -0.242 | 0.153 | 0.659 |
| HES1 CpG9 | 195 | -0.089 | -0.284 | 0.105 | 0.370 | 194 | -0.088 | -0.284 | 0.109 | 0.382 |
| **IED pre-EDS errors** | **n** | **beta** | **LCL** | **UCL** | **p-value** | **n** | **beta** | **LCL** | **UCL** | **p-value** |
| HES1 CpG1 | 185 | -0.031 | -0.100 | 0.037 | 0.372 | 184 | -0.025 | -0.095 | 0.045 | 0.477 |
| HES1 CpG2 | 178 | 0.001 | -0.070 | 0.071 | 0.980 | 177 | 0.004 | -0.067 | 0.075 | 0.912 |
| HES1 CpG3 | 178 | 0.014 | -0.057 | 0.086 | 0.691 | 177 | 0.013 | -0.058 | 0.084 | 0.720 |
| HES1 CpG4 | 177 | -0.028 | -0.099 | 0.043 | 0.439 | 176 | -0.020 | -0.096 | 0.052 | 0.581 |
| HES1 CpG5 | 177 | -0.059 | -0.127 | 0.010 | 0.094\* | 176 | -0.060 | -0.129 | 0.010 | 0.095\* |
| HES1 CpG6 | 198 | -0.019 | -0.085 | 0.048 | 0.583 | 197 | -0.016 | -0.084 | 0.051 | 0.636 |
| HES1 CpG7 | 197 | -0.014 | -0.080 | 0.052 | 0.678 | 196 | -0.014 | -0.081 | 0.053 | 0.685 |
| HES1 CpG8 | 195 | -0.047 | -0.113 | 0.019 | 0.166 | 194 | -0.046 | -0.114 | 0.021 | 0.182 |
| HES1 CpG9 | 195 | -0.030 | -0.097 | 0.037 | 0.381 | 194 | -0.028 | -0.095 | 0.040 | 0.420 |
| **SSP-span length** | **n** | **beta** | **LCL** | **UCL** | **p-value** | **n** | **beta** | **LCL** | **UCL** | **p-value** |
| HES1 CpG1 | 179 | 0.086 | -0.039 | 0.211 | 0.180 | 178 | 0.080 | -0.047 | 0.206 | 0.218 |
| HES1 CpG2 | 172 | 0.078 | -0.046 | 0.202 | 0.221 | 171 | 0.066 | -0.059 | 0.191 | 0.303 |
| HES1 CpG3 | 172 | 0.043 | -0.082 | 0.169 | 0.497 | 171 | 0.035 | -0.090 | 0.159 | 0.584 |
| HES1 CpG4 | 171 | 0.083 | -0.041 | 0.207 | 0.190 | 170 | 0.079 | -0.046 | 0.204 | 0.218 |
| HES1 CpG5 | 171 | 0.168 | 0.048 | 0.287 | 0.007\*\* | 170 | 0.149 | 0.028 | 0.270 | 0.017\*\* |
| HES1 CpG6 | 192 | 0.135 | 0.017 | 0.253 | 0.026\*\* | 191 | 0.124 | 0.005 | 0.242 | 0.042\*\* |
| HES1 CpG7 | 191 | 0.144 | 0.027 | 0.262 | 0.017\*\* | 190 | 0.130 | 0.012 | 0.249 | 0.033\*\* |
| HES1 CpG8 | 189 | 0.111 | -0.008 | 0.230 | 0.069\* | 188 | 0.094 | -0.027 | 0.215 | 0.128 |
| HES1 CpG9 | 189 | 0.103 | -0.017 | 0.222 | 0.095\* | 188 | 0.092 | -0.029 | 0.213 | 0.137 |

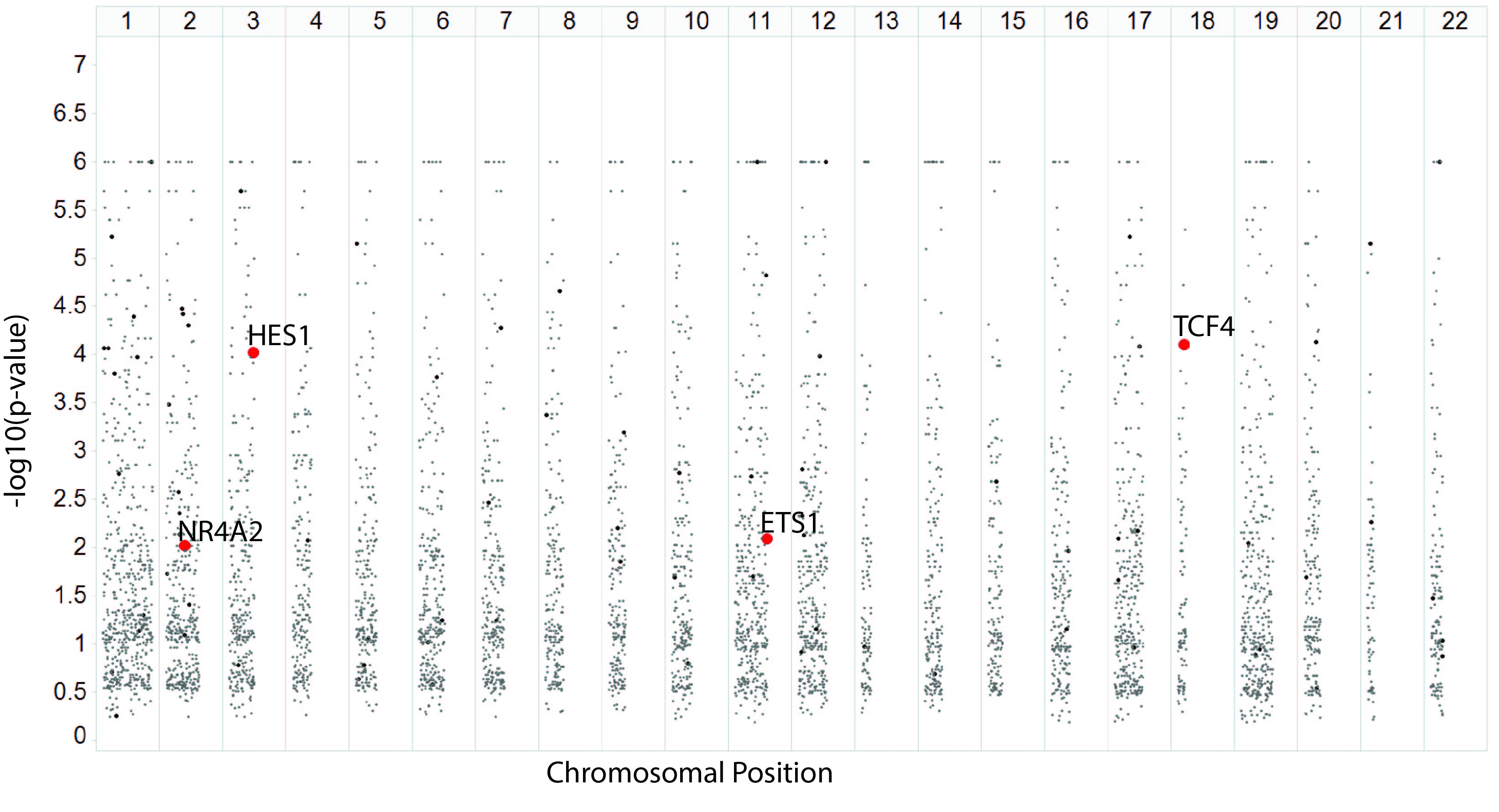
**Table 3:** Association of umbilical cord *HES1* CpG methylation within the identified DMROI with 1 year externalising in 108 children, \*\*p-value≤0.05. LCL = lower 95% confidence limit, UCL = upper 95% confidence limit.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1 year externalising,**  **adjusted for child’s sex** | | | | | **1 year externalising, adjusted for child’s sex and mother’s educational attainment** | | | | |
|  | n | β | LCL | UCL | p-value | n | β | LCL | UCL | p-value |
| ***HES1***  **CpG2** | 108 | 0.064 |  |  | *0.019\*\** | 95 | 0.053 |  |  | 0.050\*\* |
| **CpG5** | 107 | 0.025 |  |  | *0.375* | 94 | 0.025 |  |  | *0.381* |
| **CpG7** | 108 | -0.063 |  |  | *0.031\*\** | 96 | -0.068 |  |  | *0.020\*\** |

