7-21-2017

Epigenetic Marker Identification and Assessment of Methods on Cell Type Compositions at the Epigenome-Scale

Akhilesh Kaushal

EPIGENETIC MARKER IDENTIFICATION AND ASSESSMENT OF METHODS ON CELL TYPE COMPOSITIONS AT THE EPIGENOME-SCALE

by

Akhilesh Kaushal

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Epidemiology

The University of Memphis

August 2017

Preface

Chapter 2 of the this dissertation has been published as **Kaushal, A.**, Zhang, H., Karmaus, W. J., Everson, T. M., Marsit, C. J., Karagas, M. R., ... & Wang, S. L. (2017). **Genome-wide DNA methylation at birth in relation to in utero arsenic exposure and the associated health in later life.** *Environmental Health*, *16*(1), 50. I performed all the statistical analyses and drafted the manuscript along-with Zhang, H. Wang, S.L. conceived the study and collected all the data, Karmaus, W.J., provided guidance on epigenome and clinical aspects and T. M., Marsit, C. J., Karagas, M. R performed the replication study.

Chapter 4 of this dissertation has been published as **Kaushal, A.**, Zhang, H., Karmaus, W. J., Ray, M., Torres, M. A., Smith, A. K., & Wang, S. L. (2017). **Comparison of different cell type correction methods for genome-scale epigenetics studies.** *BMC bioinformatics*, *18*(1), 216. I performed all the analyses and drafted the manuscript along with Zhang, H. Karmaus, W. J. motivated the analyses and contributed to the manuscript, Ray, M. provided code for simulation scenario one and edited the manuscript . All authors were involved in editing and revising the manuscript.

Acknowledgements

I would like to express my deep appreciation and gratitude to my advisor, Dr. Hongmei Zhang, for providing me with the opportunity to perform doctoral research in her supervision. My graduate training under her guidance has been a great learning experience for me. She always encouraged me to develop independent thinking and research skills. Her guidance, mentorship, and support throughout this research journey have been invaluable.

I would like to express sincere thanks to my dissertation committee members, Drs. Wilfried JJ Karmaus and Shu-Li Wang for their valuable suggestions and their precious time.

I am also thankful to other members in my committee: Drs. Prateek Banerjee, and Meredith Ray for all their support and help from the start to the end of my graduate study journey in the School of Public health at University of Memphis.

I acknowledge all the funding sources that have sponsored my graduate studies (NSC98-2314-B-400-001-MY3, MOST103-2314-B-400-006, R01 AI091905 (PI: Wilfried Karmaus), R21AI099367 (PI: Hongmei Zhang), and 1R01AI121226 (PIs: Hongmei Zhang, John Holloway).

I am forever indebted to my parents for everything.

Abstract

Kaushal, Akhilesh. Ph.D. The University of Memphis. August 2017. Epigenetic Marker Identification and Assessment of Methods on Cell Type Compositions at the Epigenome-Scale. Major Professor: Dr. Hongmei Zhang.

Epigenetics is the study of heritable changes in genes which are caused by chemical compounds derived from natural and man-made sources. DNA methylation an epigenetic phenomenon, is most vulnerable to environmental factors during embryogenesis, which is a period of rapid cell division and epigenetic remodeling. Given the recent increase in the incidence of childhood diseases, it is crucial to understand the role of environmental factor through epigenetic study in causing adverse health effects. This dissertation revolves around three major hypotheses. In the first hypothesis we evaluated the association between in utero arsenic exposure and genome-wide DNA methylation in cord blood from the birth cohort data of Taiwan. The identified CpG sites were replicated in an independent birth cohort (New Hampshire birth cohort study; NHBCS) and further assessed longitudinal associations of DNA methylation with disease biomarkers measured at later ages in our cohort from Taiwan. In the second hypothesis we assessed the association between Immunoglobulin E (IgE) production and DNA methylation at birth via cord blood in a longitudinal study. The study was conducted from the birth cohort data of Taiwan and the findings were replicated in an independent birth cohort (Isle of Wight; IoW), and further the stability of identified CpG sites was assessed based on intra-class (ICC) correlation measure. In the third hypothesis we assessed the confounding effect in epigenome wide association study due to underlying cell composition and evaluated several methods and algorithms proposed to adjust for this confounding effect.

List of Abbreviations

CpG: 5'-cytosine-phosphate-guanine-3'; CpGs: Multiple CpG

LDL: Low density lipoprotein

tAs: Total arsenic obtained by adding inorganic arsenic (iAs), mono-methylated arsenic (MMA), di-methyl arsenic (DMA)

coeff: Coefficient; coeff.m: coefficient for main effect; coeff.int: coefficient for interaction effect.

DNA: deoxyribonucleic acid

DNA-M: DNA methylation

TSS: Transcription start site; TSS1500: within 1500 base pairs of a TSS; TSS200: within 200 base pairs of TSS.

FDR: False discovery rate

IoW: Isle of Wight

IgE: Immunoglobulin E

DAVID: Database for Annotation, Visualization and Integrated Discovery

KEGG: Kyoto Encyclopedia of Genes and Genomes

NHBCS: New Hampshire Birth Cohort Study

SVA: Surrogate variable analysis

GO: Gene ontology

ROS: Reactive oxygen species

DMRs: Differentially methylated regions

SNPs: Single nucleotide polymorphism; plural

Table of Contents

List of Tables

List of Figures

1 Introduction

The dissertation focuses on the identification of epigenetic markers from cord blood DNA methylation at birth associated with in utero arsenic exposure and immunoglobulin E (IgE). There is increasing evidence that in-utero arsenic exposure causes adverse health effects later in life [1, 2]. Arsenic is a potent human toxicant and carcinogen, but knowledge of the mechanism through which it exerts long term adverse health effects is limited. Inorganic arsenic and its methylated metabolites can easily cross the placenta and thus producing arsenic concentrations in cord blood similar to maternal blood [3]. The study of epigenetic changes such as DNA-methylation alterations that can affect gene activity may provide insight into mechanism through which arsenic exerts its adverse effects [4].

Immunoglobulin E (IgE) is known to play a major role in many of the allergic diseases such as asthma, atopic dermatitis (eczema) and hay fever. IgE production leads to type I hypersensitivity, which manifests various allergic diseases. However, the mechanism underlying IgE production is poorly understood. There is evidence that DNA methylation is associated with total IgE. An epigenome wide association study has a potential to shed more lights on the production of IgE.

Epigenome wide association studies are known to be influenced by the underlying cell compositions. Several algorithms and methods have been proposed to estimate or adjust for these underlying cell compositions. However, it is unknown which of the methods and algorithms works best.

1.1 Epigenetics

Epigenetics is the study of heritable changes in genes, which are caused by chemical compounds derived from natural and man-made sources. These heritable changes regulate genome by: (a) methylating DNA in genome, (b) modifying the histone, a protein that enables DNA to form long molecules. The chemical compounds that cause heritable changes in genome are known as epigenome. Much of the epigenome is reset when parents pass their genomes to their offspring; however, under some circumstances, some of the chemical tags on the DNA and histones of eggs and sperm may be passed on to the next generation [5-7]. When cells divide, often much of the epigenome is passed on to the next generation of cells, helping the cells remain specialized [7].

DNA methylation is an epigenetic phenomenon wherein the nucleotides in DNA is modified by addition of methyl group through covalent bond. In DNA methylation a methyl group is attached to the 5th atom in the 6-atom ring of cytosine leading to 5 methyl cytosine (5mC) or at 6^{th} position of adenine ring leading to 6-methyl adenine (6mA). The term CpG refers to the base cytosine (C) linked by a phosphate bond to the base guanine (G) in the DNA nucleotide sequence. In the human genome, it predominantly occurs at cytosine–guanine dinucleotide (CpG) sites, and serves to regulate gene expression and maintain genome stability DNA [8, 9]. Methylation is heritable and stable from one cell to another during cell division and thus leads to formation of epigenetic memory [10].

The epigenome is most vulnerable to environmental factors during embryogenesis, which is a period of rapid cell division and epigenetic remodeling [11-13]. Given the recent increase in the incidence of childhood immune-based diseases, it is crucial to understand the role of environmental stressors [14]. A stressor is a chemical or biological agent, environmental condition, external stimulus or an event that causes stress to an organism. Environmental studies

have shown that DNA methylation could be altered under environmental stress, by overall genome-wide reduction in DNA methylation content (global hypo methylation). This alteration in DNA methylation can alter the expression of underlying gene.

Arsenic (As) is one such environmental stressor whose exposure is known to alter DNA methylation both globally and in the promoter regions of certain genes [15, 16]. Upon entering the human body, inorganic As is methylated for detoxification. This detoxification process uses S-adenosyl methionine (SAM), which is a universal methyl donor for methyltransferases including DNA methyltransferases (DNMTs) that determine DNA methylation.

1.2 Arsenic

Arsenic is a metalloid found in numerous minerals usually in combination with sulfur and other metals. In reducing and oxygenated conditions, arsenite (AsIII), and arsenate (AsV), are the main oxidation states, respectively. Compounds of arsenic are divided into three major groups.

a) Inorganic arsenic compounds (arsenic trioxide, sodium arsenite, arsenic trichloride, arsenic acid, and arsenic pentoxide),

b) Organic arsenic compounds (arsanilic acid, methylarsonic acid, dimethylarsinic acid, arsenobetaine and arsenosugars), and

c) Arsine gas.

Arsenite and arsenate are the most common inorganic forms in water. Arsenic is mainly transported in the environment via water from both natural and anthropogenic sources. In some regions of the world, groundwater (used for drinking water) is naturally contaminated with arsenic due to arsenic rich geological formations. These areas include Bangladesh, China, Taiwan, West Bengal (India), and some parts of Argentina, Chile, Mexico, Vietnam, Australia, and the USA. In unaffected areas, the levels of arsenic are only a few micrograms per liter of

ground water, whereas, in affected areas, the levels may range from tens to thousands of micrograms per liter [17].

Ingestion of arsenic contaminated water is the primary route of inorganic arsenic exposure for the general population. Approximately hundred million individuals world-wide are at risk of elevated arsenic exposure, mainly via drinking water.

Most of the ingested inorganic arsenic is absorbed in the gastrointestinal tract and then reduced in the blood. In humans, inorganic arsenic is metabolized through the conversion of AsV to AsIII, followed by methylation to monomethylated and dimethylated arsenicals (MMA and DMA, respectively) [18].

1.3 Immunoglobulin E (IgE)

Immunoglobulin E (IgE) is one of five isotypes of human immunoglobulins and is produced by plasma cells. Immunoglobulin E (IgE) is known to play a major role in many of the allergic diseases such as asthma, atopic dermatitis (eczema) and hay fever. IgE production leads to type I hypersensitivity, which manifests various allergic diseases. Thus understanding the mechanism leading to the IgE production is a key to understanding the pathophysiology of various allergic disease.

1.4 Underlying cell compositions in cord blood

Cord blood is the blood that remains in the vein of the umbilical cord and placenta at the time of birth. Umbilical cord blood consists of various cell types such as nucleated red blood cells, granulocytes, monocytes, natural killer cells, B cells, $CD4^+$ T cells, and $CD8^+$ T cells. Thus, DNA methylation measured in cord blood represents weighted averages of these cell-type specific methylation levels, with weights corresponding to the proportion of the different cell types in a cord blood sample. Thus, epigenome wide association study assessing the association

between cord blood and an exposure of interest could be confounded by cellular heterogeneity. Identifying and sub setting each cell types is not practical in larger epidemiological studies. Thus, several algorithms have been developed to measure and adjust for cellular heterogeneity in whole blood.

1.5 Contribution

Epigenome wide association study can identify epigenetic markers that will help reveal the adverse developmental effect of in utero arsenic. Also the longitudinal study predicting the production of IgE associated with cord blood DNA methylation can provide useful insight into the developmental origin of immunity based disease. The main contributions of the work presented in this dissertation can be summarized as below

1. Identification of epigenetic markers at birth associated with in utero arsenic exposure.

2. Identification of epigenetic markers at birth linking in utero arsenic exposure to cardiovascular disease.

3. Identification of epigenetic markers at birth predicting the production of immunoglobulin E (IgE) at later ages.

4. Best method to adjust for the underlying cell compositions for epigenome wide association study.

1.6 Organization of the dissertation

The dissertation is organized as self-explanatory chapters related to the epigenetic marker identification. Chapter 2 presents the role of in utero arsenic exposure on fetal developmental programming and its adverse influence in later life. Chapter 3 discusses the role of DNA methylation at birth in predicting the IgE production at later ages. Chapter 4 compares several methods and algorithms to adjust for underlying cell compositions in epigenome wide association study. Finally, in Chapter 5, I summarize the work presented here.

2 Epigenetic marker identification at birth associated with in utero arsenic exposure

2.1 Abstract

Background

In utero arsenic exposure may alter fetal developmental programming by altering DNA methylation, which may result in a higher risk of disease in later life. We evaluated the association between in utero arsenic exposure and DNA methylation (DNAm) in cord blood and its influence in later life.

Methods

Genome-wide DNA methylation in cord blood from 64 subjects in the Taiwanese maternal infant and birth cohort was analyzed. Robust regressions were applied to assess the association of DNA methylation with in utero arsenic exposure. Multiple testing was adjusted by controlling false discovery rate (FDR) of 0.05. The DAVID bioinformatics tool was implemented for functional annotation analyses on the detected CpGs. The identified CpGs were further tested in an independent cohort. For the CpGs replicated in the independent cohort, linear mixed models were applied to assess the association of DNA methylation with low-density lipoprotein (LDL) at different ages (2, 5, 8, 11 and 14 years).

Results

In total, 579 out of 385,183 CpGs were identified after adjusting for multiple testing (FDR=0.05), of which ~60% were positively associated with arsenic exposure. Functional annotation analysis on these CpGs detected 17 KEGG pathways (FDR=0.05) including pathways for cardiovascular diseases (CVD) and diabetes mellitus. In the independent cohort, about 46% (252 out of 553 CpGs) of the identified CpGs showed associations consistent with those in the study cohort. In total, 12 CpGs replicated in the independent cohort were in the pathways related

to CVD and diabetes mellitus. Via longitudinal analyses, we found at 5 out of the 12 CpGs methylation was associated with LDL over time and interactions between DNA methylation and time were observed at 4 of the 5 CpGs, cg25189764 (coeff=0.157, p-value=0.047), cg04986899 (coeff. for interaction [coeff.int]=0.030, p-value=0.024), cg04903360 (coeff.int=0.026, p-value=0.032), cg08198265 (coeff.int = -0.063, p-value=0.0021), cg10473311 (coeff.int = -0.021, p-value=0.027).

Conclusion

In utero arsenic exposure was associated with cord blood DNA methylation at various CpGs. The identified CpGs may help determine pathological epigenetic mechanisms linked to in utero arsenic exposure. Five CpGs (cg25189764, cg04986899, cg04903360, cg08198265 and cg10473311) may serve as epigenetic markers for changes in LDL later in life.

2.2 Background

Arsenic, a widespread element in the environment, poses a serious threat to human health. Millions of people around the globe are exposed to arsenic from drinking water that exceeds the safe limit of 10 ppb as recommended by World Health Organizations [19]. Arsenic is known to easily pass through the placenta in humans and other mammals, producing arsenic concentrations in cord blood similar to maternal blood [3]. Epidemiological studies have reported that gestational arsenic exposure is associated with increased risk of non-cancerous and cancerous diseases in adulthood [20, 21]. For instance, a number of studies have shown that early life arsenic exposure is associated with later cardiovascular diseases (CVDs) [22-24]. In animal studies, in utero exposure to low level arsenic in the womb and in adulthood was found to be associated with diabetes mellitus [25].

The mechanisms through which in utero exposure to arsenic may result in a higher risk of various diseases are not well understood. However, harmful effects such as the generation of reactive oxygen species (ROS), which causes oxidative DNA damage, binding and inhibition of arsenic metabolites to enzymes, and perturbation of key signaling pathways, are thought to play certain roles in disease development [26]. In addition, clinical and epidemiological studies have observed that environmental exposure in early life can affect the risk of disease later in life through a phenomenon known as developmental programming [21, 27, 28]. The study of epigenetic changes such as DNA-methylation alterations that can affect gene activity may provide insight into developmental programming [4].

Studies found that chronic arsenic exposure in adults is associated with increased DNA methylation extracted from whole blood leukocytes [29, 30]. Experimental studies in animals have also shown that intra-uterine exposure to arsenic alters DNA methylation in offspring [31]. Some studies examined the association of genome-wide DNA methylation in cord blood with in utero arsenic exposure. Most of them were based on cohorts established in the United States and Bangladesh. These studies did not identify any statistically significant CpGs at the whole epigenome level and thus focused on the top 100 [32] or 500 CpG sites [16] potentially associated with in utero arsenic exposure, while the study by Kile et al [33] investigated the association of CpG sites in *p16*, *p53*, LINE-1 and Alu repetitive elements.

Our study, based on data from a prospective birth cohort study established in Taiwan, aimed to comprehensively assess genome-wide DNA methylation in cord blood in association with in utero arsenic exposures (using maternal urinary arsenic concentrations), identify CpG sites showing such statistically significant associations after adjusting for multiple testing by controlling false discovery rate (FDR), and examine possible pathways of genes involving the

identified CpGs. Additionally, we attempted to replicate our finding in an independent birth cohort (New Hampshire birth cohort study; NHBCS) and further assessed longitudinal associations of DNA methylation with disease biomarkers measured at later ages in our cohort from Taiwan. The findings will contribute to an improved understanding of the adverse mechanisms of in utero arsenic exposure on genome-wide epigenetic variation and whether epigenetic markers in cord blood can influence children's diseases risk later in life.

2.3 Methods

2.3.1 Data collection and pre-processing of birth cohort data from Taiwan

Taiwanese maternal infant and birth cohort description

The data resulted from the Maternal and Infant Cohort Study in Taiwan investigating various in utero and postnatal factors considered to affect child health outcomes [21]. All pregnant women participating in this study signed informed consent forms explaining the benefits and risks of participation. This study was approved by Human Ethical Committee of the National Health Research Institutes in Taiwan. Pregnant women who received medical care at a local medical center were invited to join this study between December 2000 and November 2001. At the beginning, 430 of 610 pregnant women volunteered to participate, on average at 8 weeks gestation. Of the 430 pregnant women, 127 were excluded due to non-compliance of providing samples. Thus, urine samples were obtained from 313 pregnant women during the third trimester (28-38 weeks of gestation). Five newborns could not be included due to loss to follow up. Of all mother-newborn pairs, 299 have cord blood samples collected and DNA methylation data was available for 64 cord blood samples with sufficient amount of good quality DNA for this epigenome assay.

Assessment of arsenic exposure

Participants provided a spot urine sample at the time of enrollment in this study (at eight weeks of gestation). Urine was frozen at -20 °C in a 10-ml polypropylene tube. Arsenite (As^{III}), arsenate (As^{V}), monomethylarsonic acid (MMA), and dimethylarsinic acid (DMA) were quantified by high-performance liquid chromatography/inductively coupled plasma mass spectrometry (HPLC-ICP-MS) and anion exchange columns (Hamilton PRP X-100 [10 mm particle size, 250 mm64.1 mm]). Total urinary arsenic (TUA) was calculated by adding iAs (As^{III} + As^{V}) + MMA + DMA. The limitations of detection (LOD) for the various species were 0.09 mg/L for As^{III}, 0.05 mg/L for As^{V}, 0.05 mg/L for MMA, and 0.04 mg/L for DMA. Creatinine was measured by the Beckman Synchron LX20 auto-system (Beckman Coulter, Brea, CA, USA) in the central lab of Chung-Ho Memorial Hospital of Kaohsiung Medical University using a spectrophotometric method with picric acid as the reactive at 520 nm. We used total arsenic (tAs) as the sum of inorganic arsenic (As^{III} + As^{V}) and organic arsenic (MMA and DMA) divided by urinary creatinine; this ratio was used in the subsequent analyses.

Assessment of creatinine

Creatinine was measured by the Beckman Synchron LX20 auto-system (Beckman Coulter, Brea, CA, USA) in the central lab of Chung-Ho Memorial Hospital of Kaohsiung Medical University using a spectrophotometric method with picric acid as the reactive at 520 nm.

Assessment of low density lipoprotein (LDL)

The LDL Cholesterol Direct (DLDL) method was used to measure LDL cholesterol from the serum and plasma of the participants using ADVIA® Chemistry systems.