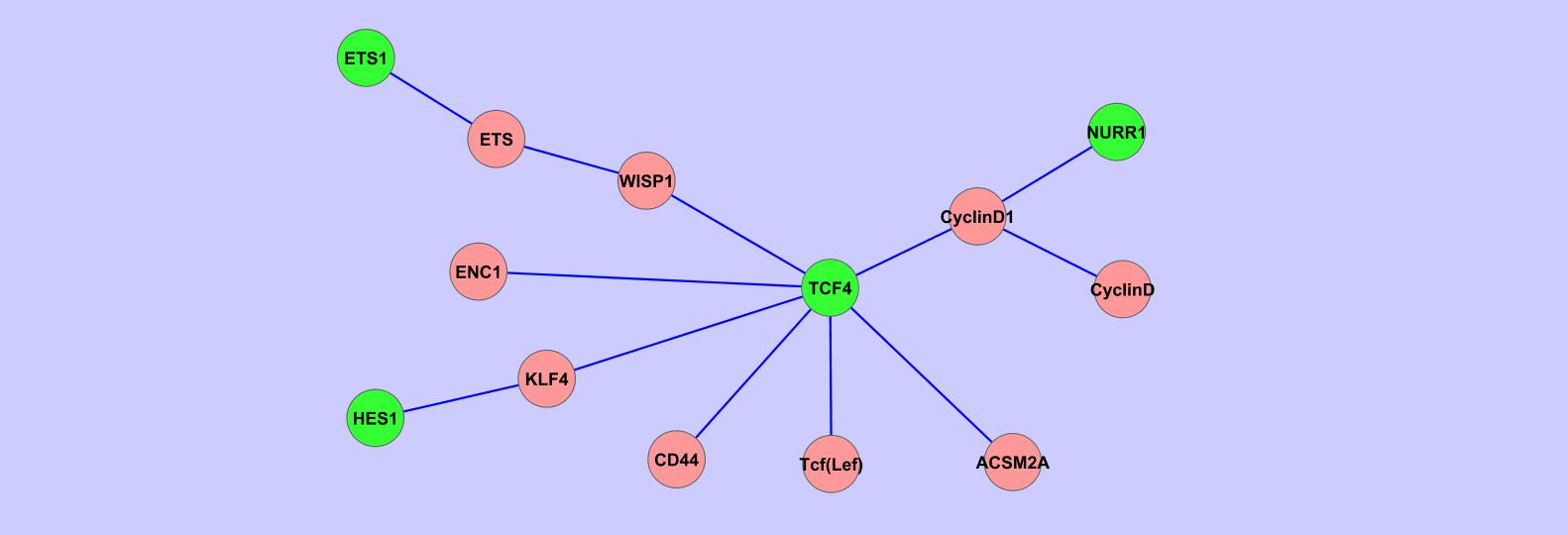
**Figure 1: Overview of Study**

****

**Figure 2: Methylation of HES1 DMR in cord at birth is associated with WPPSI IQ at 4 years of age. a)** Manhattan plot of epigenome wide methylation analysis. The X axis indicates chromosomal position, the Y axis the –log 10 p value of the Fishers Exact test. The black dots represent DMROIs and those associated with *HES1*, *NR4A2*, *ETS1* and *TCF4* are shown in red.  **b)** Diencephalon development pathway. Genes contained in the Diencephalon development GO process were connected to each other by using the direct interactions algorithm in GeneGo MetacoreTM. Genes containing DMROIs (*HES1, NURR1, ETS1* and *TCF4*) are denoted by green circles (*NR4A2* is denoted by its alternative name *NURR1)*. This figure is generated in Cytoscape. **c)** DMROI plot for *HES1* (Chr 3: 193848528 – 193853872). X-axis shows chromosomal coordinates (hg19), y-axis shows absolute %methylation difference between WPSSI groups 1 and 4. Green and red circles represent start and end of each 100 nucleotide region returned from BATMAN, respectively. 100 nucleotide regions in the dotted box were found to have >20% absolute methylation difference between WPSSI groups 1 and 4 and this region was selected for pyrosequencing in the extended sample set. The lower panel shows the positions of the *HES1* transcript and the upstream DMROI **d)** Concordance of methylation values with WPSSI scores for the 100 nt region within the *HES1* ROI selected for pyrosequencing (containing CpGs 2-8), upstream of the *HES1* coding sequence. X-axis shows WPSSI scores and y-axis shows % methylation as estimated by the Bayesian algorithm BATMAN. Sample data points are coloured by WPSSI groups (red = group 1, lowest WPSSI scores; blue = group 2, low WPSSI scores; green = group 3, high WPSSI scores; yellow = group4, highest WPSSI scores). Chromosomal coordinates of the region are detailed above the figure. WPPSI = Wechsler Pre-School and Primary Scale of Intelligence (full-scale IQ)



**a**

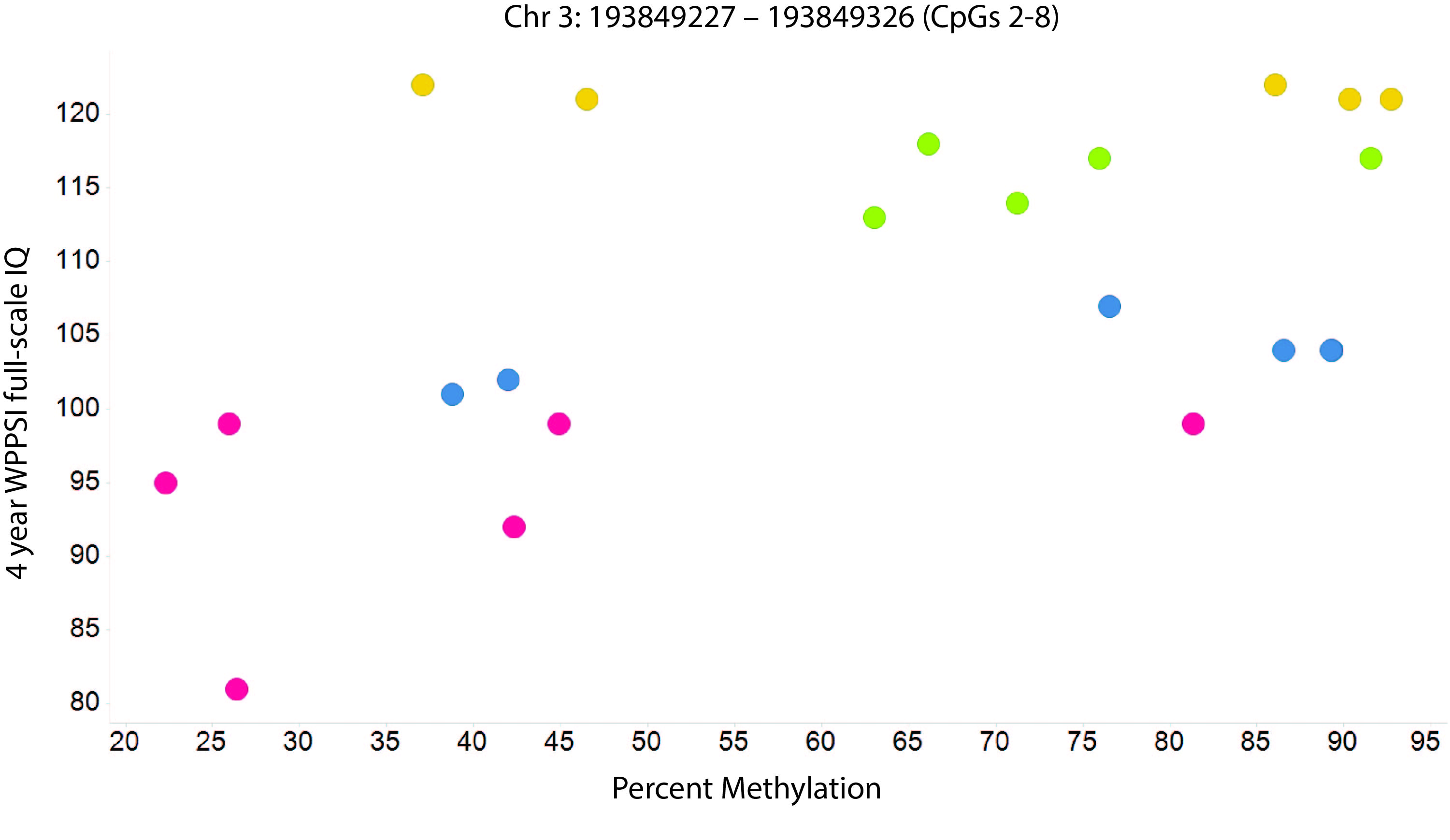


**b**



**c**

**d**



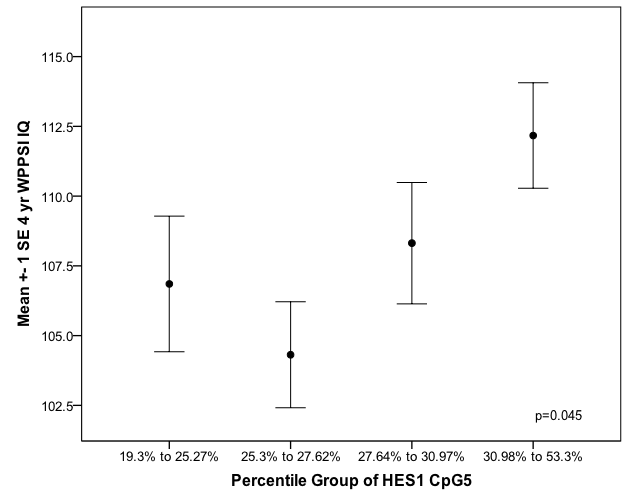
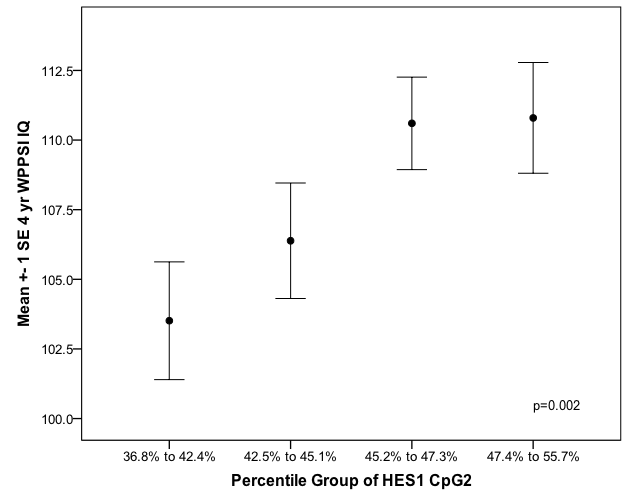
**Figure 3:** *HES1* DMROI methylation at birth is associated with childhood neuropsychological function. **a)** Association between cord *HES1* CpG2 and CpG5 methylation and Wechsler Pre-School and Primary Scale of Intelligence (WPPSI IQ) at age 4 years. **b)** Association between cord *HES1* CpG5 and CpG7 methylation and spatial span length at 7 years of age. **c)** Association between cord *HES1* CpG 5 methylation and delayed matching to sample (DMS) 12s delay total correct at 7 years of age. **d)** Association between cord *HES1* CpG7 methylation and infant externalising score. Methylation has been divided into 4 equal groups according to rank; median, interquartile range and range are plotted for each group. Methylation values greater than 1.5 times the interquartile range from either the upper or lower quartile for each group are shown as circles. P values are for regression of continuous variables adjusting for gender and mother’s IQ.

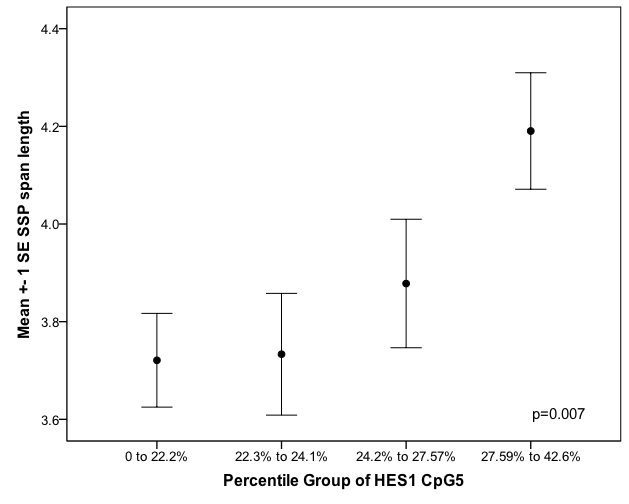
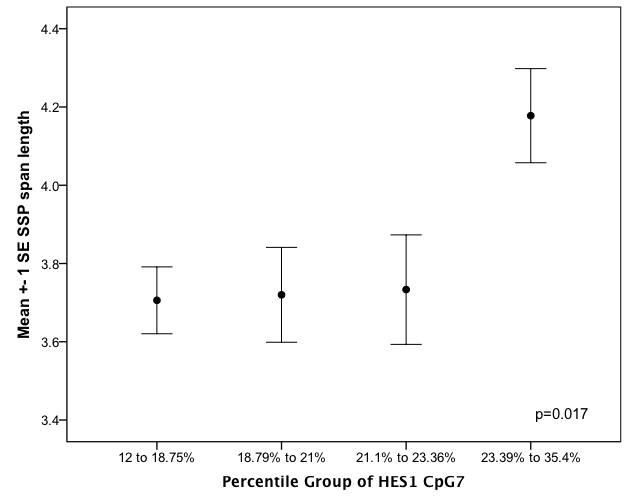
**d) CpG associations and GUSTO 1 year externalising**