DNA methylation

DNA was isolated from cord blood samples using buffy coat isolated from EDTA-treated blood (Gentra Puregene; Qiagen, Hilden, Germany) and bisulfite converted using the EZ DNA Methylation kit. Samples were randomized across several plates for epigenome-wide DNA methylation assessment using the Illumina Infinium Human Methylation 450 BeadChip which simultaneously profiles the methylation status of > 485,000 CpG sites with single-nucleotide resolution across the human genome.

DNA methylation is measured using beta values, calculated as M/ (M+U+ε), where M is the methylation signal of the target CpG, U is the unmethylated signal and ε=100, is a constant to protect division by zero. Thus, average beta-value (β) represents the percent methylation of the target CpG site and its value ranges between 0 and 1.

Quality control

The raw data of DNA methylation were pre-processed to achieve high quality for data analyses. The function *PreprocessSWAN* in the Bioconductor package *minfi* was used for normalization, background correction and peak correction. The function *preprocessSWAN* uses subset within array normalization (SWAN) technique, which normalizes Infinium type I and type II probes together within a single array. This technique reduces technical variability between arrays by accounting for differences in the comparison of the two probe types between arrays [34]. CpG sites located on sex chromosome and annotated probe SNPs within 10bp of the target CpGs were dropped to eliminate bias caused by subject's gender and bias due to genetic variability.

2.3.2 Data collection and pre-processing of birth cohort data from NHBCS (Replication Study sample)

NHBCS Cohort Description:

The New Hampshire Birth Cohort Study (NHBCS) is an ongoing prospective study that began in 2009 and includes over 1500 women from two regions of New Hampshire, USA, enrolled between 24-28 weeks gestation. Mothers were recruited into the cohort if they were literate in English, mentally competent, between 18–45 years old, and reported using a private, unregulated well as the primary source of home drinking water. Infants included in the cohort were singleton pregnancies. Pre- and post-delivery questionnaires were administered to collect self-reported sociodemographic, lifestyle, and medical history data, and a structured medical records review was employed to collect information from the pregnancy and delivery. Cord blood samples are collected on >80% of all deliveries. This study consisted of the first participants born in the study with available cord blood samples for DNA methylation analysis and mothers that were not missing for urinary arsenic or any of the covariate data (n=109).

Arsenic in maternal urine:

Measures of maternal urinary arsenic have been described thoroughly elsewhere [1]. Briefly, spot urine samples were collected between 24–28 weeks gestation with 30 μL of 10 mM diammonium diethyldithiocarbamate, then samples were frozen at –80°C until analysis. High-performance liquid chromatography inductively coupled plasma mass spectrometry (ICP-MS) system measured individual arsenic species. Samples with values below the limit of detection (LOD) were assigned a value equal to the LOD divided by the square root of two [2]. Total maternal urinary arsenic concentrations (U-As) were calculated as the sum of inorganic arsenic

(As^{III} & As^{V}), monomethylarsonic acid (MMA^{V}) and dimethylarsinic acid (DMA^{V}), which was then log_{10}-transformed prior to analyses.

Cord blood DNA-M processing and QA/QC:

DNA was bisulfite converted using the EZ DNA Methylation kit and subsequently subjected to epigenome-wide DNA methylation assessment using the Illumina Infinium® HumanMethylation450 BeadChip (Illumina, San Diego, CA) at the University of Minnesota Genomics Core Facility (Minneapolis, MN) following standardized protocols. Post-array processing was conducted in the 'minfi' package in R. Samples in which >2% of probes had poor detection p-values were excluded; probes with detection p-values > 0.01 in at least one sample were also removed. Functional normalization (funNorm) and ComBat were utilized to remove technical variations in the data; removal of batch effects was confirmed with principal components analysis. The normalized and batch-corrected beta-values were then transformed into M-values via $log_2(\beta/(1-\beta))$ prior to statistical analyses.

Covariates:

Maternal age, maternal BMI, and estimated cell proportions were included as continuous covariates. Child gender and mother's education were included as dichotomous covariates. Maternal education was defined as those with at least a college degree vs. those without a college degree. Cell type proportions were estimated with the current gold standard method [3] via the minfi package in R. In regression models, 5 of the 6 cell types were included as covariates (NK cells were excluded) to account for overall cellular heterogeneity.

Replication Sample

In total, 109 cord blood samples for which DNA-M had been obtained and complete arsenic and covariate data were available. Of the 579 CpG sites identified in the cohort in Taiwan, 553 were available for replication analyses within the NHBCS.

2.3.3 Correction for cell mixture proportion

Blood is a mixture of functionally and developmentally distinct cell populations [35]. Adjusting for this cell type will remove potential confounding effects of cell heterogeneity in DNA methylation in blood samples [36]. Cell type composition of the blood sample was calculated using function e*stimateCellCounts* in the R package *minfi* [37, 38]. IDAT files from 450k Illumina DNA methylation were used to estimate the proportion of 6 cell types: CD8T, CD4T, NK, Monocyte, Granulocyte and B-cell. Cell type proportions are provided in supplemental Table A1.2.

2.3.4 DAVID

Illumina Infinium 450K Human Methylation Beadchip array version 1.2 was used to map the significant CpG sites to USCS reference genome and identify genes associated with these CpG sites. Functional enrichment and pathway analysis of resulting genes was carried out using DAVID gene functional classification tool [39]. DAVID is a large gene-centered knowledgebase which integrates the diverse annotation resources in a centralized location. DAVID knowledgebase is based upon single-linkage algorithm called the DAVID Gene Concept, and it serves as a gene/protein IDs database. The DAVID gene functional classification tool aggregates a list of genes or associated biological terms into organized classes of related genes or biology. For a given gene list, DAVID identifies enriched functional-related gene groups and provides a visualization of genes on pathway maps such as KEGG [40]. KEGG pathway maps are

collection of manually drawn pathways based upon knowledge gained from experiments on functions of the cell and its metabolism. Genetic interaction in KEGG pathway represents the network of molecular interaction and reactions of gene products.

2.3.5 GeneMANIA

GeneMANIA [41] is used to build the network between query genes, based upon genes that are functionally similar. This connection between genes is established based upon their similar expressions and functional association across different conditions via published data. GeneMANIA uses the publicly available genomics and proteomics data, as well as organism-specific functional genomics data sets. Six organisms are currently supported by GeneMANIA (*Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Homo sapiens,* and *Saccharomyces cerevisiae*).

2.3.6 Statistical Analyses

The dataset consists of 64 samples from cord blood specimen with DNA methylation (DNA-M) data for 485,577 CpG site. After quality control using Bioconductor package *minfi* 385,183 CpG sites were retained for statistical analysis. The pre-processed DNA-M data in beta values were transformed to M values, approximated as log2 [β/(1-β)], in order to ensure a better fit to statistical model assumptions used in our analyses.

To identify CpG sites whose DNA-M is influenced by in utero arsenic exposure (tAs), robust regressions (lmFit function R-package *limma*) [42] were applied to model the association of DNA-M with urinary creatinine-adjusted total arsenic (tAs). Child's sex, batch effect, mother's age, BMI, and education level, and estimated blood cell proportions (CD8T, CD4T, NK, and B-cells, monocytes and granulocytes [37, 38]) were included as covariates. Robust regressions in *limma* package use an empirical Bayes approach to estimate sample variances

which provides stable inference when the number of arrays is small [43]. In the robust regression analyses, multiple testing is adjusted by controlling FDR of 0.05. For the replication analyses we reproduced the statistical models described above in the NHBCS sample (detailed description is given below). CpGs with regression coefficients are in the same directions were considered to be successfully replicated, and we attempted to control for multiple testing via FDR of 0.05.

To assess the association of DNA methylation at CpGs of genes in some of the identified pathways with longitudinal (2, 5, 8, 11 and 14 years) low-density lipoprotein (LDL), a biomarker for CVD and diabetes, we applied linear mixed models. Log_{10} LDL concentrations at different ages were the dependent variable and residuals of DNA methylation, age, as well as interaction between age and DNA methylation were included in the model as predictors, and child's age, sex, and birth weight were treated as covariates. A statistical significance level was set at 0.05. The residuals of DNA methylation were obtained by regressing DNA-methylation at each of 12 CpG sites on proportions of each of the six cell types (CD8T, CD4T, NK, and B-cells, monocytes and granulocytes) and batch.

2.3.6.1 Statistical Analyses in NHBCS:

For all 553 regressions, we tested the linear relationship between maternal total urinary As and cord blood DNA-M M-values while adjusting for confounders. Since batch effects were removed via ComBat during data processing, no batch variable was included in these analyses. The following model, consistent with the model used in the study based on data from Taiwan, was fit using robust regression via the lmFit function in the limma package in R (version 3.2.2), confidence intervals were extracted using the confint=TRUE option.

$$M\text{-}values = log_{10}(U\text{-}As) + Child\ Gender + Urine\ Creatinine + Mother's\ Age + Mother's\ BMI + Mother's\ Education + Estimated\ Cell\ Proportions$$

2.4 Results

The data are from a birth cohort study examining multiple in utero and postnatal factors in relation to child health outcomes as part of the nationwide Taiwan Maternal and Infant Cohort Study established in Taiwan in 2000-2001 [21]. In total, 64 subjects with genome-wide DNA methylation in cord blood, level of maternal urinary arsenic exposure, urinary creatinine, along with a child's sex, gestational age, maternal age, maternal pre-pregnancy body mass index (BMI) and the mother's educational level were available and utilized in the study. Table 2.1 presents the characteristics of pregnant women and newborns by sex. Of the 64 newborns, 38 (59.4%) were male. Maternal characteristics are comparable between male and female newborns, and there is no statistically significant difference in gestational ages between sexes of newborns.

Table 2.1. Characteristics of mothers and their newborns by newborn sex in Taiwan during 2000-2001 (n=64)

		Sex of the infant		
Characteristics	**All (n=64)[a]**	**Male (n=38)[a]**	**Female (n=26)[a]**	**p-value[b]**
Pregnant Women				
Age (years)	28.9±4.8	28.6±4.1	29.5±5.7	0.492
Pre-pregnant BMI (Kg/m2)	20.5±2.6	20.2±2.4	21.0±2.9	0.244
Urinary Creatinine (mg/dL)	63.6±41.7	70.9±46.0	53.0±32.9	0.078
Maternal Education				0.303
high school + 2 years	25(39)	13(34)	12(48)	
≥high school + 4 years	39(61)	25(66)	14(52)	
Newborns				
Gestational Age (weeks)	39±1.2	39±1.1	39±1.4	0.791

[a]Presented as the mean±SD or number (percentage).
[b]p-value for difference between male and female newborns using t-test for continuous variables and χ^2 or Fishers Exact Test for categorical variable

The levels and distribution of arsenic metabolites in maternal urine after adjusting for creatinine levels are shown in Table 2.2, distinguishing between mono-methylated arsenic (MMA), di-methylated arsenic (DMA), inorganic arsenic (iAs), and the sum of the three (total arsenic or tAs). Concentrations of each urinary arsenic species showed a large variation among the 64 mothers. We focused on tAs to represent overall arsenic exposure. The distribution of tAs

is severely skewed with a median of 23.19 µg per gram creatinine (µg g^{-1} crea [creatinine]), and 5^{th} and 95^{th} percentiles being 3.76 µg g^{-1} crea and 76.02 µg g^{-1} crea, respectively (Table 2.1 and Supplemental Figure A1.2). The results reported in this article are based on log10-transformed total arsenic concentration.

Table 2.2. Distribution of creatinine-adjusted concentrations of urinary arsenic species (iAs, MMA, and DMA) (n=64)

Exposure variables[a] \ Percentile[b]	Min	5th	25th	50th	75th	95th	Max
As metabolites (µg g^{-1} crea)							
MMA	0.06	0.08	0.19	0.40 (1.10)	1.67	6.14	28.5
DMA	0.07	3.09	11.27	20.73 (14.58)	29.5	70.75	129.1
iAs	0.11	0.19	0.41	0.83 (0.61)	1.33	4.74	6.55
tAs	0.34	3.76	12.09	23.19 (16.29)	33.29	76.02	137.5

[a]Abbreviations: iAs represents the sum of As^{3+} and As^{5+}; MMA: methylarsonic acid; DMA: dimethylarsinic acid; tAs: the sum of iAs, MMA, and DMA; µg g^{-1} crea: µg per g creatinine
[b]This study. Pregnant women from Maternal Infant cohort in Taiwan (n = 64)
LOD of detection for As^{3+} is 0.09 µg/L, As^{5+} is 0.05 µg/L, for MMA it is 0.05 µg/L and for DMA it is 0.04 µg/L
The values inside parenthesis are the average value of unadjusted arsenic expressed as µg/L.
The standard deviation for unadjusted tAs is 16.22 µg/L and interquartile range for adjusted tAs is 21.21 µg g^{-1} cre.

After pre-processing the DNA methylation data as depicted in Figure 2.1, 385,183 CpG sites were analyzed. The flow for the analyses is depicted in Figure 2.2.

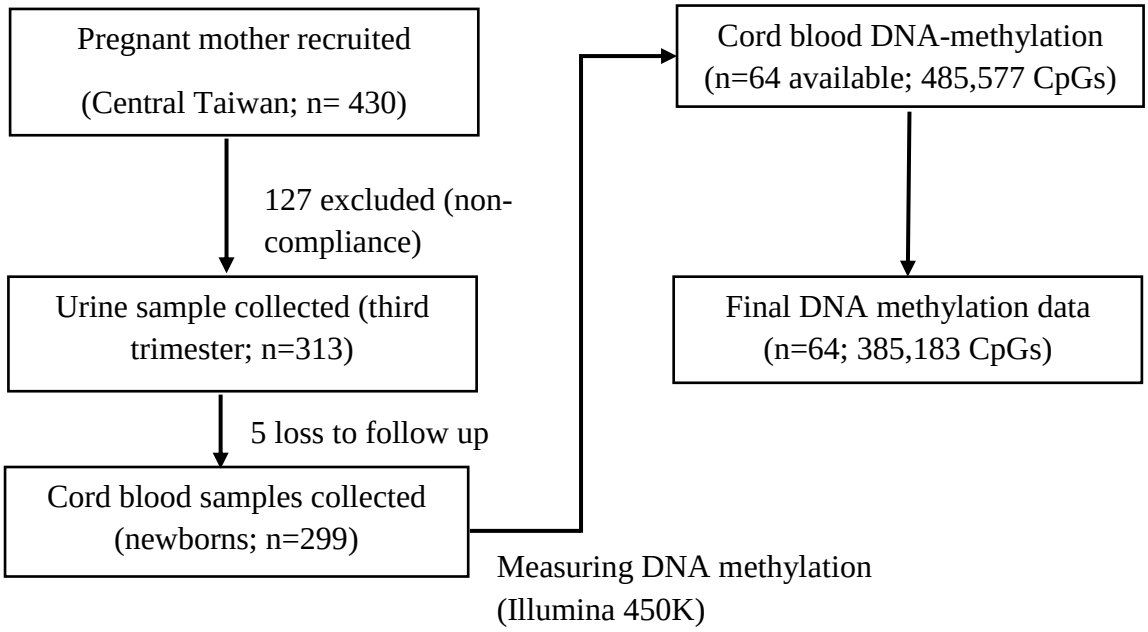


Figure 2.1. Subject recruitment and preprocessing of DNA methylation data in Taiwanese birth cohort

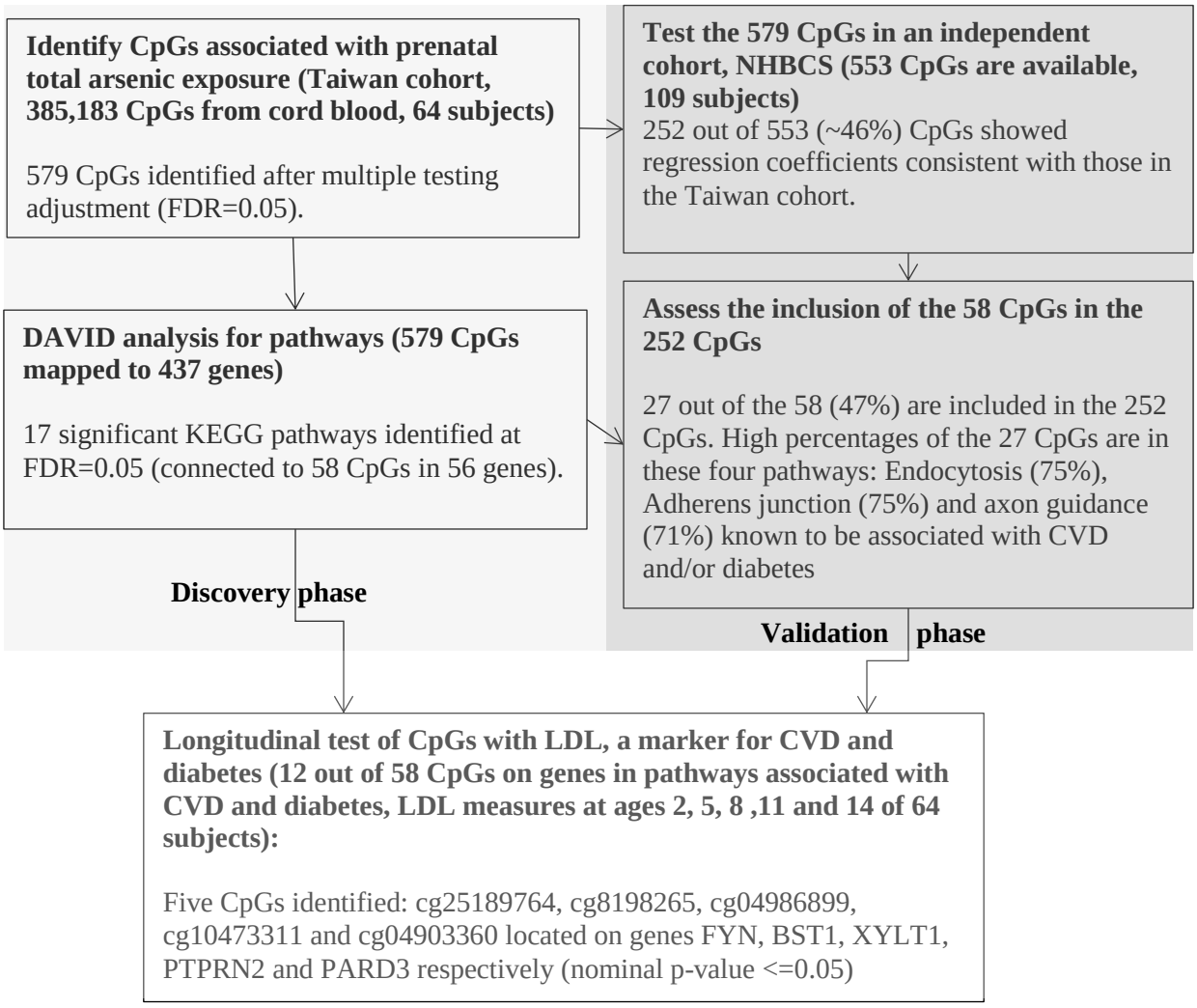


Figure 2.2. The flow of analyses performed in the study.

Epigenome-wide assessments of statistical associations between $\log_{10}$ creatinine-adjusted maternal urinary arsenic level and logit transformed DNA methylation (also noted as M values) were conducted via robust regressions. Covariates included in robust regressions were child's sex, batch of DNA methylation analyses, mother's age, mother's pre-pregnancy BMI, mother's education level, and estimated proportions of six blood cell-types (Appendix Table A1.2, related methods are in the Methods section). Figure 2.3 shows the Manhattan plot of p-values for testing on the 385,183 CpG sites, with a dashed blue line indicating the p-value threshold corresponding to FDR of p=0.05 [44]. In total, 579 CpG sites showed statistically significant associations at FDR of 0.05. Supplemental Table A1.1 lists these 579 CpG sites along with their regression coefficients, p-values, and corresponding chromosomes, locations on the chromosomes, corresponding genes, and location on the genes. About 60% of these 579 CpGs showed a positive association between DNA methylation and in utero tAs. The majority of the CpG sites located in the North shore regions of the CpG Island had higher DNA methylation associated with higher in utero tAs and about 39% of these CpG sites were located upstream of transcription start site (TSS1500, TSS200) or 1st Exon (Appendix Table A1.1).