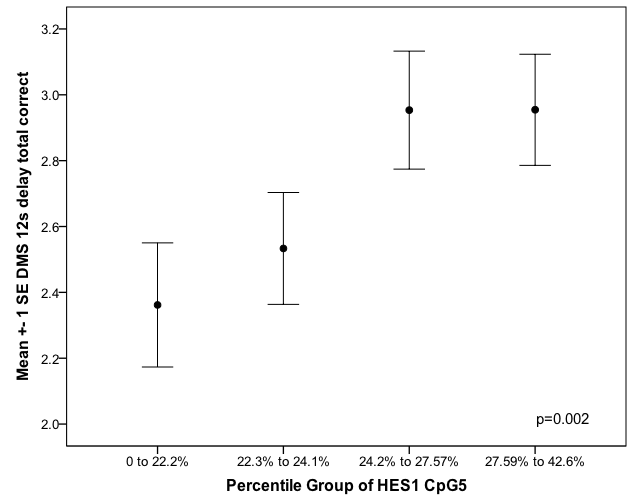
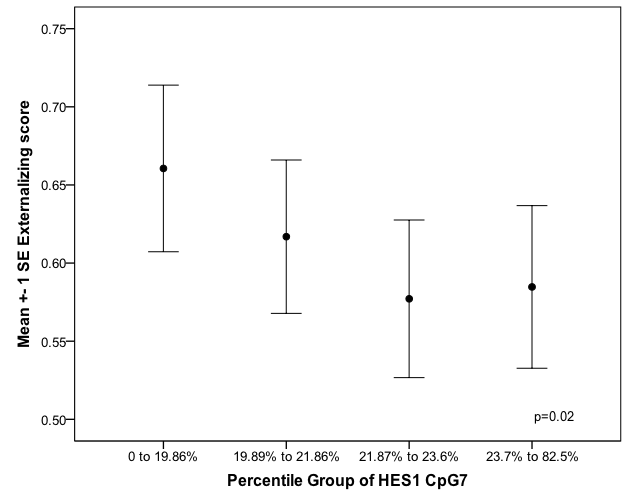
**a) CpG associations and SWS 4 year WPPSI**

**b) CpG associations and SWS 7 year SSP span length**

**c) CpG associations and SWS 7 year DMS total correct (12s)**

****

****

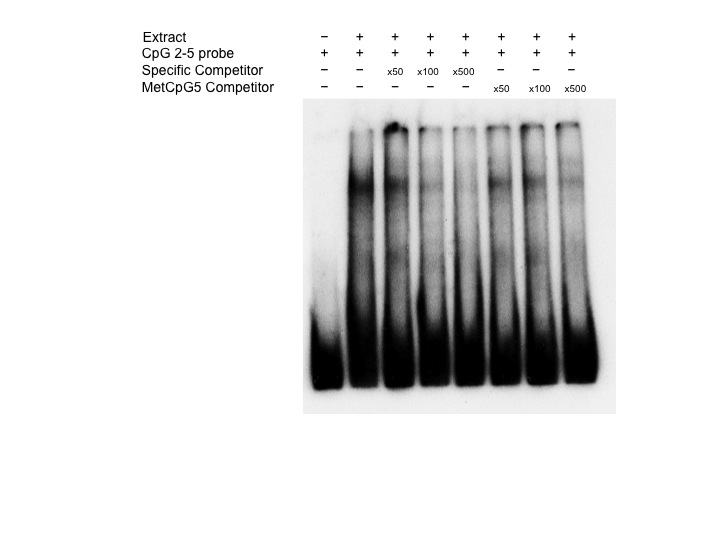
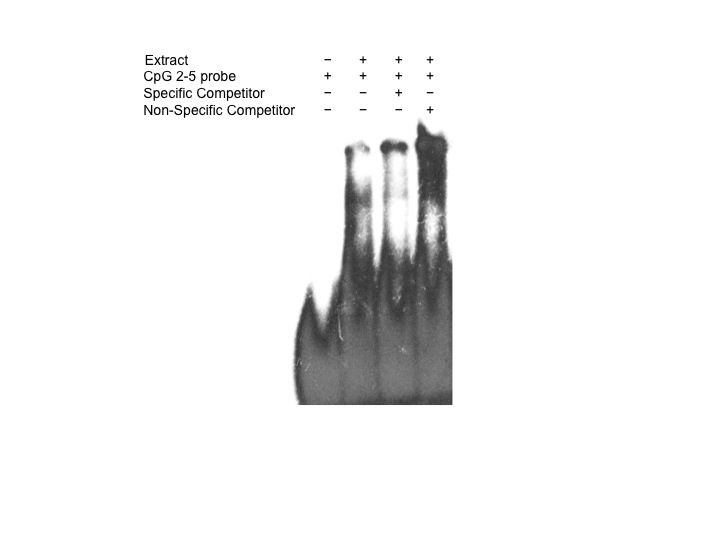
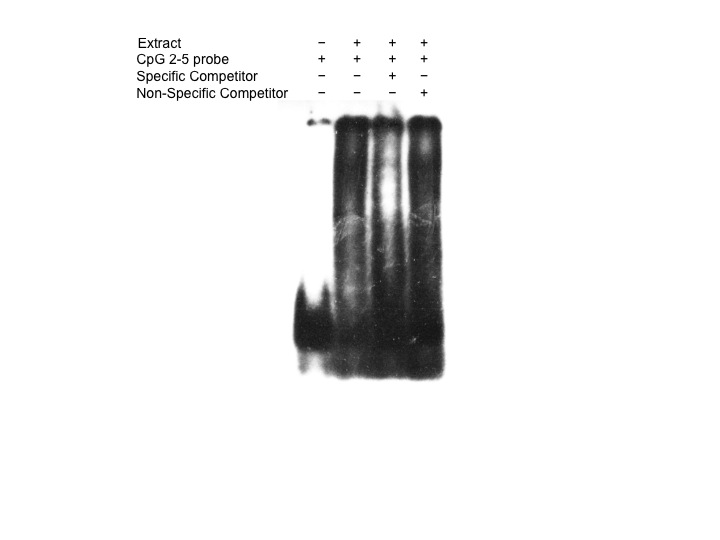
****

**Figure 4:** Methylation ofCpG5 blocks ETS transcription factor binding to the *HES1* promoter sequence. Results are typical of three analyses. **(a)** The unmethylated and methylated probes showed a strong shift upon incubation with the extract; this shift was markedly reduced by co-incubating with 500-fold excess of the unlabelled specific competitor, but not with 500-fold excess of an unlabelled non-specific competitor. **(b)** Binding to the methylated labelled probe was markedly diminished by co-incubation with 100-fold excess of an unlabelled oligonucleotide containing the core consensus sequence for ETS (GGAA). **(c)** The unmethylated biotin labelled probe was incubated with nuclear extracts from the human neuroblastoma cell line IMR32 with a 50, 100, and 500-fold excesses of the unmethylated or methylated competitor; binding to the unmethylated probe was competed out with a 50-fold excess of the methylated competitor compared to a 500-fold excess of the unmethylated competitor.