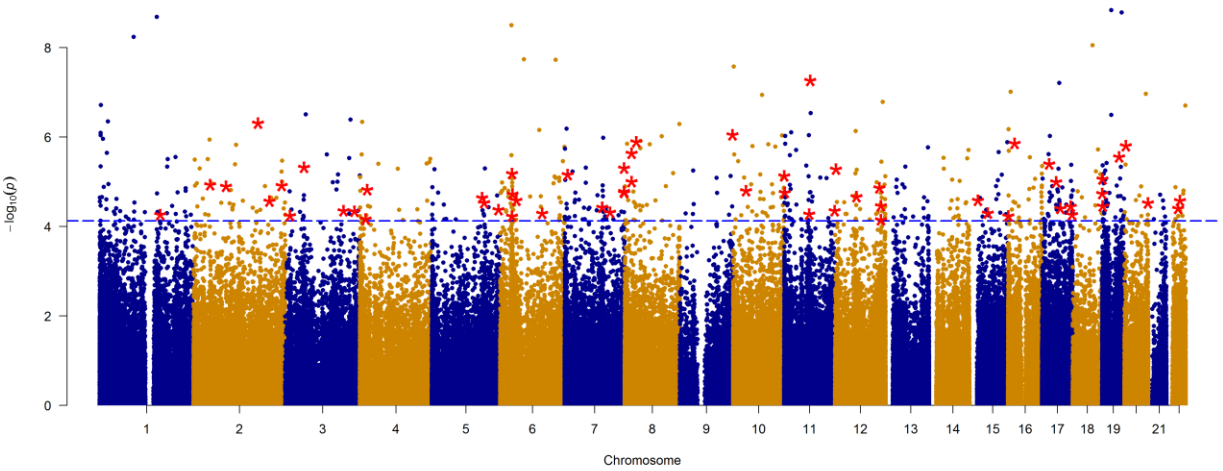


Figure 2.3. Manhattan plot for Genome-wide DNA methylation associated with creatinine adjusted urinary arsenic concentration. The horizontal dashed blue line corresponds to the significance threshold p = 7.51E-05 (FDR Adjusted p-value <= 0.05), red color stars represent the CpG sites corresponding to genes enriched in KEGG pathways from DAVID analysis. Blue and golden colors are used to differentiate the chromosomes.

The 579 CpG sites were mapped to 437 genes (Appendix Table A1.1), which were further analyzed using the bioinformatics tool DAVID [45, 46]. This analysis led to 17 significantly enriched KEGG pathways (at FDR=0.05) and 58 CpGs were within the genes involved in these pathways , including pathways connected to CVDs and diabetes [47] (e.g., Type I and Type II diabetes mellitus, focal adhesion, calcium signaling pathway, adherens junction, and chondroitin sulfate biosynthesis [48]), pathways linked to neurological and cognitive abilities (Alzheimer's disease and amyotrophic lateral sclerosis [ALS]), and pathways in cancer (the 58 CpG sites involved in these pathways are marked by red stars in Figure 2.3). The network, constructed using GeneMANIA [41] based on the genes enriched in DAVID analysis, indicated inter-connections among the genes (Supplemental Figure A1.3) via co-expression or shared pathways. Among these 58 CpG sites corresponding to the genes enriched

in KEGG pathways, most of them are located in the body region of a gene (Figure 2.4). Majority of these 58 CpGs are located in the island region (~57%) or north Shore (~22%). Furthermore, in approximately 55% out of the 58 CpG sites, we found that higher in utero tAs were linked to higher DNA methylation in cord blood, as indicated by positive regression coefficients in Figure 2.4. The strongest association between in utero tAs and cord blood DNA methylation occurred at CpG cg23767840, which is in the 5'UTR region of gene *EPN2* (coding for the Epsin-2 protein).

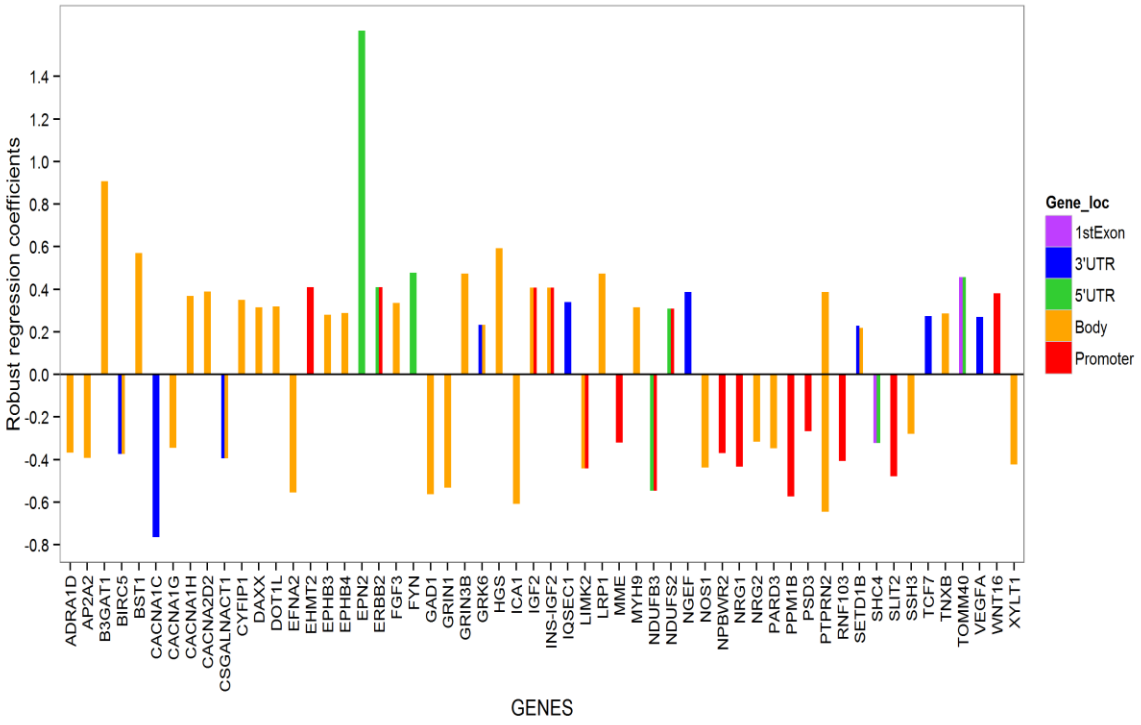


Figure 2.4. Association of arsenic exposure with the DNA methylation based on M-values of the 58 CpG sites mapped to 56 genes. The x-axis has the 56 genes enriched in KEGG pathways at FDR level of p=0.05, while the y-axis shows the estimates of total arsenic coefficients related to 58 CpG sites from robust regression. Adjusting factors include cell counts, child's sex, batch effect, mother's age, mother's BMI and mother's education level. M-values are defined as log2 [β/(1-β)]. Different colors indicate the location of the CpGs on a gene.

The resulting 579 CpG sites from our study were further tested in the independent New Hampshire Birth Cohort Study (NHBCS) (n=109). Of the 579 CpG sites 553 were available for analyses in NHBCS. We applied robust regression models with covariates comparable to those included in our study to assess the association of tAs with cord blood DNA methylation at these 553 CpG sites. At 46% of the 553 CpG sites (252 CpGs), the associations of in utero tAs with cord blood DNA-methylation levels were consistent with those found in our study in terms of direction of regression coefficients, although none survived multiple testing. In addition, 27 of these 252 CpGs are in the list of 58 CpGs (27/58=~47%) noted earlier that are involved in the enriched KEGG pathways (Table 2.3). Genes corresponding to these 27 CpGs are more often linked to pathways involved in endocytosis, adherens junction, axon guidance (a neural developmental process in which neurons send out axons to reach the correct targets) and chondroitin sulfate biosynthesis.

Table 2.3. KEGG pathways identified using DAVID that are more specific to arsenic exposure based on data from n=64 pregnant women from the maternal infant cohort in Taiwan.

KEGG-Pathways	**Genes**	**Adjusted p-value (FDR=0.05)**
Calcium signaling pathway	NOS1*, BST1, ERBB2*, GRIN1, CACNA1G, CACNA1H, CACNA1C, ADRA1D	0.000011
Endocytosis	PARD3*, AP2A2*, RNF103, PSD3, GRK6*, HGS*, IQSEC1*, EPN2	0.0000010
Axon guidance	NGEF*, LIMK2, FYN*, EFNA2*, EPHB3*, EPHB4, SLIT2*	0.000015
Alzheimer's disease	NDUFB3, NOS1*, LRP1, GRIN1, MME, CACNA1C, NDUFS2*	0.000033
MAPK signaling pathway	CACNA1G, CACNA1H, PPM1B*, CACNA1C, CACNA2D2, DAXX*, FGF3	0.00057
Regulation of actin cytoskeleton	LIMK2, INS-IGF2, SSH3, CYFIP1*, IGF2, MYH9*, FGF3	0.0019
Type I diabetes mellitus	ICA1*, INS-IGF2, PTPRN2*, IGF2, GAD1	0.0018
Amyotrophic lateral sclerosis (ALS)	NOS1*, GRIN1, TOMM40, DAXX*	0.0037
Adherens junction	TCF7, PARD3*, FYN*, ERBB2*	0.0087
Pathways in cancer	WNT16*, TCF7, ERBB2*, VEGFA, BIRC5, FGF3	0.0091
ErbB signaling pathway	ERBB2*, NRG1, NRG2, SHC4	0.01
Focal adhesion	TNXB*, FYN*, ERBB2*, VEGFA, SHC4	0.0097
Chondroitin sulfate biosynthesis	CSGALNACT1*, B3GAT1, XYLT1*	0.01
Neuroactive ligand-receptor interaction	PARD3*, GRIN1, NPBWR2, GRIN3B*, ADRA1D	0.017
Lysine degradation	DOT1L*, SETD1B, EHMT2*	0.034
Type II diabetes mellitus	INS-IGF2, CACNA1G, IGF2, CACNA1C	0.038
Huntington's disease	NDUFB3, AP2A2*, GRIN1, NDUFS2*	0.045

* CpG sites of these genes were consistently associated (in terms of regression coefficient) with total urinary arsenic exposure in an independent cohort NHBCS.

Given the connection of arsenic exposure with CVDs and diabetes [24, 25, 49, 50] , findings from the pathway analyses, and findings in the replication study, we further investigated the CpG sites of the genes enriched in KEGG pathways that are potentially linked to cardiovascular diseases and diabetes in our Taiwan cohort. In particular, 12 CpGs (located on 11 genes, Appendix table A1.1) were included in this analysis and these 12 CpGs were among the 27 CpGs replicated in the NHBCS cohort. We assessed the association of cord blood DNA methylation at these CpGs with a biomarker of CVDs and diabetes, plasma low density lipoprotein (LDL). LDL was measured at multiple ages of the children (at 2, 5, 8, 11, and 14 years). Plasma LDL concentration is the most stable in humans, with or without fasting, among blood lipids such as triglycerides. Among the 12 CpGs, cord blood DNA methylation at some CpGs showed a pattern of positive correlations with LDL at each age. While some were negatively correlated with LDL at age 2 and positively correlated at later ages (Figure 2.5), for most CpGs, the strongest correlations (positive or negative) occurred at age 2. In particular, the heatmap (Figure 2.5) indicated that DNA methylation levels at two CpGs, cg06419180 and cg25189764, were positively correlated with the LDL at different ages, while the directions of correlations at the rest of the CpG sites seemed to change over time. Via linear mixed models, we tested the association of LDL with DNA methylation (with LDL at ages 2, 5, 8, 11 and 14 as the outcome, cell type compositions and batch-effect adjusted DNA methylation as the predictor, and child's age, sex of the child, and birth weight as covariates) as well as the interaction effect between DNA methylation and age. We found that CpG cg25189764 had a statistically significant association with LDL (coefficient=0.157, p-value=0.047). DNA methylation at another 4 CpG sites showed statistically significant interaction with time, cg08198265 (coefficient for main effect [coeff.m] = 0.498, coefficient for interaction effect [coeff.int] = -

0.063, p-value=0.002), cg04986899 (coeff.m = -0.36, coeff.int = 0.030, p-value=0.024), cg10473311 (coeff.m= 0.145, coeff.int= -0.021, p-value=0.027) and cg04903360 (coeff.m= -0.189, coeff.int=0.026, p-value=0.032). It is worth noting that DNA methylation at these 5 CpG sites was found to be stable across the life course. The stability was assessed using Accessible Resource for Integrated Epigenomic Studies (ARIES) explorer [51].

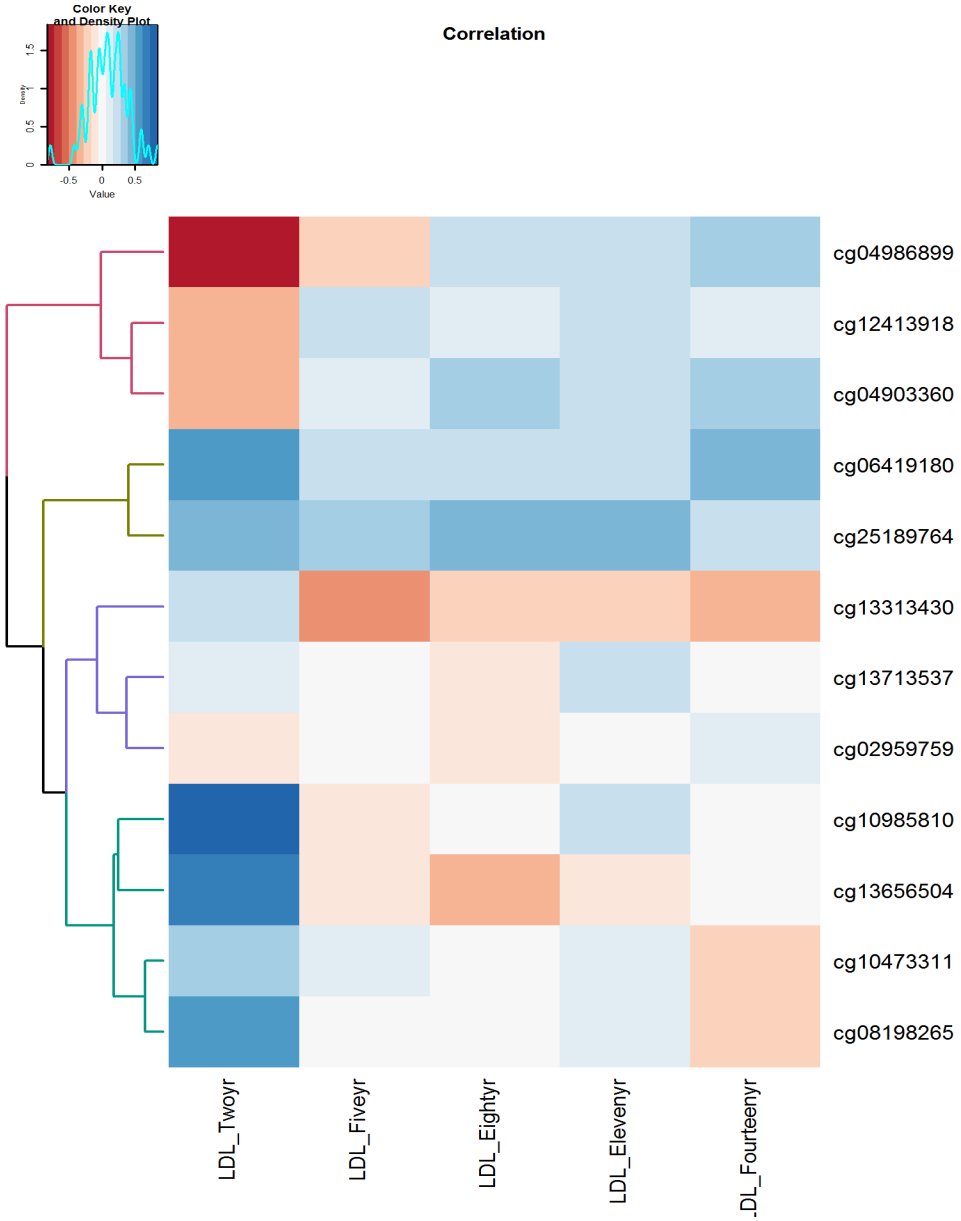


Figure 2.5. Heatmap of the correlations between cord blood DNA methylation and LDL across different ages (2, 5, 8, 11, 14 years).

2.5 Discussion

The overall aim of this study was to identify CpG sites that would represent biomarkers of possible adverse effects of arsenic in newborns and of future health outcomes. In total, at 579 CpGs identified from a cohort in Taiwan DNA methylation was associated with in utero arsenic exposure. To further understand the biological mechanisms of genes linked to these 579 CpG sites, a gene annotation analysis using DAVID was performed, which led to an identification of 17 statistically significant KEGG pathways. Genes corresponding to the identified CpGs are known to be involved in arsenic-associated diseases including neuronal [52-54], immune [55], cancer [56], cardiovascular and diabetes [25, 49, 50]. Experimental models have demonstrated a role of in utero acquired somatic epigenetic alternations in diseases [57-59]. Given the regulatory functionality of DNA methylation on different genes, the identified CpG sites may serve as epigenetic biomarkers of potential harmful effects of *in-utero* arsenic exposure among newborns.

Findings at 46% of the identified 579 CpG sites were replicated in an independent cohort, the NHBCS, with respect to directions of associations, though these did not survive multiple testing adjustments. However, the median tAs (without creatinine adjustment) in NHBCS was 2.8 µg/L with interquartile range (IQR) of 3.64 µg/L, which is substantially lower than that in the Taiwanese cohort (median= 11.51 µg/L and IQR= 16.80 µg/L). This difference, small sample sizes from both studies, differences in ancestry and unmeasured confounding may explain the limited agreement in the findings between the two cohorts.

The post hoc analysis on CpG sites replicated in the NHBCS cohort and related to genes enriched in KEGG pathways for cardiovascular disease and diabetes led to the identification of five CpG sites cg25189764, cg08198265, cg04986899, cg10473311 and cg04903360 located on genes *FYN, BST1, XYLT1, PTPRN2 and PARD3,* respectively. *FYN* is an important regulator of

whole body metabolism and is known to be associated with insulin sensitivity in mice [60]. *BST-1* is a glycosyl-phosphatidylinositol (GPI) and is expressed in abundant in pancreatic islet cells [61]. Proteins containing a *GPI* anchor play key roles in a wide variety of biological processes [62]. *XYLT1* is involved in heparan sulfate (a type of glycosaminoglycan; GAG) biosynthesis [63, 64]. *GAG*s have been studied for their role as a potential target in treating CVDs [65, 66]. Protein encoded by *PTPRN2* (also known as *IAR*) is a known autoantigen in insulin-dependent diabetes mellitus [67]. *PARD3* has been identified as candidate gene for its association with type 2 diabetes in Mexican study [68]. Out of these five CpGs, cg25189764 is located in the 5'UTR of gene *FYN*, and the other four CpGs were located in the body of the genes. We observed that most CpG sites on genes enriched in KEGG pathways were located in the body region of a gene (Figure 2.4). The regulatory functionality of DNA methylation on genes at those CpG sites is likely to be different from the functionality at CpG sites in the promoter region [69, 70]. Further assessment on their associations with gene expressions will improve our understanding of their regulatory functionality.

The temporal stability in DNA methylation at the five CpG sites (cg25189764, cg08198265, cg04986899, cg10473311 and cg04903360) showing associations with LDL across different ages raised a possibility of long term consequences of DNA methylation, established in utero, on LDL at later life. More interestingly, for the four CpGs (cg08198265, cg04986899, cg10473311 and cg04903360) showing interactions with age, the turning point of DNA methylation effects (from negative to positive effects or from positive to negative effects over time) are always around the age of 8 years. Of interest, ages 11 and 14 are during adolescence, a period of significant changes, e.g., puberty, rapid growth, and often BMI increase.

A previous study *in utero* arsenic exposure in the NHBCS was reported by Koestler et al. [32]. The top 100 CpGs identified in Koestler et al. did not overlap with the 579 CpGs, although 25% of their 100 CpGs showed statistical significance at the 0.05 level in our study (not surviving multiple testing). The disagreement could have been driven by some key differences in the analytical methods. Koestler et al. categorized arsenic exposure levels into quartiles and applied analysis of covariance with tests for trends, while our study applied robust regressions to log10-transformed arsenic concentrations to take into account possible outliers. By categorizing a continuous variable, statistical testing power for testing the associations might have been reduced. In addition, Koestler et al. did not adjust for maternal BMI, nor the cell type proportions estimated using the *minfi* R package [37, 38], though they did explore associations between urinary arsenic and estimated cell-type proportions in cord blood.

We also compared the findings from our study with another epigenome-wide study by Broberg et al [16]. The focus of that study also concentrated on the top CpG sites ranked by statistical significance on their association with in utero arsenic exposure, although none of the top CpG sites survived multiple testing corrections. The top CpG sites determined by Broberg et al. did not overlap with those identified in our study, nor overlapped with the top CpGs in Koestler et al. [32]. Broberg et al [16] utilized linear regression and did not adjust for cell type heterogeneity. In addition, some top CpG sites discussed in Broberg et al. included annotated probe-SNPs (single nucleotide polymorphisms) located within 10 base-pairs of the target CpG. They can result in biased methylation measurements, and were excluded from our analysis.

It is worth noting that the three studies we discussed herein (Koestler et al. [32], Broberg et al. [16], and ours) were conducted in different regions (United States, Bangladesh, and Taiwan, respectively) with vastly different medians in utero arsenic exposures which may have

limited replicability (for tAs, in Koestler et al., median=4.1 µg/L, in Broberg et al., median=66 µg/l, and in our study, median= 11.51 µg/L). It is also possible that ancestry, race/ethnicity or other regional differences may have contributed to the disagreement in the findings. In addition, all studies had small sample sizes (less than 200), so some of the findings are also likely to be false-positives. A large-scale study incorporating different races/ethnicities, with a wide exposure range, is well deserved. Our study had a benefit of replicating results using standard statistical approaches. Nonetheless, replicating DNA methylation analyses in additional populations, harmonizing, and comparing different DNA methylation studies on in utero arsenic exposure will help to assess the generalizability of the results. Future studies also should be directed at examining whether arsenic-related health outcomes are associated with cord blood DNA methylation in a long-term follow-up of the children in multiple cohorts.

2.6 Conclusion

We found that *in utero* arsenic exposure was associated with cord blood DNA methylation. The genes corresponding to the identified CpG sites were involved in various pathways including signaling pathways, Type I and Type II diabetes mellitus, and neuroactive ligand-receptor interactions. Cord blood DNA methylation at cg25189764, cg08198265, cg04986899, cg10473311 and cg04903360 were associated with low-density lipoprotein (LDL) at later life. These CpGs need to be studied further for their role in cardiovascular disease and diabetes in arsenic-exposed populations. Although larger studies are needed, results from this study contribute to a better understanding of epigenetic mechanism of diseases related to in utero arsenic exposure in infants.

3 Epigenetics markers at birth longitudinally associated with Immunoglobulin E (IgE)

3.1 Abstract

Background: Immunoglobulin E (IgE) is known to play a major role in allergic diseases. Epigenetic markings acquired due to modification of DNA methylation in early life may have phenotypic consequences later in development through their role in transcriptional regulation with relevance to the developmental origins of diseases including allergy. However, epigenome-wide studies on the association of cord blood DNA methylation and IgE over time are lacking.

Method: A total of 64-cord blood samples from Taiwan Maternal and Infant Cohort Study were analyzed using the Infinium Human Beadchip to obtain DNA methylation at ~450K Cytosine-phosphate-Guanine (CpG) sites. Linear mixed models were implemented to assess the association between preprocessed, batch and cell type corrected cord blood DNA methylation at >380k CpG sites with IgE levels at 5,8 and 11 years of age, adjusting for cord blood IgE. Identified statistically significant (at a false discovery rate, FDR, of 0.05) CpGs were replicated in an independent cohort, Isle of Wight (IoW) dataset. Gene ontology analysis was performed using DAVID to identify significantly enriched biological process of genes associated with resulting CpG sites. Stability assessment of the identified and replicated CpG sites was measured using ICC.

Results: DNA-methylation of 458 CpG sites were significantly (FDR <= 0.05) associated with IgE levels at different ages. Among the identified CpG sites available in both cohorts (n=241), of which, about 50% of CpGs were replicated in the IoW cohort in terms of consistency in direction of association between DNA methylation and longitudinal IgE levels. Gene ontology analysis of 84 genes linked to 124 CpG sites led to the enrichment of statistically significant biological process: PI3K-Akt signaling pathway, pathways in cancer, metabolic pathways, polymorphism,

alternative splicing, phosphoprotein, disease mutation, glycoprotein, protein transport, transcription regulation. Further temporal stability assessment of the 124 CpG sites identified 59 CpGs with significant ICC values (p-values<=0.05) at least 0.5.

Conclusion: Biological finding combined with the temporal stability measures for 59 CpG sites suggest them as a potential epigenetic marker for predicting later IgE production.

3.2 Background:

The prevalence of allergic diseases is increasing worldwide and the severity of allergic diseases, including asthma, continues to increase in children and young adults. About fifty percent of school children are sensitized to one or more common allergens [71]. Allergic disease is known to be hereditary implying individual's susceptibility to the genetic factors [72].

The gene-environment interaction during critical periods of immune development is assumed to be one of the causes of this disease later in life. Epigenetic variation is postulated to be an important mechanism through which these interactions are mediated [73]. Epigenetic processes regulate gene expression during immune development, and evidence suggests disruption in these processes can modify disease risk in a manner analogous to single nucleotide polymorphisms (SNPs) [74]. DNA methylation is one such epigenetic process which is associated with gene silencing and with the patterning of gene expression that determines cellular types and functions.

Immunoglobulin E (IgE) is known to play a major role in many of the allergic diseases such as Asthma, atopic dermatitis (eczema) and hay fever. IgE production leads to type I hypersensitivity, which manifests various allergic diseases. However, the mechanism underlying IgE production is poorly understood. There is evidence that DNA methylation is associated with total IgE. For instance, epigenome-wide study using Illumina methylation 27k array have

identified CpG loci from peripheral blood associated with total serum IgE [75]. Cross sectional study using peripheral blood of 18 year old men and women was used to identify association between CpG loci and serum IgE [76]. However, longitudinal epigenome-wide studies on the association of DNA methylation at birth via cord blood and IgE over time are lacking. Thus, it is unknown whether and how DNA methylation in cord blood is associated with IgE during the course of early life.

In this study we aimed at assessing genome-wide DNA methylation with respect to their association with IgE levels of child at ages 5, 8 and 11 years via linear mixed models adjusting for multiple testing by controlling FDR, and examine possible pathways of genes involved in the identified CpGs. We tested the identified CpGs in an independent cohort, the Isle of Wight (IoW) birth cohort in the United Kingdom. The findings will contribute to an improved understanding of in utero epigenetic development (or mechanism) leading to IgE changes later in life.

3.3. Method

3.3.1 Data collection and pre-processing of birth cohort data from Taiwan

In this section, we focus on the assessment of IgE. For other contents of this cohort including information related to DNA methylation measurement, quality control, and cell type compositions, please refer to Chapter 2, the Method Section.

Assessment of Immunoglobulin E

To determine serum IgE levels in children at 5, 8 and 11 years, blood samples (0.5 mL) obtained via venipuncture were centrifuged and the sera stored at −20°C prior to analysis. Serum total IgE (tIgE) levels were measured using the ADVIA Centaur chemiluminescence immunoassay system (Siemens Healthcare Diagnostics; Deerfield, IL, USA). The assay range was approximately

1.5–3000 IU/mL.

3.3.2 Data collection and pre-processing of birth cohort data from IoW

Participant selection and sample collection

The Isle of Wight (IOW) birth cohort was established to study the natural history of asthma and allergies in children born between January 1, 1989 and February 28, 1990 on the Isle of Wight, UK. The study was approved by the local research ethics committee (now named the National Research Ethics Service, NRES Committee South Central – Southampton B, 06/Q1701/34) and written informed consent was provided by the infants' parents (F1 generation). Details about the birth cohort have been described in detail elsewhere [32–34].

Immunoglobulin E (IgE) measurement:

Cord blood was taken from 1023 infants. Blood was collected from umbilical cord using fine needle in a specimen bottle containing dipotassium EDTA as anti-coagulant. Duplicate measurements of cord IgE were made using ULTRA EIA® kit (Pharmacia Diagnostics AB, Uppsala, Sweden) unmodified, designed to measure IgE between 0.2 and 50 ku/l on 0.1 ml of serum or plasma. Total IgE was measured in samples of serum collected at age 10 (n = 923) and age 18. IgE at age 10 and at age 18 were determined using PRIST® (Pharmacia Diagnostics AB, Uppsala, Sweden) designed to measure IgE between 2.0 to 1000 kU/L.

DNA sample collection and processing

Blood samples for the F1 generation were obtained from Guthrie card (n=34) of newborn babies. For methylation assays, DNA was extracted from whole blood using a standard salting out procedure [38]. DNA concentration was determined by the PicoGreen dsDNA quantitation kit (Molecular Probes, Inc., OR, USA). One microgram of DNA was bisulfite-treated for cytosine to thymine conversion using the EZ 96-DNA methylation kit (Zymo Research, CA, USA), following the manufacturer's standard protocol. This process converted unmethylated cytosines

into thymines (T) while leaving methylated cytosines (C) unaltered, allowing the array technology described below to distinguish between unmethylated and methylated sites by sequence (C vs. T) recognition [39]. Probes, 50 nucleotides in length, were developed to target specified CpGs throughout the genome with the CpG at the very end of the probe. Multiple copies of each probe were attached to a BeadChip, allowing for the interrogation of multiple DNA reads for each target CpG site [39,40]. After bisulfite conversion, the DNA samples underwent whole genome amplification, then each sample was washed over an array containing many BeadChips to identify the proportion of methylated probes for each target CpG.

Genome-wide DNA methylation was assessed using the Illumina Infinium HumanMethylation450K BeadChip (Illumina, Inc., CA, USA), which interrogates >484,000 CpG sites associated with approximately 24,000 genes. Arrays were processed using a standard protocol as described elsewhere [40], with multiple identical control samples assigned to each bisulphite conversion batch to assess assay variability and samples randomly distributed on microarrays to control against batch effects. The BeadChips were scanned using a BeadStation, and the Methylation Module of BeadStudio software calculated the methylation level for each queried CpG as beta (β) values. They represent the proportions of methylated (M) over methylated (M) and unmethylated (U) sites ($\beta = M/[c+M+U]$ with constant c introduced for the situation of too small $M+U$).

Quality control (QC) measures were employed to improve the reliability of data prior to analysis. In our study, the detection *P*-value reported by BeadStudio (Illumina software to process raw intensities) was used as a QC measure of probe performance, in which large *P*-values were deemed to be unreliable measures of methylation. Probes whose detection *P*-values > 0.01 in >10% of the samples were removed [41]. The methylation data were then preprocessed

using the Bioconductor IMA package for peak correction, background noise removal and batch effect correction [77]. The program for data cleaning was written in R (R Development Core Team, 2012). Some probes may overlap with known single nucleotide polymorphisms (SNPs), which may result in measurement errors [43]. SNPs at the target CpG [44,45] or within close proximity to the target CpG appear to be most likely to affect measurement of methylation, thus probes with SNPs within 5 nucleotides of the target CpG [46,47] and within the probe binding region were excluded from further analyses. Probes more than 5 nucleotides from the target CpG, but still within the binding region will be flagged with an indicator variable, and followed up if selected from the GWAS analysis.

3.3.3 Functional annotation and pathway analysis

DAVID analyses were performed in this study, and the method was discussed in Chapter 2. In the following, I focus on the statistical methods.

3.3.4 Statistical analyses

The final dataset consisted of 64 samples from cord blood specimen with DNA methylation data for 485,577 CpG site. After quality control using Bioconductor package *minfi* 385,183 CpG sites were retained for statistical analysis. The pre-processed DNA-M data in beta values were transformed to M values, approximated as log2 [β/(1-β)], in order to ensure a better fit to statistical model assumptions used in our analyses. To identify CpG sites whose DNA-M could predict IgE levels in children at 5, 8 and 11 years of age, the analysis was performed in two stages. In stage 1 we obtained the residuals of DNA-M by regressing DNA-M of each CpG (385,183) on cell proportions and batch effect. In stage 2 we used residuals of DNA-M to predict the IgE in longitudinal setting using linear mixed model, while adjusting for cord blood IgE, birth weight and gender of the child. Linear mixed model was fitted using proc mixed in SAS

9.4. Covariance structure was determined by comparing the Schwarz's Bayesian information criterion (BIC) of four covariance structure: Compound symmetry, Toeplitz, Unstructured and Auto-regressive1. To determine the best covariance structure 100 CpG sites were randomly selected to fit mixed model under each of the four covariance structure, and covariance structure with the smallest BIC was chosen for the final model. In all analyses, multiple testing is adjusted by controlling false discovery rate of p=0.05. Statistically significant CpG sites were further replicated in an independent cohort Isle of Wight (IoW), using the statistical method similar to our study.

An R package *irr* was used to estimate intra-class correlations of DNA-M at 124 CpG sites at birth via Guthrie card, at ages 10 and age 18 years in the IoW cohort.

3.3.4.1 Statistical analyses in the IoW

Analysis similar to described in 3.3.5 was performed for data from the IoW cohort. In the IoW, DNA methylation at birth was assessed from Guthrie cards. IgE was assessed at birth from Cord blood, at age 10 and age 18 years. Longitudinal association between residuals of DNA methylation and $\log_{10}$ IgE was assessed using proc mixed in SAS 9.4.

3.4 Results

The data are from a birth cohort study examining multiple prenatal and postnatal factors in relation to child health outcomes as part of the nationwide Taiwan Maternal and Infant Cohort Study [78, 79] established in Taiwan in 2000-2001. In total, 64 subjects with genome-wide DNA methylation from cord blood, child's gender, batch effect, and birth weight were available and utilized in the study.

Table 2.1 as described in Chapter 2 presents the characteristics of pregnant women and newborns by sex. Of the 64 newborns, 38 (59%) were male and 26 (40%) were female. The levels

and distribution of cord blood and serum IgE for children's at ages 5, 8 and 11 years is provided in Table 3.1.

Table 3.1. Distribution of IgE across different time points

Outcome Variable\Percentile	**Min**	**5th**	**25th**	**50th**	**75th**	**95th**	**Max**
IgE in Cord blood (IU/ml)	0.03	0.03	0.06	0.215	0.885	22.9	61.4
IgE at 5 years (IU/ml)	5.7	7.59	17.95	61.4	125.13	303.65	524
IgE at 8 years (IU/ml)	6.38	7.78	11	96.2	169	695	921
IgE at 11 years (IU/ml)	6	6	15.25	55	185.25	616.35	946

Epigenome-wide assessments of statistical associations between $\log_{10}$ IgE and residuals of DNA methylation (DNA-M) in cord blood at 385,183 CpG sites were conducted in a longitudinal setting via linear mixed modeling. This analysis was performed in two stages. In stage 1 we obtained the residuals of DNA-M by regressing DNA-M of each CpG (385,183) on cell proportions (CD4T, CD8T, NK, Mono, Bcell, Gran, Eos) and an indicator variables associated with different batches of DNA methylation data. In stage 2 we used residuals of DNA-M in cord blood to predict IgE at ages 5, 8, and 11 years using linear mixed models, adjusting for cord blood IgE, birth weight and gender of the child. In total, 458 CpG sites showed statistically significant associations with total IgE, after correcting for multiple testing by controlling FDR of 0.05. Figure 3.1 shows the Manhattan plot of p-values for testing on the 385,183 CpG sites, with a red line indicating the p-value threshold corresponding to FDR of 0.05 [44]. Appendix Table A2.1 lists the 458 CpG sites with their regression coefficients along with p-values, chromosomes they belong to, location on the chromosome, their corresponding genes and their location on the genes.

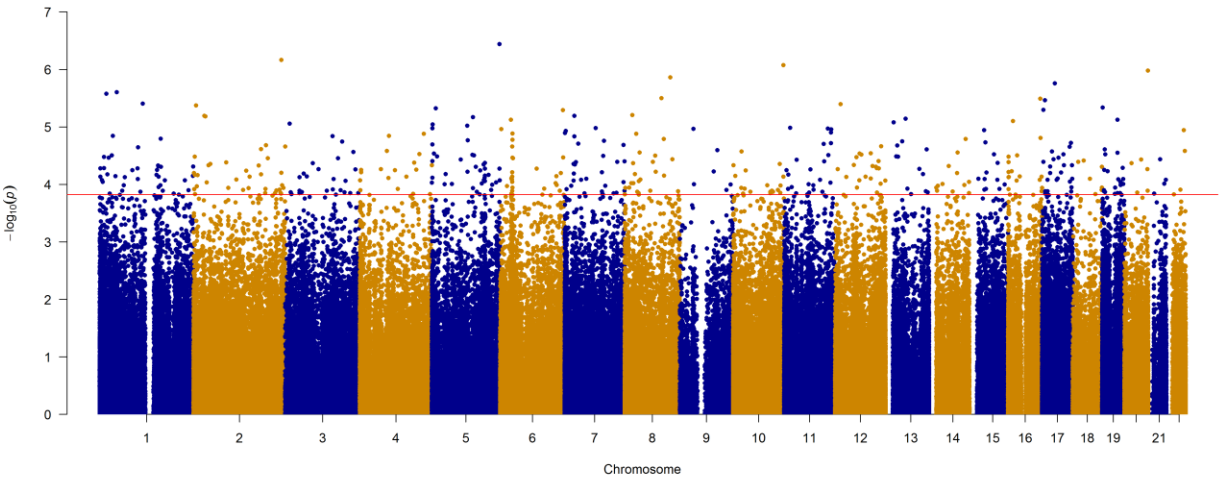


Figure 3.1 Manhattan plot for the longitudinal association of Genome-wide DNA methylation with $\log_{10}$ Immunoglobulin E (IgE). The horizontal dashed red line corresponds to the significance threshold p = 7.51E-05 (FDR Adjusted p-value <= 0.05). Blue and golden colors are used to differentiate the chromosomes.

The resulting 458 CpG sites from our study were further tested in an independent cohort, the Isle of Wight (IoW) birth cohort. Of the 458 CpG sites 241 were available for analyses in IoW. The flow of this study is provided in Figure 3.2. We used linear mixed models to assess the longitudinal association between residuals of DNA-methylation of 241 CpG sites from Guthrie card blood samples with serum IgE of 30 children measured at ages 10 and 18 years as outcome. The analysis was performed in two stages similar to the main study and used the same additional covariates. At about 51% of the 241 CpG sites (124 CpGs), the longitudinal associations of Guthrie card blood DNA-methylation with IgE over time were consistent with those found in our study in terms of directions of regression coefficients, although none survived multiple testing. Majority of the 124 CpGs were located in body (~50%) and promoter regions (~48%) of the genes, Figure 3.3. In addition, 22 of these 124 CpGs were nominally significant with p-value <=0.05. Functional annotation analysis using DAVID of 84 genes corresponding to 124 CpG sites led to identification of the following statistically significant (FDR <=0.05) KEGG pathways

and functional categories: PI3K-Akt signaling pathway, pathways in cancer, metabolic pathways, polymorphism, alternative splicing, phosphoprotein, disease mutation, glycoprotein, protein transport, transcription regulation.