**a**

**b**

**c**



|  |  |  |
| --- | --- | --- |
| **Characteristic** | **% or Median (5th,95th percentile) for four-year SWS cohort** | **% or Median (5th, 95th percentile) for seven-year SWS cohort** |
|  |  |  |
| **Mother** |  |  |
| Full Scale IQ (WASI) | 108 (90 to 126) |  |
| Educational qualifications (%) |  |  |
| None | 1.1% | 1.0% |
| CSE | 10.8% | 10.8% |
| O levels | 24.4% | 27.6% |
| A levels | 30.8% | 35.0% |
| HND | 8.5% | 5.4% |
| Degree | 24.4% | 20.2% |
| Social class (%) |  |  |
| Professional | 5.1% | 5.5% |
| Management and technical | 41.8% | 37.2% |
| Skilled non-manual | 34.9% | 34.7% |
| Skilled manual | 8.0% | 6.0% |
| Partly skilled | 9.1% | 14.6% |
| Unskilled | 1.1% | 2.0% |
| Primiparous | 49.4% | 53.4% |
| Age at birth, years | 30.4 (24.5 to 35.6) | 32.2 (25.5 to 36.9) |
| Smoker | 25.0% | 34.4% |
| BMI | 24.5 (20.6 to 33.8) | 24.3 (19.2 to 33.4 ) |
|  |  |  |
| **Child** |  |  |
| Age at follow-up, years | 4.4 (4.2 to 4.6) | 7.0 (6.8 to 7.2) |
| Female (%) | 44.3% | 51.5% |
| Birth order (%) |  |  |
| 1st | 49.4% | 53.4% |
| 2nd | 36.9% | 33.8% |
| 3rd or higher | 13.7% | 12.8% |
| Birth weight, kg | 3.5 (2.7 to 4.4) | 3.4 (2.6 to 4.3) |
| Gestational age, weeks | 40.0 (36.7 to 41.9) | 40.1 (36.9 to 42) |
| Full Scale IQ (WPPSI) | 110 (85 to 127) |  |
| CANTAB assessments |  |  |
| DMS total correct (12s) |  | 3 (1 to 5) |
| IED total errors (stage 1) |  | 2 (0 to 4) |
| IED total errors (stage 8)  IED pre-EDS errors  SSP span length |  | 23 (2 to 31)  8.5 (5 to 23)  4 (2 to 5) |

**Supplementary Table 1: Characteristics of the SWS study participants with four-year (n=175) and seven-year (n=200) cognitive and neuropsychological assessment.** Median birth weight (3.4 kg), maternal age (30.8 years) and pre-pregnancy body mass index (24.1 kg/m2) were similar in the whole SWS cohort (n=3,159). WPPSI = Wechsler Pre-School and Primary Scale of Intelligence (full-scale IQ), WASI = Wechsler Abbreviated Scale of Intelligence (full-scale IQ), CANTAB = Cambridge Neuropsychological Test Automated Battery.

|  |  |
| --- | --- |
| **Characteristic** | **% or Median (5th, 95th percentile) for one-year GUSTO cohort** |
|  |  |
| **Mother** |  |
| Maternal Highest Education % |  |
| None/Primary | 4.6% |
| Secondary/Technical Education | 22.7% |
| GCE A level/Polytechnic  University/Others | 41.8%  30.9% |
| Ethnicity % |  |
| Chinese | 58.9% |
| Malays | 25.0% |
| Indians | 16.1% |
| Household Monthly Income (SGD) % |  |
| 0-1999 | 15.4% |
| 2000-5999 | 59.1% |
| >6000 | 25.5% |
| Primiparous | 51.6% |
| Age at birth, years | 31.7 (22.7 to 39.4) |
| Smoker | 13.7% |
| BMI at 26 weeks | 25.1 (20.5 to 33.2) |
|  |  |
| **Child** |  |
| Female % | 50.0% |
| Birth order % |  |
| 1st | 51.6% |
| 2nd | 30.7% |
| 3rd or higher | 17.7% |
| Birth weight, kg | 3.09 (2.40 to 3.85) |
| Gestational age, weeks | 38.7 (37.0 to 40.4) |
| Age, years | 0.99 ( 0.92 to 1.12) |
|  |  |
| ITSEA Externalising score | 0.59 (0.13 to 1.04) |
|  |  |
|  |  |

**Supplementary Table 2:** **Characteristics of the 124 GUSTO study participants with one-year neuropsychological assesment.** Median birth weight (3.08 kg), maternal age (30.6 years) and pre-pregnancy body mass index (26.1 kg/m2) were also similar in the whole GUSTO cohort (n=1162). ITSEA = Infant Toddler Socio-Emotional Assessment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Primers for bisulphite pyrosequencing** | | | | | |
| **Gene** | **Primer** | **Sequence (5’-3’)** | **Genomic co-ordinates (UCSC, hg19, Feb 2009 assembly)** | **Amplicon length (bp)** | **No. CpGs** |
| *HES1* | Forward  Reverse †  Sequencing 1  Sequencing 2 | AGGGGATAAAGGGGAGTT  TCACTTCTTTAATCCCCCTATAACACCA  GGTTTGAAAGTAAATAGGT  TTGTGGGTGGAGATAA | Chr3:193849141-193849361 + | 221 | 9 |
| **Primers for EMSAs** | | | | | |
|  | **Primer** | **Sequence (5’-3’)** |  | | |
| *HES1* | HES1 CpG2-5 | AGTCGCCCTTCCGGGGCGGGGGTGGGGGGACGCTG |
| HES1 CpG2-5 methylated2 | AGT*[5MedC]*GCCCTTCCGGGGCGGGGGTGGGGGGACGCTG |
| HES1 CpG2-5 methylated5 | AGTCGCCCTTCCGGGGCGGGGGTGGGGGGA*[5MedC]*GCTG |
| *ETS* | Consensus sequence  Mutated consensus | GGGCTGCTTGAGGAAGTATAAGAAT  GGGCTGCTTGAAAAAGTATAAGAAT |

**Supplementary Table 3:** **Primers for bisulphite pyrosequencing and electrophoretic mobility shift assays**. † denotes biotinylated primer.