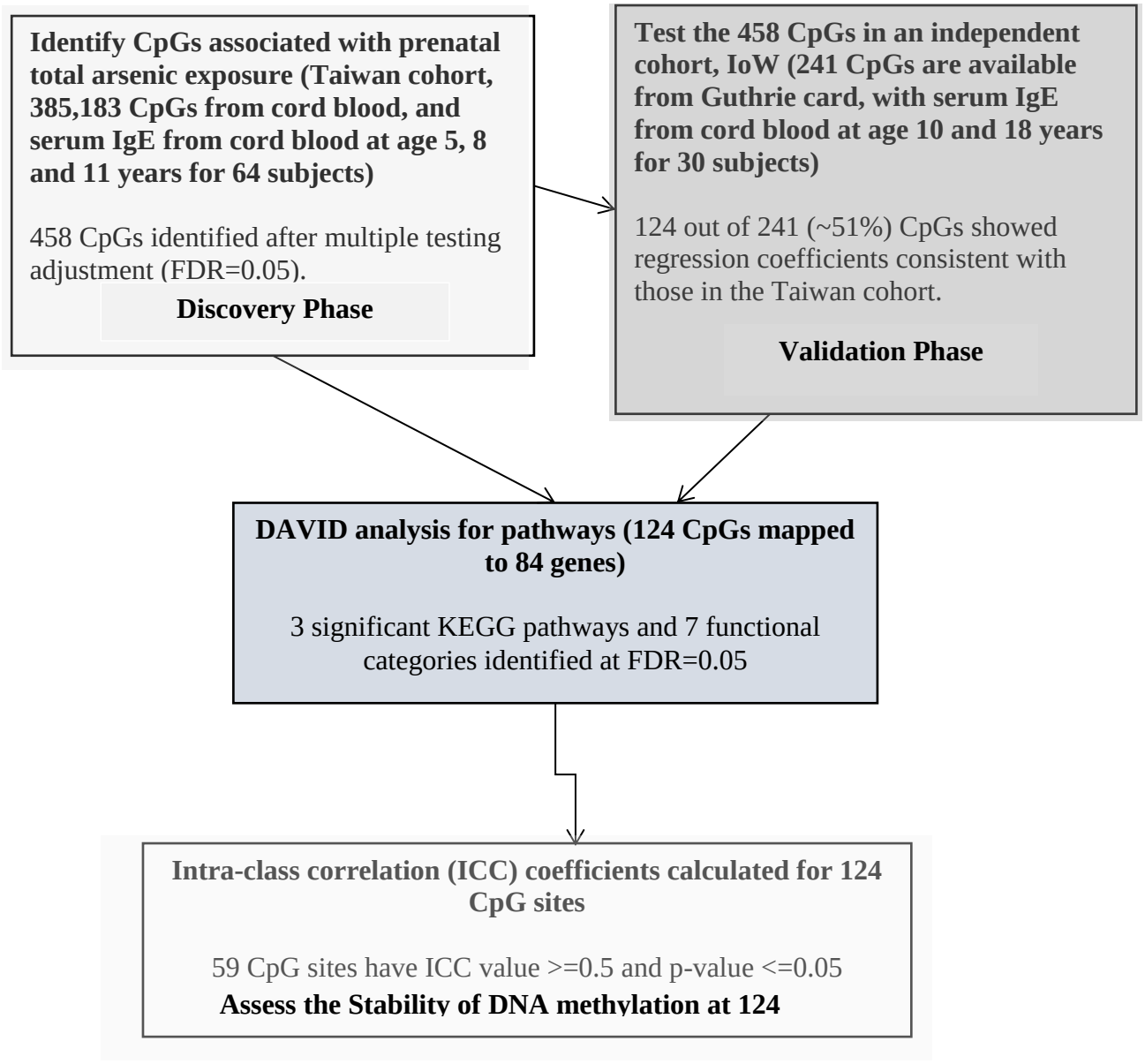


Figure 3.2 Flow of analysis.

The stability of these 124 CpG sites were inferred via intra-class correlation (ICC) based on DNA methylation data in the IoW cohort. In particular, DNA methylation of these 124 CpG sites at birth from Guthrie cards, at age 10 and age 18 years for five subjects were included in this assessment. Among the 124 CpGs, DNA-M at 59 CpG sites had ICC at least 0.5 (p-value<=0.05) (Appendix Table T2.2).

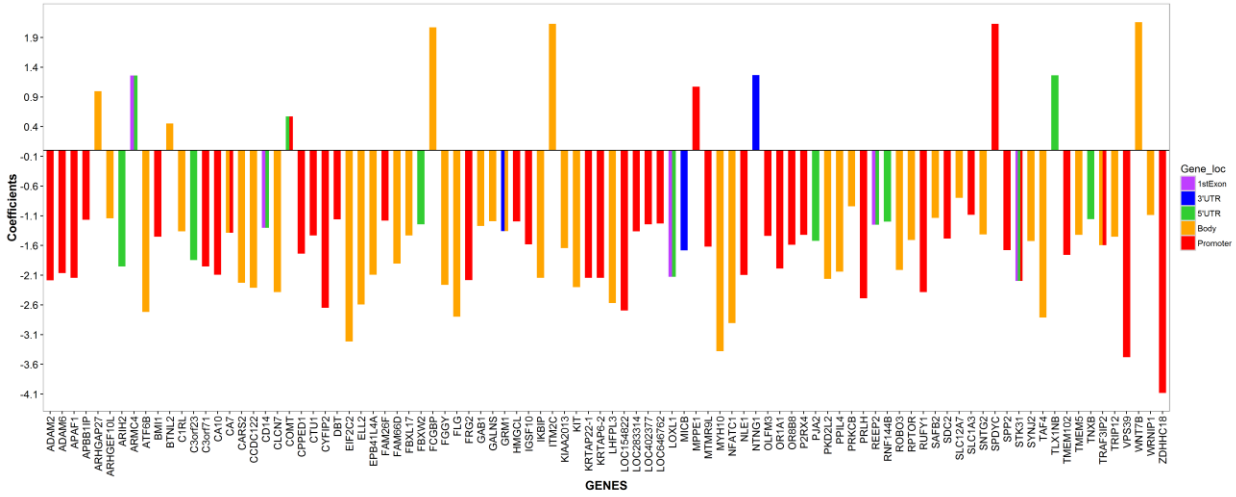


Figure 3.3 Longitudinal association of the residual of DNA methylation with $\log_{10}$ Immunoglobulin E (IgE) of the 124 CpG sites mapped to 89 genes. Please refer to method section for the detail of analysis. Different colors indicate the location of the CpGs on a gene.

3.5 Discussion

The overall aim of this study was to identify CpG sites of which DNA methylation at birth measured in cord blood could potentially predict the level of IgE at later ages. We identified 458 CpG sites (at FDR cutoff of 0.05) longitudinally associated with IgE in discovery cohort, 241 of these 458 CpG sites were available in the replication IoW birth cohort, and 124 out of the 241

CpG sites showed results consistent with those from the discovery cohort in terms of direction of association with IgE. These 124 CpGs were on 84 genes.

Functional analyses indicated that a number of genes associated with the 124 CpG sites were involved in biological processes related to IgE production. Phosphoinositide 3-kinase (*PI3K*) signaling is known to play a crucial role in IgE production, blockade of *P13K* enhances IgE levels [80]. IgE production is also known to be altered in cancer patients compare to control groups [81] and thus have been studied for their role in tumor surveillance and immunotherapy of cancer patients [82, 83]. Gene *CD14*, one of the 84 genes has been shown to elevate the IgE production [84]. Phosphodiesterase 11A (*PDE11A*) has been suggested for its role in asthma pathogenesis [85, 86], which indicates a potential role of this gene in IgE production.

Stability assessment of DNA methylation based on ICC at 124 CpG sites from birth to age 18 revealed that DNA methylation at 59 CpG sites is likely to be stable. SDA1 Domain Containing 1 (*SDAD1*) gene associated with one these 59 CpG sites has been known to contribute towards development of seasonal allergic rhinitis in Japanese population [87]. Fc Fragment Of IgG Binding Protein (*FCGBP*) another gene with CpG site having stable DNA methylation is known to share significant sequence homology with the carboxyl terminal of IgE binding protein [88]. Similarly, Immunoglobulin Superfamily Member 10 (*IGSF10*) and Butyrophilin Like 2 (*BTNL2*) have been known to be associated with IgE [89] and they both have CpG sites with stable DNA methylation.

Combining this finding with their associations with IgE at different ages, these 59 CpG sites have the potential to serve as epigenetic biomarkers for IgE changes over time.

3.6 Conclusion

Genome wide longitudinal assessment of 385,183 CpG sites with IgE production led to identification of number CpG sites. Among the identified CpG sites, 124 were replicated in an independent cohort for their longitudinal association with IgE production at later ages. Genes associated with these CpG sites were enriched in DAVID pathways and categories known to influence the IgE production. Most importantly 59 of 124 CpG sites have stable DNA-methylation across different time points. Thus, the identified CpG sites have the potential to serve as an epigenetic biomarker for IgE changes over time and can serve as candidates for future studies related to IgE production.

4 Assessment of methods for cell type correction in epigenome wide association study

4.1 Abstract

Background

Whole blood is frequently utilized in genome-wide association studies of DNA methylation patterns in relation to environmental exposures or clinical outcomes. These associations can be confounded by the cellular heterogeneity. Several algorithms have been developed to measure or adjust for this heterogeneity. However, it is unknown whether these approaches are consistent and if not, which method(s) perform better.

Results

Methods: We compared eight cell-type correction methods including the method implemented in the minfi R package, the method by Houseman et al., the Removing unwanted variation (RUV) approach, the methods implemented in FaST-LMM-EWASher, ReFACTor, RefFreeEWAS, and RefFreeCellMix R programs, along with one approach utilizing surrogate variables (SVAs). In the first comparison, we evaluated the association of DNA methylation at each CpG across the whole genome with prenatal arsenic exposure levels and with cancer status, adjusted for estimated cell-type information obtained from different methods. We then compared CpGs showing statistical significance from different approaches. For the methods implemented in minfi and proposed by Houseman et al., we utilized homogeneous data with composition of some blood cells available and compared them with the estimated cell compositions. Finally, for methods not explicitly estimating cell compositions, we evaluated their performance using simulated DNA methylation data with a set of latent variables representing "cell types".

Results: Results from the SVA-based method overall showed the highest agreement with all other methods except for FaST-LMM-EWASher. Using homogeneous data, minfi provided

better estimations on cell types compared to the originally proposed method by Houseman et al. Further simulation studies on methods free of reference data revealed that SVA provided good sensitivities and specificities, RefFreeCellMix in general produced high sensitivities but specificities tended to be low when confounding present, and FaST-LMM-EWASher gave the lowest sensitivity but highest specificity.

Conclusions

Results from real data and simulations indicated that SVA is recommended when the focus is on the identification of informative CpGs. When appropriate reference data are available, the method implemented in the minfi package is recommended. However, if no such reference data are available or if the focus is not on estimating cell proportions, the SVA method is suggested.

4. 2 Background

Whole blood is frequently utilized in genome-wide association studies of DNA methylation patterns in relation to environmental exposures or clinical outcomes. However, for DNA methylation assessed from whole blood, the association between DNA methylation and an exposure of interest could be confounded by cellular heterogeneity [90, 91]. In larger epidemiological studies, it is not feasible to isolate and profile every individual cell subset. Thus, several algorithms have been developed to measure and adjust for cellular heterogeneity in whole blood.

Houseman et al. proposed a method to infer the cell mixture proportions based on a regression calibration technique, which uses an external validation dataset to calibrate the model and correct for the bias [38]. Their approach was specifically designed for the Illumina 27k beadchip [92]. Jaffe and Irizarry [37] modified the Houseman et al.'s algorithm and tailored it to predict cell mixture composition of DNA-methylation profiles obtained from Illumina 450k

beadchip (450K array; Illumina, Inc., San Diego, CA, USA). This cell type correction method is implemented in Bioconductor [93] package minfi [94]. The above two approaches require external validation datasets and are designed to identify cell mixtures in tissues such as whole blood.

Apart from these two reference-based techniques, non-reference-based methods have also been developed. An advantage is that these non-reference based methods can also be applied to any other tissue in addition to blood. Zou et. al developed a non-reference-based method FaST-LMM-EWASher. This method is built upon linear mixed models with top principal components as the covariates. RefFreeEWAS [95] and its recently improved version (RefFreeCellMix)[96] are another two non-reference-based methods. They both utilize singular value decompositions (SVDs) and extract latent subject and cell-specific effects, but RefFreeCellMix incorporated additional constraints and utilities aiming to reduce the occurrence of false positives. Surrogate variable analysis (SVA) [97], based on SVDs of residuals in linear regressions, uses permutations to identify statistically significant eigen-vectors and consequently infer potential confounding factors (surrogate variables). A Bioconductor package is available to estimate surrogate variables using this approach [97]. ReFACTor [98]is another method that is free of reference database and it is based on principal component analyses on a set of potentially informative CpG sites.

Removing unwanted variation (RUV) is an approach designed to estimate cell type heterogeneity and built upon factor analyses. This approach utilizes reference CpGs inferred from reference database, based on which factor analyses are conducted. The factors are then included in subsequent analyses for the purpose of adjusting for cell type effects. Although

reference database is needed, this method does not estimate cell type proportions as in the minfi package and in the Houseman et al. method.

Among all these methods (eight approaches in total), it was unknown whether the existing methods were comparable to one another, and if not, which method(s) might perform better. To this end, we applied each cell type correction method (Houseman et al., minfi, RUV, FaST-LMM-EWASher, ReFACTor, RefFreeEWAS, and RefFreeCellMix) as well as the surrogate variable analyses (SVA) to real data sets. We evaluated the association between genome-scale DNA methylation and a variable of interest adjusting for cell type compositions. Then we compared the methods with regard to agreement on the number of CpG sites identified as being statistically significant. In addition, for the methods implemented in the minfi package and proposed by Houseman et al., we utilized homogeneous data with some blood cells composition available and compared these with the estimated cell compositions. For methods free of reference groups (FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, ReFACTor, and SVA), we further utilized simulated data generated under different scenarios to compare different methods, which, combined with findings from the real data, enabled us to demonstrate the quality of each method.

4.3 Method

In this section we briefly describe the existing techniques for estimating cell proportions or inferring latent variables due to cell compositions, data sets (real and simulated) to assess these methods, and statistical methods used in the analyses. All the analyses were programmed in R and a tutorial website including all the programs demonstrating the methods is available at https://akhilesh362.wordpress.com/.

Existing methods for cell compositions

4.3.1 Reference-based methods

Houseman et al. [38] developed a method for cell type correction that capitalizes on the idea that differentially methylated regions (DMRs) can serve as a signature for the distribution of different types of white blood cells. It uses these DMRs as a surrogate in a regression calibration based technique to identify the cell mixture distribution. Regression calibration technique can lead to bias estimate, thus external validation data is used to calibrate the model and to correct for the bias [99]. Their method was specifically for the Illumina 27k beadchip array.

The method by Jaffe and Irizarry [37] was adapted from the Houseman et al. [38] method and is tailored for Illumina450k along with 27k array. The algorithm in Houseman et al. identifies 500 CpG sites used to estimate cell mixture proportions from the Illumina 27k array. The modification of Jaffe and Irizarry was motivated because of the existence of probe SNPs in the 500 CpG sites and the inconsistency of CpG sites between the 27k and 450k arrays. In addition, the flow-sorted data of the six adult male subjects were used as references [35] .

The method of removing unwanted variation (RUV) uses information from reference database, but it does not estimate cell type proportions. Instead, this approach bases on the information of negative control probes and performs factor analysis on these probes to identify factors due to unmeasured confounders. These factors are then included in subsequent analyses to adjust for cell type effects. The negative control probes were chosen as top 500 CpG sites from the reference databases of DNA methylation known to be correlated with the cell types [100].

4.3.2 Reference-free methods

In total, four commonly used or recently developed reference-free methods are implemented in our study, FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, and ReFACTor. These methods do not need any external validation datasets and have the potential to adjust for cell mixture arising from any other tissue in addition to blood. FaST-LMM-EWASher [101] applies the maximum likelihood (ML) approach in linear mixed models and optimize spectral decomposition to estimate cell types [102]. RefFreeEWAS utilizes singular value decomposition (SVD) to decompose the residuals of unadjusted linear model along with unadjusted linear coefficient estimates, and estimate latent subject and cell-specific effects. Bootstrap estimates for coefficient standard errors are used to account for the correlation in the error structure.

Surrogate variable analysis (SVA) estimates potential confounding factors from a singular value decomposition (SVD) of residuals and was initially applied to gene expression data [103]. SVA utilizes the concept of expression heterogeneity while estimating surrogate variables. Expression heterogeneity (EH) refers to certain plausible biological profiles of the subject, which may not be captured by the covariates in study. Compared to the method in RefFreeEWAS, SVA decomposes the residual matrix and utilize permutations to identify statistically significant eigen-vectors which serve as a representative of EH (the so-called eigengenes), and then infer surrogate variables based on theses "eigengenes". Surrogate variables from SVA have the potential to cover information on cell types in DNA methylation from blood cells.

The method built in the R package RefFreeCellMix is improved from that in RefFreeEWAS. It uses a variant of non-negative matrix factorization to decompose the total methylation sites into CpG-specific methylation states for a pre-specified number of cell types

and subject-specific cell-type distributions [96]. Another approach in the R package, ReFACTor, a variant form of principal component analysis (PCA) to adjust for the cell type effects. This method assumes that a small number of methylation sites are affected by underlying cell mixtures. It filters out CpGs if the variation is not large enough (the default cutoff is standard deviation=0.02). To avoid too many CpGs filtered out, in our analyses, we excluded CpGs such that their standard deviations were in the lower 5^{th} percentile. By default this method searches for top 500 most informative methylation sites and performs PCA with a fixed number of components on these CpG sites to obtain the components. These ReFACTor components can be used as a covariate in epigenome wide association study or can be added one at time to remove the inflation due to cell type composition [98].

4.3.3 Three real data sets used to compare the approaches

These three data sets include data on prenatal arsenic exposure and DNA methylation, an example data from FaST-LMM-EWASher, and data on breast cancer status and DNA methylation. The first two data sets were utilized to demonstrate each of the five methods for cell type compositions and their agreement in terms of identified CpGs potentially associated with a variable of interest. The third data set was used to assess the agreement between the estimated cell type proportions (using the Houseman et al. method and the method in minfi) and the physical counts of the cells. This data set served as a benchmark and was critical for the comparison between the Houseman et al. method and the method in minfi. The benchmark data used to compare reference-free methods were simulated data, as discussed in the next section.

Prenatal arsenic exposure and DNA methylation data: The data were from a birth cohort study examining multiple prenatal and postnatal factors in relation to child health outcomes, part of the nationwide Taiwan prenatal and infant cohort study [78, 79] established in Taiwan in 2000-2001.

In total, 64 subjects with genome-scale DNA methylation and level of prenatal arsenic exposure were included in our study. DNA methylation data were pre-processed including quantile normalization, probe-type correction, and probe SNPs exclusion. After pre-processing, in total, 385,183 CpG sites were included in the analyses. All the five methods were applied to this data set. This and the following example data set were used to compare the performance of the five methods.

An example data from FaST-LMM-EWASher: This is an example data provided by the FaST-LMM-EWASher package [104]. It was originally used to illustrate the method incorporated into FaST-LMM-EWASher. In total, 204 subjects with cancer status and DNA methylation from Illumina 27K array on 25,978 CpG sites are available.

Breast cancer status and DNA methylation data: This data set has been previously described [105] and has genome-scale DNA methylation and breast cancer status available on 61 subjects at baseline and 39 subjects at six month follow-up along with complete blood counts. After pre-processing, 484,489 CpG sites were included in the study. In this article, we focus on granulocytes, monocyte and lymphocytes cells since proportions of these cells can be estimated by use of the minfi package and the original Houseman et al. approach. In our study, proportions of these cells from the physical counts were compared to the cell proportions estimated by minfi and the Houseman et al. method.

4.3.4 Simulated data sets to compare the methods

To further evaluate the three reference-free methods (FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, ReFACTor, and SVA), we simulated DNA methylation data under different settings with "latent" variables representing "cell types". These data sets served as benchmark

data for comparing reference-methods because the underlying truth was known. Two simulation scenarios were employed to evaluate the methods.

Scenario 1: We simulated DNA methylation data at 2,000 CpG sites across 600 samples, of which the first *n* CpG sites were associated with covariates of interest (e.g., level of arsenic exposure) and a set of latent variables, and the remaining CpG sites were only associated with the latent variables. The set of latent variables represent "cell types". One covariate of interest was considered and generated from a Normal distribution with mean 1 and variance 1 ($N(0, 1)$), The coefficients of this covariate was set at 0.3 and the intercept in the regressions was 0.5. Five "latent" variables were used and generated from five different Normal distributions: $N(0,5)$, $N(3,1)$, $N(0,1)$, $N(2,4)$, $N(0,3)$, respectively. The association of DNA methylation and the latent variables was assumed linear and the coefficients were generated from $N(0.5, 0.01)$. The distribution of random errors in the linear regressions was assumed to be Normal with mean 0 and variance 1.2 for the *n* CpGs, mean 0 and variance of 1.2 for the next 100 CpGs, and mean 0 and variance 2 for the remaining CpGs. The last setting with larger variance in random errors was for situations that the influence of cell types on DNA methylation was weaker.

We took three values of *n*, n=50, 100, and 150, representing different sparsity levels (from high to low) of informative CpGs. In total, 100 data sets for each *n* were simulated. Note that under this scenario, the covariates and latent variables were generated separately and had no correlations.

Scenario 2: Latent variables generated under this scenario have potential confounding effects. The overall setting is the same as in Scenario 1, except that the covariate of interest and the five latent variables (6 variables in total) were correlated such that correlation is equal to $0.7^{|i-}$

$^{j|}$, i, j = 1, 2, 3, 4, 5, 6. For instance, the correlation of the continuous covariate with the first latent variable was 0.7, and with the second latent variable was 0.7^2=0.49.

4.3.5 Statistical analyses

Linear regression-based analyses were used to assess the associations of DNA methylation with variables of interest with cell type heterogeneity adjusted using eight different methods. In the analyses of the two real data sets (the arsenic and DNA methylation data, and the FasT-LMM-EWASher example data), we recorded CpG sites showing statistically significant association with variables of interest (i.e., arsenic exposure and cancer) after implementing different cell type heterogeneity inference methods. We also inferred the number of statistically significant CpG sites without adjusting for cell type heterogeneity. To compare the eight cell type heterogeneity inference methods (Houseman et al., minfi, FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, ReFACTor, RUV, and SVA), we assessed the percentage of overlap between different methods in the number of identified CpG sites that showed statistical significance, and calculated a similarity index, Jaccard index (J-index) [106]. The percentage of overlap is calculated as the number of identified CpGs overlapped with that from SVA divided by the number of CpGs identified by SVA. We used Fisher exact test to assess the significance of overlap. Jaccard index measures the similarity between two finite sample sets. We used a Bioconductor package GeneOverlap to calculate this index. To assess whether the CpGs uniquely identified by the SVA approach are informative, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) [45, 46] to analyze the enrichment in Gene ontology (GO) [107] categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) [40, 108] pathways.

As for each simulated data set, we calculated sensitivity and specificity of the selected CpG sites for each cell type heterogeneity inference method. They were calculated by comparing the detected CpGs with the truly important CpGs. For each of the five methods, median of sensitivity and specificity along with 95% empirical intervals across 100 data sets were recorded for each setting under each simulation scenario.

4.3 Results

4.3.1 Findings from prenatal arsenic exposure and DNA-methylation data

We used genome-scale DNA methylation data from a birth cohort study consisting of 64 cord blood samples examining multiple prenatal factors in relation to child health outcomes, pilot of the nationwide Taiwan Maternal and Infant Cohort Study [78, 79].

We assessed the association of DNA methylation at each CpG site across the whole genome with prenatal urinary arsenic exposure levels (a continuous measure), adjusting for cell-type effects with cell type information inferred from one of the eight methods. For each method, the number of CpGs was recorded showing statistically significant associations with prenatal urinary arsenic exposure after adjusting for multiple testing by controlling false discovery rate (FDR) at 0.05. ReFACTor identified the largest number of CpGs (~60,000) and no CpGs were detected by FaST-LMM-EWASher (Table 4.1). RefFreeCellMix also identified a large number of CpGs (~3000). SVA and RefFreeEWAS detected more CpGs compared to the remaining methods. (Table 4.1). Next, we assessed the number of identified CpGs that overlap between different methods. The diagram in Figure 4.1 shows the overlap of CpG sites from four approaches (Houseman et al., minfi, RefFreeEWAS, and SVA) as well as the analyses without adjusting for cell types. Results from SVA showed the best agreement with findings from the other four analyses (Figure 4.1). Two identified CpG sites cg06434480 and cg10662395 were

common to all these five analytical methods labeled in Figure 4.1. Further comparisons indicated that CpG site cg10662395 was also identified by RefFreeCellMix and RUV, and this is the only CpG site overlapped among all the seven analyses (Houseman et al., minfi, RefFreeEWAS, SVA, RefFreeCellMix and RUV, as well as the analyses without adjusting for cell types). Although ReFACTor identified the largest number of CpGs, they did not overlap with the joint findings from the aforementioned seven analyses. Overall, CpGs identified via SVA overlapped with those from the Houseman et al. method, minfi and RefFreeEWAS (p-value<0.0001, Table 4.1, Figure 4.1. The definition of percentage overlap is given in the Methods section). One of the two CpGs (cg06434480 and cg10662395), cg06434480 is located within 200 base pairs of transcription start site of gene *HMGCR* (3-hydroxy-3-methylglutaryl-CoA reductase) known to be associated with inorganic arsenic exposure [109]. While in a study conducted in humans Mono-methylated arsenic (MMA) it was found to downregulate the gene expression of *HMGCR*, a gene involved in cholesterol biosynthesis [110]. The other CpG cg10662395 is located in the body region of gene *HCN2* (hyperpolarization activated cyclic nucleotide gated potassium channel 2). This gene was not found to be directly associated with arsenic exposure in the literature, but *HCN2* has been known to regulate pacemaker activity in the heart and the brain of mouse and human [111, 112], and arsenic has been found to induce QT interval (i.e., time between initial deflection of QRS complex to the end of T wave) prolongation probably by altering potassium ion channel [113].

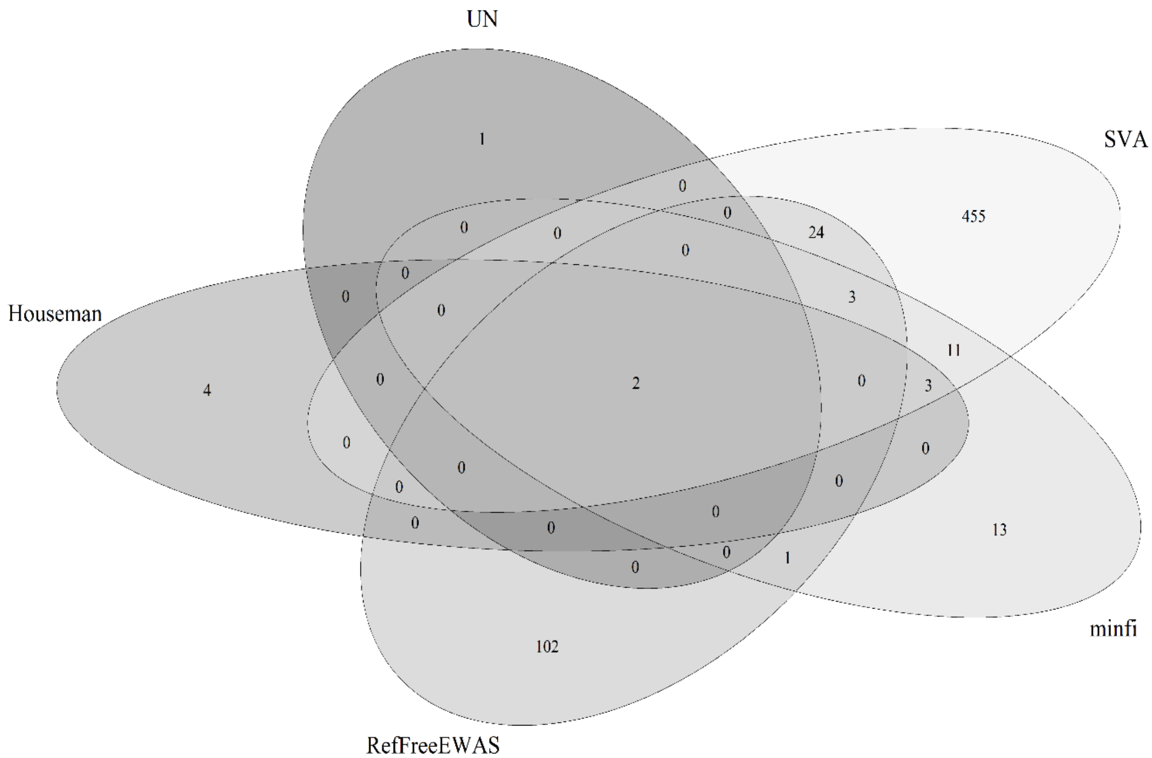


Figure 4.1. Venn diagram illustrating the overlap of identified CpG sites that are associated with prenatal arsenic exposure at FDR level of 0.05 after incorporating estimated cell type compositions by different methods for the association study of prenatal arsenic exposure with DNA-methylation. “UN”: results from an analysis without adjusting for cell type compositions.

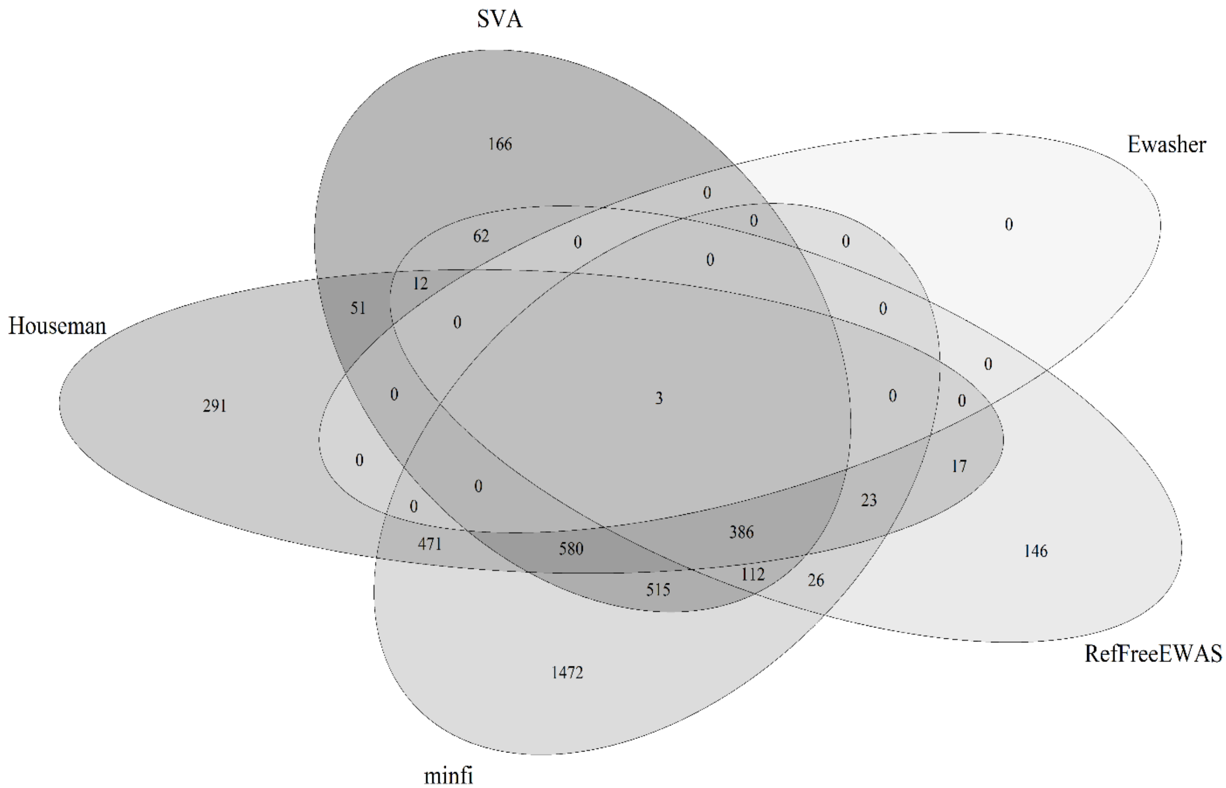


Figure 4.2. Venn diagram illustrating the overlap of identified CpG sites that are associated with cancer status at FDR level of 0.05 after incorporating estimated cell type compositions by different methods for the association study of cancer status with DNA-methylation.

The motivation of adjusting for cell types was due to the potential confounding effects of cell type compositions with respect to the association of arsenic exposure with DNA methylation, caused by the association of arsenic exposure with cell type compositions [114-117]. Our assessment on the correlations between total arsenic exposure and estimated cell type proportions also supported the potential confounding-role of cell types (Appendix Figure A4.1). To support the existence of such confounding effects, we assessed the associations with and without adjusting for cell type proportions at all CpG sites. We found that at more than 99% of all the CpGs the effects (regression coefficients) of prenatal arsenic exposure changed by more than

10% from cell type unadjusted (the median of the coefficients was 2.32 with 5^{th} percentile of 0.40 and 95^{th} percentile of 3.46) to cell types adjusted (the corresponding statistics were 0.080, 0.0073 and 0.25), indicating a need of adjusting for cell types.

Overall, the analysis based on SVA identified CpG sites that had better overlap with the CpGs identified by other methods. To acquire the biological relevance of CpGs uniquely identified by use of SVA, we implemented DAVID to perform Gene Ontology (GO) analysis and to identify KEGG pathways. The 455 (out of 498) significant CpG identified uniquely by SVA were mapped to genes using Illumina annotation file for 450K DNA methylation array. Of great interest, GO categories related to transcription and regulation of RNA metabolic process were enriched after controlling FDR at 0.05, as well as three KEGG pathways, endocytosis, cancer pathway and MAPK signaling pathway. A discussion on the connection of arsenic exposures and the identified GO categories and KEGG pathways is presented in the Discussion section.

Table 4.1. Number of significant CpG sites with and without cell type correction and overlap with the SVA method (data on prenatal arsenic exposure and DNA methylation).

Method	Identified CpGs (N)[#]	Overlap with SVA (%)	p-value[##]
Houseman et al.	10	1.20	<0.0001
minfi	57	4..62	<0.0001
SVA	498	---	---
RefFreeEWAS	133	6.01	<0.0001
RefFreeCellMix	2,932	0.60	1.0
ReFACTor	58,871	13.03	1.0
EWASher*	0	0.0	---
RUV	356	0.20	1.0
Unadjusted**	3	0.60	<0.0001

* The FasT-LMM-EWASher method.
** Unadjusted: cell type compositions were not included in the analyses.
[#]The selection of CpG sites is based on FDR-adjusted p-values (FDR is controlled at 0.05).
[##]P-value is based on Fishers exact test for overlap with results from SVA. The null hypothesis is that there is no overlap with the CpGs identified based on SVA.

4.3.2 Findings from example data

We repeated the same analysis on an example data set provided by the FasT-LMM-EWASher package. A tutorial website for applying all the cell type composition inference methods to this example data is available at https://akhilesh362.wordpress.com/. This data set includes DNA methylation from the Illumina 27K array and measures of a binary variable (cancer status) for 204 subjects. In total, 7,648 CpGs were included in our study based on initial screening done by the FasT-LMM-EWASher package. In this example data, cell type proportions were likely to be different on average between subjects with cancer and those without cancer, based on two-sample t-tests applied to logit-transformed sample proportions, explaining the potential need to adjust for their confounding effects. Since Illumina 27K focuses more on cancer genes, DNA

methylation at a large number of CpG sites showed statistically significant associations with cancer status (Table 4.2). Some similar findings as in Table 4.1 were observed. ReFACTor identified a large number of CpGs, Fast-LMM-EWASher identified the least number of CpG sites, and SVA agreed nicely with minfi (Jaccard similarity index=0.4). A unique observation from this analysis is that RUV identified the largest number of CpGs (6,008 CpGs, close to the number of CpGs in the candidate pool, 7,648 CpGs). Since the original Houseman et al. method was designed specifically for Illumina 27K platform, it is reasonable that SVA also showed a large overlap with results from this approach (Jaccard similarity index=0.4). In total, 3 identified CpGs (cg22029275 located in the 1st Exon of *FAM123A* gene, cg07080358 located in 1st Exon of *CNRIP1*, and cg15202954 located within 200 base pair of transcription start site of *NALCN* gene) were common to all the eight cell correction methods as well as to the analyses without cell type composition adjusted. There is evidence that these three genes (*FAM123A*, *CNRIP1 and NALCN)* are associated with the risk of colorectal cancer [118-120].

Table 4.2. Number of significant CpG sites with and without cell-correction methods and overlap of CpG sites with those from the SVA method (example data from FasT-LMM-EWASher package).

Method	**Identified CpGs (N)[#]**	**Overlap with SVA (%)**	**p-value[##]**	**J-index[###]**
Houseman et al.	1,835	54.71	<0.0001	0.40
minfi	3,589	84.59	<0.0001	0.40
SVA	1,888	---	---	---
RefFreeEWAS	788	30.51	<0.0001	0.30
RefFreeCellMix	1,006	18.38	<0.0001	0.10
ReFACTor	4,224	87.45	<0.0001	0.40
EWASher*	3	0.16	<0.0001	0
RUV	6,008	99.95	<0.0001	0.30
Unadjusted**	3,768	82.89	<0.0001	0.40

* The FasT-LMM-EWASher method.
** Unadjusted: cell type compositions were not incorporated into the analyses.
[#]The selection of CpG sites is based on FDR-adjusted p-values (FDR is controlled at 0.05).
[##] P-value is based on Fishers exact test for overlap. The null hypothesis is that there is no overlap with the CpGs identified based on SVA.
[###] J-index is Jaccard index.

DAVID analysis of genes associated with the significant CpGs identified uniquely by SVA led to the identification of three GO categories related to plasma membrane at FDR of 0.05 (integral to plasma membrane, intrinsic to plasma membrane, and plasma membrane part), as well as KEGG pathways such as pathways in cancer and signaling pathways, which indicates that genes corresponding to these CpG sites may play a role in the regulation of cancer.

4.3.3 Findings from breast cancer status and DNA-methylation data

This analysis uses a data set discussed in Smith et al. [105]. Breast cancer status, DNA-methylation, and cell counts for granulocytes, monocyte, and lymphocytes for 61 subjects at baseline and a subset of 39 subjects at six months follow up are implemented in the analyses. Among all the methods discussed, the method implemented in the minfi package and the original Houseman et al. method are able to estimate cell proportions. We used minfi and the Houseman et al. approach to estimate the proportions of granulocyte, monocyte and lymphocyte cells. Lymphocytes proportion were derived by adding the proportions of B cell, T cell and Natural Killer (NK) cells. For the three cell types (granulocyte, monocyte and lymphocyte), Pearson correlations between estimated (minfi) and true cell proportions were 0.85, 0.79, 0.88 at baseline and 0.84, 0.78, 0.87 at the six month follow up, respectively. For the correlations based on the Houseman et al. method, they were 0.84, 0.78 and 0.88 at baseline and 0.78, 0.73 and 0.83 at the six month follow up, respectively. All the correlations showed statistically significant difference from zero (p-value<0.05).

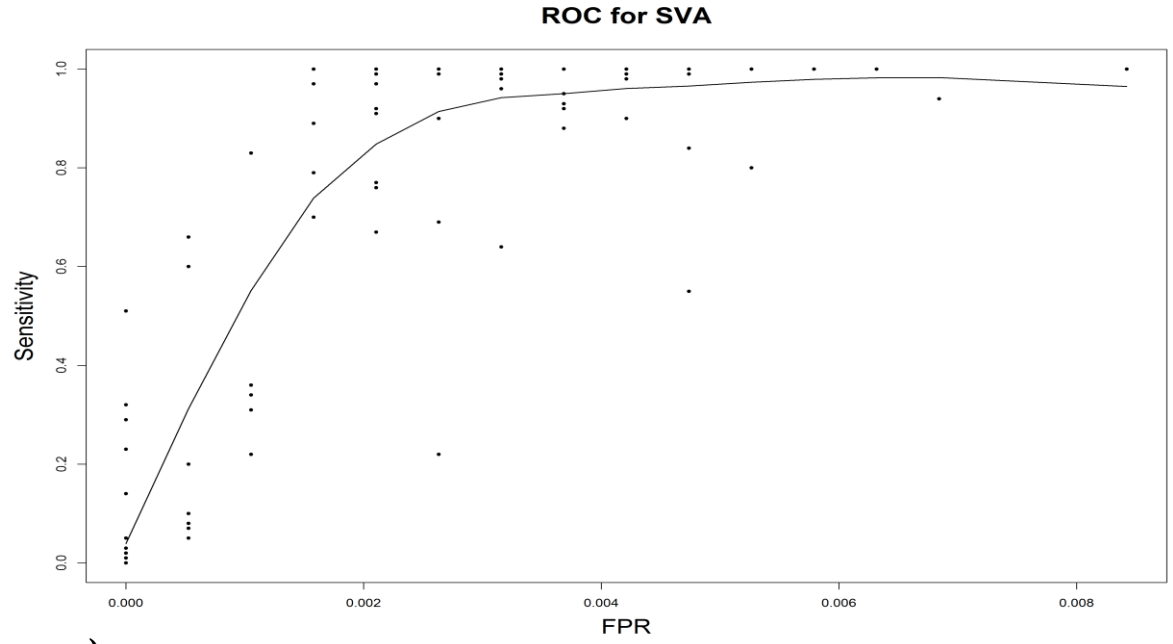
4.3.4 Findings from simulated data

We simulated data applying two scenarios with the first scenario focusing on latent variable effects (comparable to effects of cell composition), and the second focusing on latent variable effects with confounding (comparable to effects of cell composition as well as confounding effects). In total, 100 data sets were simulated under each scenario. Details of the simulation scenarios are given in the Methods section. The simulated data were used to evaluate the five methods that do not estimate cell proportions nor need reference databases, specifically, FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, ReFACTor, and SVA.