**Supplementary Table 4:** **Genes containing differentially methylated regions of interests (DMROIs) identified from the MBD array using Fisher Exact tests, sorted by Fisher Exact test p-value.**

|  |  |  |
| --- | --- | --- |
| **Chr** | **Gene symbol** | **Fisher Exact**  **p-value** |
| *22* | *FAM83F* | 1.00E-06 |
| *1* | *OR6F1* | 1.00E-06 |
| *11* | *SERPINH1* | 1.00E-06 |
| *12* | *BRI3BP* | 1.00E-06 |
| *10* | *FANK1* | 2.00E-06 |
| *17* | *MPP3* | 6.00E-06 |
| *5* | *CMBL* | 7.00E-06 |
| *21* | *C21orf49* | 7.00E-06 |
| *11* | *C11orf61* | 1.50E-05 |
| *8* | *TMEM65* | 2.20E-05 |
| *2* | *IL1RN* | 3.30E-05 |
| *2* | *SLC35F5* | 3.80E-05 |
| *1* | *OR6N2* | 4.00E-05 |
| *2* | *NFE2L2* | 5.00E-05 |
| *7* | *TTC26* | 5.30E-05 |
| *20* | *EYA2* | 7.40E-05 |
| *18* | *TCF4* | 7.80E-05 |
| *3* | *HES1* | 9.60E-05 |
| *12* | *GAS2L3* | 0.000104 |
| *1* | *ZMPSTE24* | 0.000158 |
| *2* | *CIB4* | 0.000325 |
| *8* | *LONRF1* | 0.000422 |

|  |  |  |
| --- | --- | --- |
| **Chr** | **Gene symbol** | **Fisher Exact**  **p-value** |
| *12* | *KLRC4* | 0.001545 |
| *11* | *RTN3* | 0.001821 |
| *10* | *FGFBP3* | 0.00213 |
| *2* | *RNF103* | 0.002676 |
| *7* | *INHBA-AS1* | 0.003401 |
| *2* | *CIAO1* | 0.004382 |
| *12* | *CLEC12A* | 0.004732 |
| *21* | *SIM2* | 0.005469 |
| *10* | *ACADSB* | 0.005942 |
| *9* | *C9orf5* | 0.006332 |
| *17* | *ABCA5* | 0.006665 |
| *2* | *FER1L5* | 0.007339 |
| *12* | *DERA* | 0.007395 |
| *17* | *FGF11* | 0.008121 |
| *11* | *ETS1* | 0.008175 |
| *4* | *GAB1* | 0.008434 |
| *19* | *CNN1* | 0.009007 |
| *2* | *NR4A2* | 0.00962 |
| *11* | *OR1S1* | \* |
|  |  |  |
|  |  |  |
|  |  |  |

\*The *OR1S1* region contained on the array included just one 100nt region so it was not possible to conduct a fisher’s exact test. However, that 1 region passed the criteria for a DMR

**Supplementary Table 5: Gene Ontology Processes enrichment analysis of the DMROIs (ontologies with p<0.0005).**

|  |  |  |  |
| --- | --- | --- | --- |
| **GO Process** | **p value** | **Ratio (pathway genes including a DMROI / all genes in pathway** | **DMROIs in each ontology** |
| Diencephalon development | 4.430E-05 | 4/71 | *NR4A2, HES1, TCF4, ETS1* |
| Pituitary gland development | 1.88E-04 | 3/40 | *TCF4, HES1, ETS1* |
| Negative regulation of glial cell proliferation | 2.153E-04 | 2/8 | *HES, RNF10* |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Beta** | **Std. Err** | **t-value** | **p-value** | **LCL** | **UCL** |
| **HES1 CpG2** | 2.954 | 1.048 | 2.82 | 0.005 | 0.883 | 5.026 |
| **Sex** | 4.201 | 2.048 | 2.05 | 0.042 | 0.153 | 8.248 |
| **Mother’s WASI** | 0.241 | 0.090 | 2.68 | 0.008 | 0.063 | 0.419 |
| **Parity** | -4.228 | 1.988 | -2.13 | 0.035 | -8.157 | -0.298 |
| **Current Smoking** | 0.567 | 2.350 | 0.24 | 0.810 | -4.077 | 5.211 |
| **Birthweight** | 0.003 | 0.002 | 1.36 | 0.175 | -0.001 | 0.007 |
| **Age at test** | -3.834 | 7.857 | -0.49 | 0.626 | -19.362 | 11.693 |
| **Mother’s BMI** | -0.189 | 0.240 | -0.79 | 0.432 | -0.663 | 0.285 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Beta** | **Std. Err** | **t-value** | **p-value** | **LCL** | **UCL** |
| **HES1 CpG5** | 1.766 | 1.108 | 1.59 | 0.113 | -0.426 | 3.957 |
| **Sex** | 5.976 | 2.160 | 2.73 | 0.007 | 1.643 | 10.310 |
| **Mother’s WASI** | 0.251 | 0.100 | 2.51 | 0.013 | 0.053 | 0.448 |
| **Parity** | -3.474 | 2.161 | -1.59 | 0.115 | -7.809 | 0.861 |
| **Current Smoking** | 1.411 | 2.523 | 0.56 | 0.577 | -3.580 | 6.402 |
| **Birthweight** | 0.002 | 0.002 | 0.92 | 0.360 | -0.003 | 0.007 |
| **Age at test** | -3.256 | 8.524 | -0.38 | 0.703 | -20.123 | 13.611 |
| **Mother’s BMI** | -0.141 | 0.257 | -0.55 | 0.586 | -0.650 | 0.369 |

**Supplementary Table 6: Adjustment covariates for HES1 CpG2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **SWS subjects at four- years** | **SWS subjects at**  **seven- years** | **GUSTO subjects** |
|  | **Genomic co-ordinates (hg19)**  **and distance from TSS** | Methylation %, median  (5th, 95th percentile) | Methylation %, median  (5th, 95th percentile) | Methylation %, median  (5th, 95th percentile) |
| ***HES1***  **CpG1** | chr3:193849210+ (-4721) | 47.55 (41.4, 53.76) | 45.99 (40.61,52.66) | 47.33 (42.42, 52.04) |
| **CpG2** | chr3:193849227+ (-4704) | 45.16 (39.76, 50.75) | 43.04 (35.76,48.41) | 44.66 (39.60, 48.99) |
| **CpG3** | chr3:193849235+ (-4696) | 35.43 (30.74, 43.23) | 34.29 (29.37,41.57) | 33.70 (29.03, 37.61) |
| **CpG4** | chr3:193849240+ (-4691) | 42.97 (37.64, 48.33) | 42.16 (35.85,50.13) | 42.39 (37.03, 46.98) |
| **CpG5** | chr3:193849254+ (-4679) | 27.64 (21.93, 35.77) | 24.16 (19.05,34.77) | 23.35 (19.34, 28.11) |
| **CpG6** | chr3:193849275+ (-4656) | 32.18 (26.21, 42.12) | 31.20 (26.05,42.16) | 29.04 (24.96, 39.10) |
| **CpG7** | chr3:193849309+ (-4622) | 24.51 (19.13, 31.83) | 20.77 (16.15,28.30) | 21.89 (18.20, 30.24) |
| **CpG8** | chr3:193849318+ (-4613) | 23.02 (17.54, 32.12) | 19.42 (14.71,30.57) | 19.60 (15.62, 29.52) |
| **CpG9** | chr3:193849328+ (-4603) | 32.60 (27.08, 41.59) | 29.12 (23.23,38.74) | 30.69 (26.27, 40.07) |

**Supplementary Table 7**: **Umbilical cord methylation range within the DMROI of *HES1* as measured by bisulphite pyrosequencing in the SWS and GUSTO cohort.**

**Supplementary Table 8**: **Correlation of the differentially methylated HES1 CpGs at age 4- and 7-years in the SWS cohort**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Four-year cohort** | | | | | | | | | | |
|  |  | **CpG1** | **CpG2** | **CpG3** | **CpG4** | **CpG5** | **CpG6** | **CpG7** | **CpG8** | **CpG9** |
| **CpG1** | **Correlation Coefficient** | 1.000 | .634\*\* | .538\*\* | .510\*\* | .412\*\* | .446\*\* | .666\*\* | .568\*\* | .637\*\* |
|  | **Sig (2-tailed)** |  | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 168 | 157 | 154 | 146 | 139 | 163 | 148 | 140 | 131 |
| **CpG2** | **Correlation Coefficient** | .634\*\* | 1.000 | .611\*\* | .573\*\* | .578\*\* | .589\*\* | .659\*\* | .605\*\* | .669\*\* |
|  | **Sig (2-tailed)** | .000 |  | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 157 | 157 | 154 | 146 | 139 | 152 | 137 | 129 | 122 |
| **CpG3** | **Correlation Coefficient** | .538\*\* | .611\*\* | 1.000 | .467\*\* | .571\*\* | .457\*\* | .498\*\* | .503\*\* | .623\*\* |
|  | **Sig (2-tailed)** | .000 | .000 |  | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 154 | 154 | 154 | 146 | 139 | 149 | 134 | 126 | 120 |
| **CpG4** | **Correlation Coefficient** | .510\*\* | .573\*\* | .467\*\* | 1.000 | .636\*\* | .518\*\* | .644\*\* | .578\*\* | .595\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 |  | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 146 | 146 | 146 | 146 | 137 | 141 | 127 | 120 | 115 |
| **CpG5** | **Correlation Coefficient** | .412\*\* | .578\*\* | .571\*\* | .636\*\* | 1.000 | .691\*\* | .554\*\* | .619\*\* | .647\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 |  | .000 | .000 | .000 | .000 |
|  | **n** | 139 | 139 | 139 | 137 | 139 | 135 | 121 | 113 | 109 |
| **CpG6** | **Correlation Coefficient** | .446\*\* | .589\*\* | .457\*\* | .518\*\* | .691\*\* | 1.000 | .546\*\* | .560\*\* | .582\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 |  | .000 | .000 | .000 |
|  | **n** | 163 | 152 | 149 | 141 | 135 | 170 | 155 | 146 | 138 |
| **CpG7** | **Correlation Coefficient** | .666\*\* | .659\*\* | .498\*\* | .644\*\* | .554\*\* | .546\*\* | 1.000 | .850\*\* | .798\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 |  | .000 | .000 |
|  | **n** | 148 | 137 | 134 | 127 | 121 | 155 | 155 | 146 | 138 |
| **CpG8** | **Correlation Coefficient** | .568\*\* | .605\*\* | .503\*\* | .578\*\* | .619\*\* | .560\*\* | .850\*\* | 1.000 | .808\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 | .000 |  | .000 |
|  | **n** | 140 | 129 | 126 | 120 | 113 | 146 | 146 | 146 | 136 |
| **CPG9** | **Correlation Coefficient** | .637\*\* | .669\*\* | .623\*\* | .595\*\* | .647\*\* | .582\*\* | .798\*\* | .808\*\* | 1.000 |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |  |
|  | **n** | 131 | 122 | 120 | 115 | 109 | 138 | 138 | 136 | 138 |
| **Seven-year cohort** | | | | | | | | | | |
|  |  | **CpG1** | **CpG2** | **CpG3** | **CpG4** | **CpG5** | **CpG6** | **CpG7** | **CpG8** | **CpG9** |
| **CpG1** | **Correlation Coefficient** | 1.000 | .636\*\* | .592\*\* | .510\*\* | .490\*\* | .538\*\* | .498\*\* | .420\*\* | .512\*\* |
|  | **Sig (2-tailed)** |  | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 218 | 210 | 210 | 209 | 209 | 218 | 216 | 215 | 215 |
| **CpG2** | **Correlation Coefficient** | .636\*\* | 1.000 | .540\*\* | .488\*\* | .479\*\* | .487\*\* | .455\*\* | .410\*\* | .450\*\* |
|  | **Sig (2-tailed)** | .000 |  | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 210 | 210 | 210 | 209 | 209 | 210 | 208 | 207 | 207 |
| **CpG3** | **Correlation Coefficient** | .592\*\* | .540\*\* | 1.000 | .307\*\* | .451\*\* | .497\*\* | .471\*\* | .468\*\* | .581\*\* |
|  | **Sig (2-tailed)** | .000 | .000 |  | .000 | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 210 | 210 | 210 | 209 | 209 | 210 | 208 | 207 | 207 |
| **CpG4** | **Correlation Coefficient** | .510\*\* | .488\*\* | .307\*\* | 1.000 | .576\*\* | .510\*\* | .455\*\* | .476\*\* | .437\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 |  | .000 | .000 | .000 | .000 | .000 |
|  | **n** | 209 | 209 | 209 | 209 | 209 | 209 | 207 | 206 | 206 |
| **CpG5** | **Correlation Coefficient** | .490\*\* | .479\*\* | .451\*\* | .576\*\* | 1.000 | .730\*\* | .697\*\* | .697\*\* | .630\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 |  | .000 | .000 | .000 | .000 |
|  | **n** | 209 | 209 | 209 | 209 | 209 | 209 | 207 | 206 | 206 |
| **CpG6** | **Correlation Coefficient** | .538\*\* | .487\*\* | .497\*\* | .510\*\* | .730\*\* | 1.000 | .735\*\* | .733\*\* | .747\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 |  | .000 | .000 | .000 |
|  | **n** | 218 | 210 | 210 | 209 | 209 | 231 | 229 | 227 | 227 |
| **CpG7** | **Correlation Coefficient** | .498\*\* | .455\*\* | .471\*\* | .455\*\* | .697\*\* | .735\*\* | 1.000 | .811\*\* | .828\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 |  | .000 | .000 |
|  | **n** | 216 | 208 | 208 | 207 | 207 | 229 | 229 | 227 | 227 |
| **CpG8** | **Correlation Coefficient** | .420\*\* | .410\*\* | .468\*\* | .476\*\* | .697\*\* | .733\*\* | .811\*\* | 1.000 | .830\*\* |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 | .000 |  | .000 |
|  | **n** | 215 | 207 | 207 | 206 | 206 | 227 | 227 | 227 | 227 |
| **CPG9** | **Correlation Coefficient** | .512\*\* | .450\*\* | .581\*\* | .437\*\* | .630\*\* | .747\*\* | .828\*\* | .830\*\* | 1.000 |
|  | **Sig (2-tailed)** | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |  |
|  | **n** | 215 | 207 | 207 | 206 | 206 | 227 | 227 | 227 | 227 |

\*\*. Correlation is significant at the 0.01 level (2-tailed).

**Supplementary Methods 1: The SWS cohort and neurological assesment**

***Southampton Women’s Survey: participants***

The Southampton Women’s Survey (SWS) is a prospective mother offspring cohort study that has assessed the diet, body composition, physical activity and social circumstances of a large group of non-pregnant women aged 20 to 34 years living in the city of Southampton, UK. Women were recruited through General Practices across the city between April 1998 and December 2002. Each woman was invited to take part by letter, followed by a telephone call when an interview date was arranged; 12,583 women agreed to take part, 75% of all women contacted. Trained research nurses visited the women at home and collected information about their health, diet and lifestyles, as well as taking anthropometric measurements. Women who subsequently became pregnant were followed up at 11, 19 and 34 weeks gestation and their offspring were studied in infancy and childhood. Details of mothers’ parity, educational attainment (defined in six groups according to highest academic qualification) and social class were obtained at the pre-pregnancy interview, and height and weight were measured. Amongst women who became pregnant, smoking status in pregnancy was ascertained at the 11 and 34 week interviews. A total of 1981 women became pregnant and delivered a live-born singleton infant before the end of 2003. Six infants died in the neonatal period and two had major congenital growth abnormalities, which left 1973 mother-offspring pairs. Follow-up of the children and sample collection/analysis was carried out under Institutional Review Board approval (Southampton and SW Hampshire Research Ethics Committee) with written informed consent. Investigations were conducted according to the principles expressed in the Declaration of Helsinki.

*SWS cognitive and neuropsychological assessment at seven-years of age*

The Cambridge Neuropsychological Test Automated Battery (CANTAB®) is designed to test specific components of executive function. To reduce the likelihood of chance findings we focused on five CANTAB® outcomes with the strongest track record of associations with executive function in the published literature (Delayed Matching to Sample (DMS) 12 s delay total correct, Intra-Extra Dimensional Set Shift (IED) Stage 1 errors, IED Pre- extra-dimensional shift (EDS) errors, IED Stage 8 errors, Spatial Span (SSP) span length). DMS assesses forced choice recognition memory for novel non-verbalisable patterns, and tests both simultaneous and short term visual memory; 12 s delay DMS total correct ranged between 0 and 5, with a median of 3. IED involves a total of nine stages and assesses the ability to engage in deliberate, goal-directed thought and action. The number of errors committed on stage 1 indicates proficiency in detecting and learning the implicit rule of the task based on feedback from the experimenter as to whether the choice was correct. The total number of errors from Stages 1 to 7 is referred to as Pre-EDS errors and indicates proficiency in maintaining selective attention. Successful completion of stages indicates ability to maintain attention and the flexibly to shift in response to the demands of the task. At Stage 8 (the EDS stage) participants must learn to shift attention from the previously correct dimension (the shape of the stimulus) to the newly correct dimension (the line); the number of errors at this stage indicates proficiency in extra-dimensional set-shifting. Two children completed only stage 1 of the task and the data from these participants were removed from the IED analyses. Stage 1 errors ranged from 0 to 23; for statistical analysis the data were analysed in 5 groups. Stage 1-7 errors were positively skewed and were log-transformed for statistical analysis. Stage 8 errors ranged from 0 to 35; for statistical analysis the data was analysed in five groups. SSP assesses the working memory capacity aspect of executive function. White squares are shown, some of which briefly change colour in a variable sequence. The participant must then touch the boxes which changed colour in the same order that they were displayed by the computer (for clinical mode) or in the reverse order (for reverse mode). The number of boxes (and level of difficulty) increases from two at the start of the test to nine at the end, and the sequence and colour are varied through the test. We used the span length (the longest sequence of the pattern the participant is able to follow) as a measure of working memory capacity, which is shorter in children with bipolar disorder. Spatial span length ranged between two and six, with a median of four. Direct assessments of the mother’s cognitive function were not available for analyses relating to this group of seven-year old children; we therefore controlled for the mother’s level of educational attainment as a principal potential confounding factor for child’s cognitive function.

**Supplementary Methods 2: The GUSTO cohort and neurological assesment**

***Growing Up in Singapore Towards Healthy Outcomes (GUSTO): participants***

Mothers were recruited from the KK Women’s and Children’s Hospital and the National University Hospital in Singapore. 3751 families were screened and 2034 met eligibility criteria. Ineligibility was principally accounted for by an intention to deliver outside the 2 study hospitals or not to remain in Singapore for the next 5 years, booking beyond the first trimester, or non-homogenous parental ethnic background. Of the 1247 women (response rate 61.3%) recruited, 1162 conceived naturally, while 85 conceived through in vitro fertilisation (IVF). At baseline, 55.9% were Chinese, 26.1% Malay and 18% Indian. Mean maternal age at recruitment was 30.6 years (range: 18 to 46). Gestational age was defined from a dating ultrasound (10–12 weeks) followed by an additional scan at 18–22 weeks. The average birth weight in the GUSTO cohort was 3081 g, which is comparable to the average across a larger Singaporean sample of 3183 g for a term infant (unpublished data). Written parental consent to participate in the study was given and hard copies are stored by the GUSTO data team. Ethical approval for the study and the consent forms and contents was granted, by the ethics boards of both KKH and NUH, which are centralised Institute Review Board (CIRB) and Domain Specific Review Board (DSRB), respectively.

*GUSTO neuropsychological assessment at one year of age*

The ITSEA was administered to mothers via questionnaire format. The ITSEA detects social-emotional and behaviour problems and delays in the acquisition of competencies in infants and toddlers. It is designed to be applicable to a wide range of parents including those with limited education and from different cultural backgrounds. The Externalising domain considers early manifestations of socially disruptive behaviours such as aggression and defiance. Here, we include ITSEA data from the 124 GUSTO children for whom umbilical cord DNA was also available.

**Supplementary Methods 3. Whole Genome methylation analysis**

After methyl capture, the labelled methylated DNA and input DNA was hybridised to the Agilent Human Promoter Whole-Genome ChIP-on-chip array (G4489A).This contains probes which are split across 2 plates, one accommodating chromosomes 1 to 10 and the other accommodating chromosomes 10/11 to 22 together with X and Y. Microarray hybridisation of the methylated DNA and input DNA (sonicated, total DNA) was carried out by Oxford Gene Technology (OGT, Oxford UK) in accordance with the company’s quality control procedures using standard protocols for labelling, hybridisation and washing. Microarray slides were scanned at 5μM resolution using the extended dynamic range (high 100%, low 10%). The slides were then feature extracted using Agilent Feature Extraction Software (v.9.5.3.1). All arrays were normalised per spot and per chip by an intensity dependent normalisation (Lowess normalisation) using Agilent Genespring Software.

A minority of samples from each plate (most of which were technical replicates), deviated from the expected beta distribution and were discarded. 22 samples for plate one and 21 samples for plate two, remained. All the remaining samples had median absolute deviation (MAD) scores above -5 and did not cluster into strongly defined separate groups in unsupervised hierarchical clustering or principal component analysis.