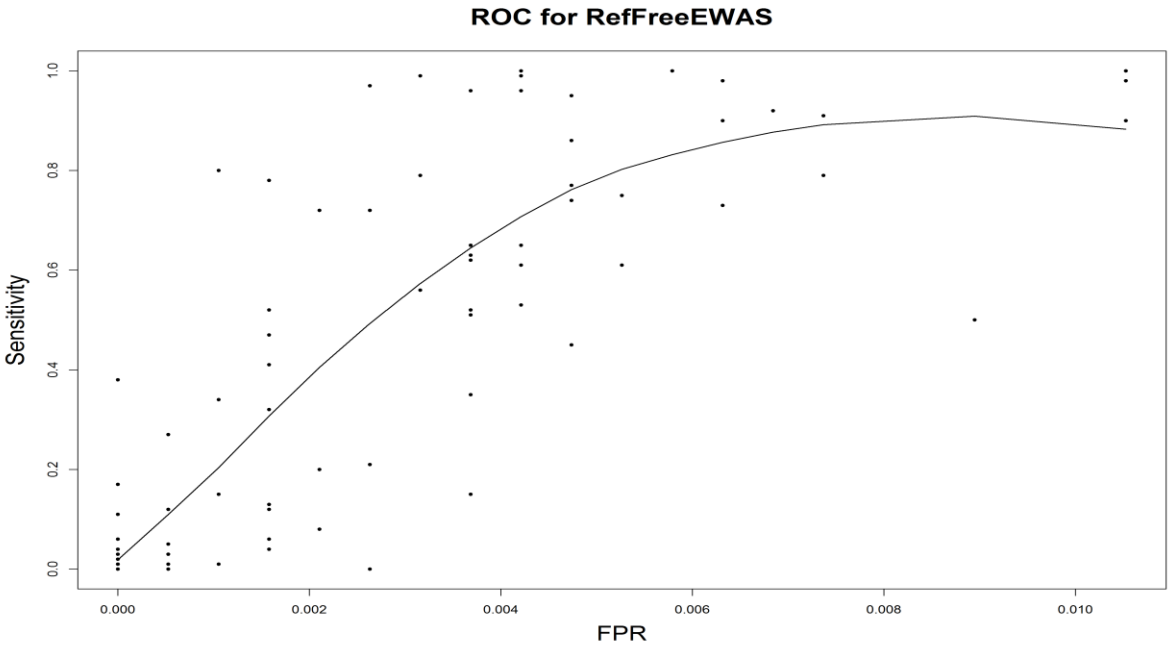
For data under all scenarios, we applied each of the five methods to each simulated data to draw information on cell compositions. We then incorporated the information to assess the associations of "DNA methylation" with the variable of interest at each pseudo CpG site, and compared each method by assessing the sensitivity and specificity of the selected CpG sites across all 100 data sets. Regardless of the number of important CpGs, FaST-LMM-EWASher resulted in the lowest sensitivity but the highest specificity for both scenarios, consistent with findings from real data (Table 4.3). Findings from RefFreeEWAS, RefFreeCellMix, ReFACTor, and SVA are, in general, comparable for data simulated under scenario 1, but SVA gives consistently higher sensitivity and specificity in all settings (Table 4.3). For data simulated under scenario 2 with high correlations (ρ=0.7), SVA outperformed FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix and ReFACTor and had higher sensitivity and specificity. Compared with RefFreeEWAS, overall RefFreeCellMix outperformed RefFreeEWAS when confounding effects present, showing much higher sensitivities with relatively lower specificities. Results from ReFACTor indicated extremely low specificity under scenario 2, which is consistent with the rather large numbers of CpGs identified in real data. The performance of FaST-LMM-EWASher was similar between the two scenarios and was inferior to all other methods. On the other hand, the SVA method performed well under both scenarios, followed by RefFreeEWAS and RefFreeCellMix with RefFreeEWAS being weaker in capturing confounding effects. We also considered a situation with ρ=0.3, mimicking a situation of moderate confounding, and similar patterns observed as those from the relatively two extreme cases (ρ=0 and ρ=0.7).

In the above simulations, we fixed the regression coefficients of the important CpGs. To demonstrate the pattern of sensitivity and specificity, we implemented receiver operating

characteristic (ROC) plots. In total, 100 data sets were simulated under scenario 1 with regression coefficients for the variable of interest ranged from 0.01 to 0.3. For each data set, we calculated sensitivity and specificity of selected CpGs, based on which we estimated the ROC curves. Sensitivities from FaST-LMM-EWASher were substantially low and were not considered in this demonstration. The performance of RefFreeEWAS, RefFreeCellMix, and ReFACTor was comparable under scenario 1 (Table 4.3). We therefore only presented ROC curves for ReFreeEWAS and SVA for the purpose of comparison (Figure 4.2). The findings are consistent with what we observed from Table 4.3 for scenario 1, that is, SVA performed better than RefFreeEWAS. In addition, the results indicated that both SVA and RefFreeEWAS have high specificity regardless of the underlying regression coefficients, indicating the conservativism when selecting informative CpGs.



a)



b)

Figure 4.3. Plots of sensitivity v.s. 1-specificity and estimated ROC curves, a) SVA. b) RefFreeEWAS

Table 4.3. Summary of sensitivity, specificity of FaST-LMM-EWASher, RefFreeEWAS, RefFreeCellMix, ReFACTor, and SVA for 100 simulated data across three settings.

	Sensitivity **(Median, 95% interval)**		Specificity **(Median, 95% interval)**	
	Number of Important CpGs =50			
	Scenario 1 (ρ = 0)	**Scenario 2** (ρ = 0.7)	**Scenario 1** (ρ = 0)	**Scenario 2** (ρ = 0.7)
Ewasher[a]	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
RefEWASRef[b]	1 (0.96, 1.00)	0.00 (0.00,0.49)	1.00 (0.99,1.00)	0.58 (0.06,1.00)
CellMix[c]	1.00 (0.98, 1.00)	1.00 (1.00, 1.00)	1.00 (0.99, 1.00)	0.55 (0.20, 0.92)
ReFACTor	1.00 (0.96, 1.00)	1.00 (1.00, 1.00)	1.00 (0.83, 1.00)	0 (0.00, 0.00)
SVA[d]	1.00 (0.98, 1.00)	1.00 (0.96, 1.00)	1.00 (0.996, 1.00)	1.00 (0.996, 1.00)
	Number of Important CpGs =100			
Ewasher[a]	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Ref[b]RefEWAS	1.00 (0.97,1.00)	0.00 (0.00,0.40)	1.00 (0.99, 1.00)	0.52 (0.01,1.00)
CellMix	1.00 (0.98, 1.00)	1.00 (1.00, 1.00)	0.99 (0.97, 1.00)	0.21 (0.05, 0.53)
ReFACTor	1.00 (0.97, 1.00)	1 (1.00, 1.00)	0.99 (0.81, 1.00)	0.00 (0.00, 0.00)
SVA[c]	1.00 (0.99,1.00)	0.99 (0.97, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
	Number of Important CpGs =150			
Ewasher	0.00 (0.00,0.00)	0.00 (0.00,0.00)	1.00 (1.00,1.00)	1.00 (1.00,1.00)
RefRefEWAS	0.99 (0.97,1.00)	0.00 (0.00,0.29)	0.99 (0.99,1.00)	0.50 (0.01,1.00)
CellMix	1.00 (0.98, 1.00)	1.00 (1.00, 1.00)	0.98 (0.93, 0.99)	0.10 (0.02, 0.29)
ReFACTor	1.00 (0.98, 1.00)	1.00 (1.00, 1.00)	0.99 (0.79, 1.00)	0.00 (0.00, 0.00)
SVA	1.00 (0.99,1.00)	0.99 (0.97, 1.00)	0.99 (0.99,1.00)	0.99 (0.99, 1.00)

Footnote: **ρ = correlation between primary covariate and latent variables. ρ = 0 corresponds to data simulated from Scenario 1, while ρ=0.7 corresponds to data simulated from Scenario 2. a= FasT-LMM-EWASher, b= RefFreeEWAS, c= RefFreeCellMix, d=Surrogate variable analysis.**

Discussion

We compared eight cell-type correction methods using real and simulated data. Based on DNA methylation in a cohort study, the methods in ReFACTor identified the largest number of CpGs (~60K CpGs), none of which overlapped with the common CpGs detected by other methods including the analysis without adjusting for cell type compositions (but excluding the method in FaST-LMM-EWASher). The method in FaST-LMM-EWASher did not identify any CpG sites. Except for ReFACTor and FaST-LMM-EWASher, at least one detected CpG was shared between all the other methods. More than 50% of CpGs identified by the Houseman et al. method and by the approach implemented in minfi were also detected by the SVA method; The overlap in CpGs was much less between these two methods and the remaining methods. The genes associated with CpGs uniquely identified by using SVA with prenatal urinary arsenic as primary exposure led to the enrichment of GO categories and KEGG pathways that were consistent with our understanding with respect to the effect of arsenic on DNA methylation. Arsenic exposure leads to generation of reactive oxygen species (ROS) which induces DNA damage [121]. This reactive oxygen species play a crucial role in signal transduction pathways, transcription factor regulation [122], and mitogen activated protein kinases (MAPKs) signal transduction pathway is one such pathway that is affected by ROS [123]. DAVID analysis of genes associated with the CpGs uniquely identified by SVA for FaST-LMM-EWASher example dataset led to enrichment of KEGG pathways in cancer. All these imply that the CpGs uniquely

identified by using SVA are potentially informative. Using the example dataset provided by FasT-LMM-EWASher method, we found that all methods except for FasT-LMM-EWASher identified a large number of CpG sites. This was likely due to the platform used to measure DNA methylation levels (Illumina 27K), which is centered more on cancer genes. However, CpGs identified based on ReFACTor and RUV were close to the number of CpGs in the pool of candidate CpGs, indicating possible inflations. On the other hand, results from minfi showed the greatest overlap with the SVA method (Table 4.2). Based on these two real data sets, results from the method in the minfi package and those from SVA were most agreeable. However, for real data, the underlying truth was unknown, which was the motivation of incorporating a data set with cell counts known and the use of a series of simulation studies. Findings from these data were further discussed in this section.

Using the available cell counts in the cancer status and DNA methylation dataset we observed agreements between cell types estimated by Houseman et al. and minfi, but minfi showed a better agreement. The Houseman et al. approach was designed for the Illumina 27K beadchip array, which may not fit the 450K array as noted in the literature [37]. The modification of the Houseman et al. approach implemented in the minfi package, on the other hand, is suitable for both 27K and 450K array. The reference data were from six adult white European males. It has been shown that DNA methylation patterns vary by sex, age and ancestry [124-128]. Generalizing the cell mixtures estimated by minfi to studies with both genders and non-Europeans of different age groups may potentially introduce bias.

Further simulations investigating reference-free methods supported the findings from real data. Regardless of the number of important CpGs, FaST-LMM-EWASher showed the lowest sensitivity, indicating low power to identify truly important CpGs if using that method to adjust

for cell type compositions. ReFACTor produced lowest specificity when confounding effects were present, supporting the rather low overlapping with findings from other methods. On the other hand, findings from ReFACTor, RefFreeEWAS, RefFreeCellMix and SVA were in general comparable for data simulated under scenario 1 (no-confounding effects), but SVA gave consistently higher sensitivity and specificity when cofounding effects present.

The SVA approach does not provide estimates on cell type compositions; however, our ultimate goal was not to estimate cell counts. The goal was to identify an approach that best assesses DNA methylation differentiation due to exposure or diseases, corrected for a potential cell type bias. From this viewpoint and the findings from real data and the high sensitivities and specificities from simulations (under both scenarios, confounding and no confounding), using SVA to adjust for cell type compositions seems to be an appropriate method and may perform better than the existing methods. It is worth noting that information included in the surrogate variables produced by the SVA method may also include other information in addition to cell type compositions. There is a potential of over-adjustment by use of this approach. Furthermore, we would like to point out that all these reference-free methods can be directly applied to genome-wide bisulfite sequencing data and we expect similar findings in terms of their ability in inferring cell type compositions.

Conclusion

When appropriate reference data are available and if inferences on cell type compositions are needed, the method implemented in the minfi package is recommended. However, if no such reference data are available or if the focus is not on estimating cell proportions, the SVA method is suggested to correct for bias resulting from varying cell mixtures.

5 Summary

The work presented in this dissertation is distributed in the ongoing project related to epigenome wide association studies. The epigenetic markers identified in this study will benefit researchers in epigenetic studies in that they will help elucidate the pathophysiology of disease associated with in utero arsenic exposure, and provide insight into the production of immunoglobulin E. The replications of identified CpG sites in independent cohorts strengthen the validity of the findings. I summarize the highlights of the work as follows:

1. The 252 CpG sites identified and replicated in an independent cohort can serve as an epigenetic marker for the adverse health effect of in utero arsenic exposure on the newborn subjects.

2. Of the 252 CpG sites, 5 CpG sites were found to be longitudinally associated with low density lipoprotein measures of the subjects at ages 2, 5, 8, 11 and 14. The genes corresponding to these five CpGs (cg25189764, cg04986899, cg04903360, cg08198265 and cg10473311) also had literature support for their association with cardiovascular disease or diabetes. The DNA methylation measurements at the identified five CpG sites are known to be stable across different age group. Thus, these five CpG sites have the potential to serve as an epigenetic marker for the adverse effect of in-utero arsenic exposure and its association with LDL measures at later ages.

3. The 124 CpG sites longitudinally associated with IgE in the main cohort and replicated in an independent cohort can serve as an epigenetic marker predicting the production of IgE. Of 124 the DNA methylation measures at 59 CpG sites were found to be stable at birth, age 10 and age 18 in IoW cohort. Thus, these 59 CpG sites are more reliable to serve as an epigenetic marker explaining the production of IgE and could eventually help in revealing the pathophysiology of the developmental immune based disease.

4. In the assessment of the methods for cell type adjustment we found that Houseman's algorithm implemented in Bioconductor package "minfi" is the best choice if reference dataset is available. Although R package "sva" performed best compared to other methods in most of the situations, but given the possibility of over fit it is recommended in the situation where reference dataset is not available.

References:

1. Farzan SF, Karagas MR, Chen Y: **In utero and early life arsenic exposure in relation to long-term health and disease**. *Toxicology and applied pharmacology* 2013, **272**(2):384-390.
2. Farzan SF, Li Z, Korrick SA, Spiegelman D, Enelow R, Nadeau K, Baker E, Karagas MR: **Infant Infections and Respiratory Symptoms in Relation to in Utero Arsenic Exposure in a U.S. Cohort**. *Environmental health perspectives* 2016, **124**(6):840-847.
3. Guan H, Piao F, Zhang X, Li X, Li Q, Xu L, Kitamura F, Yokoyama K: **Prenatal exposure to arsenic and its effects on fetal development in the general population of Dalian**. *Biological trace element research* 2012, **149**(1):10-15.
4. O'Sullivan L, Combes AN, Moritz KM: **Epigenetics and developmental programming of adult onset diseases**. *Pediatric nephrology* 2012, **27**(12):2175-2182.
5. Youngson NA, Whitelaw E: **Transgenerational epigenetic effects**. *Annual review of genomics and human genetics* 2008, **9**:233-257.
6. Chong S, Whitelaw E: **Epigenetic germline inheritance**. *Current opinion in genetics & development* 2004, **14**(6):692-696.
7. Egger G, Liang G, Aparicio A, Jones PA: **Epigenetics in human disease and prospects for epigenetic therapy**. *Nature* 2004, **429**(6990):457-463.
8. Holliday R, Pugh JE: **DNA modification mechanisms and gene activity during development**. *Science* 1975, **187**(4173):226-232.
9. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites**. *Trends in genetics : TIG* 1997, **13**(8):335-340.
10. Bird A: **DNA methylation patterns and epigenetic memory**. *Genes & development* 2002, **16**(1):6-21.
11. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R: **Prospects for epigenetic epidemiology**. *American journal of epidemiology* 2009, **169**(4):389-400.
12. Wang SC, Oelze B, Schumacher A: **Age-specific epigenetic drift in late-onset Alzheimer's disease**. *PloS one* 2008, **3**(7):e2698.
13. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL: **Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome**. *Environ Health Perspect* 2006, **114**(4):567-572.
14. Dietert RR, Dietert JM: **Potential for early-life immune insult including developmental immunotoxicity in autism and autism spectrum disorders: focus on critical windows of immune vulnerability**. *Journal of toxicology and environmental health Part B, Critical reviews* 2008, **11**(8):660-680.
15. Bailey KA, Wu MC, Ward WO, Smeester L, Rager JE, Garcia-Vargas G, Del Razo LM, Drobna Z, Styblo M, Fry RC: **Arsenic and the epigenome: interindividual differences in arsenic metabolism related to distinct patterns of DNA methylation**. *Journal of biochemical and molecular toxicology* 2013, **27**(2):106-115.
16. Broberg K, Ahmed S, Engstrom K, Hossain MB, Jurkovic Mlakar S, Bottai M, Grander M, Raqib R, Vahter M: **Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys**. *Journal of developmental origins of health and disease* 2014, **5**(4):288-298.
17. **A review of human carcinogens. Part C: Arsenic, metals, fibres, and dusts / IARC Working Group on the Evaluation of Carcinogenic Risks to Humans**. *International Agency for Research on Cancer* 2012, **100C**.
18. Vahter M: **Effects of arsenic on maternal and fetal health**. *Annual review of nutrition* 2009, **29**:381-399.

19. Nordstrom DK: **Public health. Worldwide occurrences of arsenic in ground water**. *Science* 2002, **296**(5576):2143-2145.
20. Smith AH, Marshall G, Liaw J, Yuan Y, Ferreccio C, Steinmaus C: **Mortality in young adults following in utero and childhood exposure to arsenic in drinking water**. *Environmental health perspectives* 2012, **120**(11):1527-1531.
21. Chou WC, Chung YT, Chen HY, Wang CJ, Ying TH, Chuang CY, Tseng YC, Wang SL: **Maternal arsenic exposure and DNA damage biomarkers, and the associations with birth outcomes in a general population from Taiwan**. *PloS one* 2014, **9**(2):e86398.
22. Rosenberg HG: **Systemic arterial disease and chronic arsenicism in infants**. *Archives of pathology* 1974, **97**(6):360-365.
23. Hawkesworth S, Wagatsuma Y, Kippler M, Fulford AJ, Arifeen SE, Persson LA, Moore SE, Vahter M: **Early exposure to toxic metals has a limited effect on blood pressure or kidney function in later childhood, rural Bangladesh**. *International journal of epidemiology* 2013, **42**(1):176-185.
24. Yuan Y, Marshall G, Ferreccio C, Steinmaus C, Selvin S, Liaw J, Bates MN, Smith AH: **Acute myocardial infarction mortality in comparison with lung and bladder cancer mortality in arsenic-exposed region II of Chile from 1950 to 2000**. *Am J Epidemiol* 2007, **166**(12):1381-1391.
25. Davila-Esqueda ME, Morales JM, Jimenez-Capdeville ME, De la Cruz E, Falcon-Escobedo R, Chi-Ahumada E, Martin-Perez S: **Low-level subchronic arsenic exposure from prenatal developmental stages to adult life results in an impaired glucose homeostasis**. *Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association* 2011, **119**(10):613-617.
26. Rossman TG, Klein CB: **Genetic and epigenetic effects of environmental arsenicals**. *Metallomics : integrated biometal science* 2011, **3**(11):1135-1141.
27. Gluckman PD: **Epigenetics and metabolism in 2011: Epigenetics, the life-course and metabolic disease**. *Nature reviews Endocrinology* 2012, **8**(2):74-76.
28. Vickers MH: **Early life nutrition, epigenetics and programming of later life disease**. *Nutrients* 2014, **6**(6):2165-2178.
29. Majumdar S, Chanda S, Ganguli B, Mazumder DN, Lahiri S, Dasgupta UB: **Arsenic exposure induces genomic hypermethylation**. *Environmental toxicology* 2010, **25**(3):315-318.
30. Smeester L, Rager JE, Bailey KA, Guan X, Smith N, Garcia-Vargas G, Del Razo LM, Drobna Z, Kelkar H, Styblo M *et al*: **Epigenetic changes in individuals with arsenicosis**. *Chemical research in toxicology* 2011, **24**(2):165-167.
31. Xie Y, Liu J, Benbrahim-Tallaa L, Ward JM, Logsdon D, Diwan BA, Waalkes MP: **Aberrant DNA methylation and gene expression in livers of newborn mice transplacentally exposed to a hepatocarcinogenic dose of inorganic arsenic**. *Toxicology* 2007, **236**(1-2):7-15.
32. Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ: **Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero**. *Environmental health perspectives* 2013, **121**(8):971-977.
33. Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Cardenas A, Wright RO, Christiani DC: **Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood**. *Epigenetics* 2014, **9**(5):774-782.
34. Maksimovic J, Gordon L, Oshlack A: **SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips**. *Genome biology* 2012, **13**(6):R44.
35. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J: **Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility**. *PloS one* 2012, **7**(7):e41361.

36. Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, Wiencke JK, Houseman EA: **Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis**. *Epigenetics* 2013, **8**(8):816-826.
37. Jaffe AE, Irizarry RA: **Accounting for cellular heterogeneity is critical in epigenome-wide association studies**. *Genome biology* 2014, **15**(2):R31.
38. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT: **DNA methylation arrays as surrogate measures of cell mixture distribution**. *BMC bioinformatics* 2012, **13**:86.
39. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA: **The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists**. *Genome biology* 2007, **8**(9):R183.
40. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 2000, **28**(1):27-30.
41. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT *et al*: **The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function**. *Nucleic acids research* 2010, **38**(Web Server issue):W214-220.
42. Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S: **Bioinformatics and computational biology solutions using R and Bioconductor**, vol. 746718470: Springer; 2005.
43. Smyth GK, Yang YH, Speed T: **Statistical issues in cDNA microarray data analysis**. *Methods Mol Biol* 2003, **224**:111-136.
44. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal Statistical Society Series B (Methodological)* 1995:289-300.
45. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nature protocols* 2009, **4**(1):44-57.
46. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic acids research* 2009, **37**(1):1-13.
47. Chan KH, Huang YT, Meng Q, Wu C, Reiner A, Sobel EM, Tinker L, Lusis AJ, Yang X, Liu S: **Shared molecular pathways and gene networks for cardiovascular disease and type 2 diabetes mellitus in women across diverse ethnicities**. *Circ Cardiovasc Genet* 2014, **7**(6):911-919.
48. Gowd V, Gurukar A, Chilkunda ND: **Glycosaminoglycan remodeling during diabetes and the role of dietary factors in their modulation**. *World J Diabetes* 2016, **7**(4):67-73.
49. Wang SL, Chiou JM, Chen CJ, Tseng CH, Chou WL, Wang CC, Wu TN, Chang LW: **Prevalence of non-insulin-dependent diabetes mellitus and related vascular diseases in southwestern arseniasis-endemic and nonendemic areas in Taiwan**. *Environmental health perspectives* 2003, **111**(2):155-159.
50. Gribble MO, Howard BV, Umans JG, Shara NM, Francesconi KA, Goessler W, Crainiceanu CM, Silbergeld EK, Guallar E, Navas-Acien A: **Arsenic exposure, diabetes prevalence, and diabetes control in the Strong Heart Study**. *American journal of epidemiology* 2012, **176**(10):865-874.
51. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, Woodward G, Lyttleton O, Evans DM, Reik W *et al*: **Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)**. *Int J Epidemiol* 2015, **44**(4):1181-1190.
52. Luo J, Shu W: **Arsenic-Induced Developmental Neurotoxicity**. *Handbook of Arsenic Toxicology* 2014:363.
53. Gong G, O'Bryant SE: **The arsenic exposure hypothesis for Alzheimer disease**. *Alzheimer disease and associated disorders* 2010, **24**(4):311-316.

54. Vahidnia A, Romijn F, van der Voet GB, de Wolff FA: **Arsenic-induced neurotoxicity in relation to toxicokinetics: effects on sciatic nerve proteins**. *Chemico-biological interactions* 2008, **176**(2-3):188-195.
55. Lemarie A, Morzadec C, Bourdonnay E, Fardel O, Vernhet L: **Human macrophages constitute targets for immunotoxic inorganic arsenic**. *Journal of immunology* 2006, **177**(5):3019-3027.
56. Hsu WL, Tsai MH, Lin MW, Chiu YC, Lu JH, Chang CH, Yu HS, Yoshioka T: **Differential effects of arsenic on calcium signaling in primary keratinocytes and malignant (HSC-1) cells**. *Cell calcium* 2012, **52**(2):161-169.
57. Perera F, Herbstman J: **Prenatal environmental exposures, epigenetics, and disease**. *Reproductive toxicology* 2011, **31**(3):363-373.
58. Skinner MK: **Role of epigenetics in developmental biology and transgenerational inheritance**. *Birth Defects Research Part C: Embryo Today: Reviews* 2011, **93**(1):51-55.
59. Skinner MK: **Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability**. *Epigenetics* 2011, **6**(7):838-842.
60. Lee TW, Kwon H, Zong H, Yamada E, Vatish M, Pessin JE, Bastie CC: **Fyn deficiency promotes a preferential increase in subcutaneous adipose tissue mass and decreased visceral adipose tissue inflammation**. *Diabetes* 2013, **62**(5):1537-1546.
61. Kajimoto Y, Miyagawa J, Ishihara K, Okuyama Y, Fujitani Y, Itoh M, Yoshida H, Kaisho T, Matsuoka T, Watada H *et al*: **Pancreatic islet cells express BST-1, a CD38-like surface molecule having ADP-ribosyl cyclase activity**. *Biochemical and biophysical research communications* 1996, **219**(3):941-946.
62. Paulick MG, Bertozzi CR: **The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins**. *Biochemistry* 2008, **47**(27):6991-7000.
63. Pedersen LC, Tsuchida K, Kitagawa H, Sugahara K, Darden TA, Negishi M: **Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I**. *The Journal of biological chemistry* 2000, **275**(44):34580-34585.
64. Kreuger J, Kjellen L: **Heparan sulfate biosynthesis: regulation and variability**. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2012, **60**(12):898-907.
65. Grande-Allen KJ, Osman N, Ballinger ML, Dadlani H, Marasco S, Little PJ: **Glycosaminoglycan synthesis and structure as targets for the prevention of calcific aortic valve disease**. *Cardiovascular research* 2007, **76**(1):19-28.
66. Ballinger ML, Nigro J, Frontanilla KV, Dart AM, Little PJ: **Regulation of glycosaminoglycan structure and atherogenesis**. *Cellular and molecular life sciences : CMLS* 2004, **61**(11):1296-1306.
67. Schmidli RS, Colman PG, Cui L, Yu WP, Kewming K, Jankulovski C, Harrison LC, Pallen CJ, DeAizpurua HJ: **Antibodies to the protein tyrosine phosphatases IAR and IA-2 are associated with progression to insulin-dependent diabetes (IDDM) in first-degree relatives at-risk for IDDM**. *Autoimmunity* 1998, **28**(1):15-23.
68. Below JE, Gamazon ER, Morrison JV, Konkashbaev A, Pluzhnikov A, McKeigue PM, Parra EJ, Elbein SC, Hallman DM, Nicolae DL *et al*: **Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals**. *Diabetologia* 2011, **54**(8):2047-2055.
69. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D: **Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome**. *Nature genetics* 2007, **39**(4):457-466.

70. Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y: **A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues**. *PLoS genetics* 2011, **7**(2):e1001316.
71. Pawankar R, Canonica GW, Holgate ST, Lockey RF, Organization WH: **White book on allergy 2011-2012 executive summary**. *World Allergy Organization* 2011.
72. Tezza G, Mazzei F, Boner A: **Epigenetics of allergy**. *Early human development* 2013, **89**:S20-S21.
73. Greer JM, McCombe PA: **The role of epigenetic mechanisms and processes in autoimmune disorders**. *Biologics : targets & therapy* 2012, **6**:307-327.
74. Javierre BM, Hernando H, Ballestar E: **Environmental triggers and epigenetic deregulation in autoimmune disease**. *Discovery medicine* 2011, **12**(67):535-545.
75. Liang L, Willis-Owen SAG, Laprise C, Wong KCC, Davies GA, Hudson TJ: **An epigenome-wide association study of total serum immunoglobulin E concentration**. *Nature* 2015, **520**.
76. Everson TM, Lyons G, Zhang H, Soto-Ramirez N, Lockett GA, Patil VK, Merid SK, Soderhall C, Melen E, Holloway JW *et al*: **DNA methylation loci associated with atopy and high serum IgE: a genome-wide application of recursive Random Forest feature selection**. *Genome Med* 2015, **7**:89.
77. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S: **IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data**. *Bioinformatics (Oxford, England)* 2012, **28**(5):729-730.
78. Lin L-C, Wang S-L, Chang Y-C, Huang P-C, Cheng J-T, Su P-H, Liao P-C: **Associations between maternal phthalate exposure and cord sex hormones in human infants**. *Chemosphere* 2011, **83**(8):1192-1199.
79. Wang S-L, Su P-H, Jong S-B, Guo YL, Chou W-L, Päpke O: **In utero exposure to dioxins and polychlorinated biphenyls and its relations to thyroid function and growth hormone in newborns**. *Environmental health perspectives* 2005:1645-1650.
80. Zhang TT, Okkenhaug K, Nashed BF, Puri KD, Knight ZA, Shokat KM, Vanhaesebroeck B, Marshall AJ: **Genetic or pharmaceutical blockade of p110delta phosphoinositide 3-kinase enhances IgE production**. *J Allergy Clin Immunol* 2008, **122**(4):811-819 e812.
81. Fu SL, Pierre J, Smith-Norowitz TA, Hagler M, Bowne W, Pincus MR, Mueller CM, Zenilman ME, Bluth MH: **Immunoglobulin E antibodies from pancreatic cancer patients mediate antibody-dependent cell-mediated cytotoxicity against pancreatic cancer cells**. *Clin Exp Immunol* 2008, **153**(3):401-409.
82. Jensen-Jarolim E, Achatz G, Turner MC, Karagiannis S, Legrand F, Capron M, Penichet ML, Rodriguez JA, Siccardi AG, Vangelista L *et al*: **AllergoOncology: the role of IgE-mediated allergy in cancer**. *Allergy* 2008, **63**(10):1255-1266.
83. Matta GM, Battaglio S, Dibello C, Napoli P, Baldi C, Ciccone G, Coscia M, Boccadoro M, Massaia M: **Polyclonal immunoglobulin E levels are correlated with hemoglobin values and overall survival in patients with multiple myeloma**. *Clin Cancer Res* 2007, **13**(18 Pt 1):5348-5354.
84. Nabih ES, Kamel HF, Kamel TB: **Association Between CD14 Polymorphism (-1145G/A) and Childhood Bronchial Asthma**. *Biochem Genet* 2016, **54**(1):50-60.
85. March ME, Sleiman PM, Hakonarson H: **Genetic polymorphisms and associated susceptibility to asthma**. *Int J Gen Med* 2013, **6**:253-265.
86. DeWan AT, Triche EW, Xu X, Hsu LI, Zhao C, Belanger K, Hellenbrand K, Willis-Owen SA, Moffatt M, Cookson WO *et al*: **PDE11A associations with asthma: results of a genome-wide association scan**. *J Allergy Clin Immunol* 2010, **126**(4):871-873 e879.
87. Zhang J, Noguchi E, Migita O, Yokouchi Y, Nakayama J, Shibasaki M, Arinami T: **Association of a haplotype block spanning SDAD1 gene and CXC chemokine genes with allergic rhinitis**. *J Allergy Clin Immunol* 2005, **115**(3):548-554.

88. Albrandt K, Orida NK, Liu FT: **An IgE-binding protein with a distinctive repetitive sequence and homology with an IgG receptor**. *Proc Natl Acad Sci U S A* 1987, **84**(19):6859-6863.
89. Konno S, Takahashi D, Hizawa N, Hattori T, Takahashi A, Isada A, Maeda Y, Huang SK, Nishimura M: **Genetic impact of a butyrophilin-like 2 (BTNL2) gene variation on specific IgE responsiveness to Dermatophagoides farinae (Der f) in Japanese**. *Allergol Int* 2009, **58**(1):29-35.
90. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, Gudnason V: **Heterogeneity in white blood cells has potential to confound DNA methylation measurements**. *PloS one* 2012, **7**(10):e46705.
91. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, Putter H, Slagboom PE, Heijmans BT: **Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2010, **24**(9):3135-3144.
92. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium(R) assay**. *Epigenomics* 2009, **1**(1):177-200.
93. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome Biol* 2004, **5**(10):R80.
94. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays**. *Bioinformatics (Oxford, England)* 2014, **30**(10):1363-1369.
95. Houseman EA, Molitor J, Marsit CJ: **Reference-free cell mixture adjustments in analysis of DNA methylation data**. *Bioinformatics (Oxford, England)* 2014, **30**(10):1431-1439.
96. Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ: **Reference-free deconvolution of DNA methylation data and mediation by cell composition effects**. *BMC Bioinformatics* 2016, **17**:259.
97. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis**. *PLoS genetics* 2007, **3**(9):1724-1735.
98. Rahmani E, Zaitlen N, Baran Y, Eng C, Hu D, Galanter J, Oh S, Burchard EG, Eskin E, Zou J *et al*: **Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies**. *Nat Methods* 2016, **13**(5):443-445.
99. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM: **Measurement error in nonlinear models: a modern perspective**: CRC press; 2012.
100. Gagnon-Bartsch JA, Speed TP: **Using control genes to correct for unwanted variation in microarray data**. *Biostatistics* 2012, **13**(3):539-552.
101. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J: **Epigenome-wide association studies without the need for cell-type composition**. *Nature methods* 2014, **11**(3):309-311.
102. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D: **FaST linear mixed models for genome-wide association studies**. *Nature methods* 2011, **8**(10):833-835.
103. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis**. *PLoS genetics* 2007, **3**(9):e161.
104. Zou JY: **Correcting for Sample Heterogeneity in Methylome-Wide Association Studies**. *Methods in molecular biology* 2015.
105. Smith AK, Conneely KN, Pace TW, Mister D, Felger JC, Kilaru V, Akel MJ, Vertino PM, Miller AH, Torres MA: **Epigenetic changes associated with inflammation in breast cancer patients treated with chemotherapy**. *Brain, behavior, and immunity* 2014, **38**:227-236.
106. Jaccard P: **Distribution de la Flore Alpine: dans le Bassin des dranses et dans quelques régions voisines**: Rouge; 1901.

107. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nature genetics* 2000, **25**(1):25-29.
108. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG**. *Nucleic acids research* 2014, **42**(Database issue):D199-205.
109. Liu S, Guo X, Wu B, Yu H, Zhang X, Li M: **Arsenic induces diabetic effects through beta-cell dysfunction and increased gluconeogenesis in mice**. *Scientific reports* 2014, **4**:6894.
110. Guo L, Xiao Y, Wang Y: **Monomethylarsonous acid inhibited endogenous cholesterol biosynthesis in human skin fibroblasts**. *Toxicology and applied pharmacology* 2014, **277**(1):21-29.
111. Small EM, Frost RJ, Olson EN: **MicroRNAs add a new dimension to cardiovascular disease**. *Circulation* 2010, **121**(8):1022-1032.
112. Elinder F, Mannikko R, Pandey S, Larsson HP: **Mode shifts in the voltage gating of the mouse and human HCN2 and HCN4 channels**. *The Journal of physiology* 2006, **575**(Pt 2):417-431.
113. Mumford JL, Wu K, Xia Y, Kwok R, Yang Z, Foster J, Sanders WE: **Chronic arsenic exposure and cardiac repolarization abnormalities with QT interval prolongation in a population-based study**. *Environmental health perspectives* 2007, **115**(5):690-694.
114. Hernandez-Castro B, Doniz-Padilla LM, Salgado-Bustamante M, Rocha D, Ortiz-Perez MD, Jimenez-Capdeville ME, Portales-Perez DP, Quintanar-Stephano A, Gonzalez-Amaro R: **Effect of arsenic on regulatory T cells**. *Journal of clinical immunology* 2009, **29**(4):461-469.
115. Biswas D, Banerjee M, Sen G, Das JK, Banerjee A, Sau TJ, Pandit S, Giri AK, Biswas T: **Mechanism of erythrocyte death in human population exposed to arsenic through drinking water**. *Toxicology and applied pharmacology* 2008, **230**(1):57-66.
116. Andrew AS, Jewell DA, Mason RA, Whitfield ML, Moore JH, Karagas MR: **Drinking-water arsenic exposure modulates gene expression in human lymphocytes from a U.S. population**. *Environ Health Perspect* 2008, **116**(4):524-531.
117. Soto-Pena GA, Luna AL, Acosta-Saavedra L, Conde P, Lopez-Carrillo L, Cebrian ME, Bastida M, Calderon-Aranda ES, Vega L: **Assessment of lymphocyte subpopulations and cytokine secretion in children exposed to arsenic**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2006, **20**(6):779-781.
118. Bethge N, Lothe RA, Honne H, Andresen K, Troen G, Eknaes M, Liestol K, Holte H, Delabie J, Smeland EB *et al*: **Colorectal cancer DNA methylation marker panel validated with high performance in Non-Hodgkin lymphoma**. *Epigenetics : official journal of the DNA Methylation Society* 2014, **9**(3):428-436.
119. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N *et al*: **The consensus coding sequences of human breast and colorectal cancers**. *Science* 2006, **314**(5797):268-274.
120. Blot-Chabaud M, Wanstok F, Bonvalet JP, Farman N: **Cell sodium-induced recruitment of Na(+)-K(+)-ATPase pumps in rabbit cortical collecting tubules is aldosterone-dependent**. *The Journal of biological chemistry* 1990, **265**(20):11676-11681.
121. Li D, Morimoto K, Takeshita T, Lu Y: **Arsenic induces DNA damage via reactive oxygen species in human cells**. *Environmental health and preventive medicine* 2001, **6**(1):27-32.
122. Martindale JL, Holbrook NJ: **Cellular response to oxidative stress: signaling for suicide and survival**. *Journal of cellular physiology* 2002, **192**(1):1-15.
123. Son Y, Kim S, Chung HT, Pae HO: **Reactive oxygen species in the activation of MAP kinases**. *Methods in enzymology* 2013, **528**:27-48.

124. El-Maarri O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, Wienker T, Oldenburg J: **Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males**. *Human genetics* 2007, **122**(5):505-514.
125. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA: **The relationship of DNA methylation with age, gender and genotype in twins and healthy controls**. *PloS one* 2009, **4**(8):e6767.
126. Teschendorff AE, West J, Beck S: **Age-associated epigenetic drift: implications, and a case of epigenetic thrift?** *Human molecular genetics* 2013, **22**(R1):R7-R15.
127. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, Klengel T, Mehta D, Binder EB, Epstein MP *et al*: **Accounting for population stratification in DNA methylation studies**. *Genetic epidemiology* 2014, **38**(3):231-241.
128. Fraser HB, Lam LL, Neumann SM, Kobor MS: **Population-specificity of human DNA methylation**. *Genome biology* 2012, **13**(2):R8.

Appendices

Table A1.1

	TAIWAN COHORT			NHBCS			CpG Information				
Name	**Coef (CI)**	**P-value**	**FDR P-value**	**Coef (CI)**	**P-value**	**FDR P-value**	**Gene**	**MAPINFO**	**CHR**	**Relation to**	**CpG Islands Name**
cg19407236	-0.66 (-0.84, -0.49)	1.44E-09	0	-0.07(-0.33,0.2)	6.13E-01	0.992	RPSAP58	23946205	19	S_Shore	chr19:23945713-23946128
cg17923401	-0.84 (-1.07, -0.62)	1.62E-09	0	-0.16(-0.46,0.13)	2.78E-01	0.967	SIGLEC7;SIGLEC7	51644168	19		
cg06515634	-0.54 (-0.68, -0.39)	2.04E-09	0	-0.04(-0.14,0.06)	4.57E-01	0.988		1.51E+08	1	Island	chr1:151445871-151446142
cg03403996	1.55 (1.12, 1.97)	3.14E-09	0	0.66(-0.42,1.74)	2.30E-01	0.967	TRIM31	30074162	6	S_Shelf	chr6:30071225-30071428
cg07600884	-0.55 (-0.71, -0.4)	5.75E-09	0	0.02(-0.16,0.2)	8.16E-01	0.992	LRRC8D;LRRC8D	90286190	1	Island	chr1:90286189-90287222
cg26241877	0.54 (0.38, 0.69)	8.78E-09	0.001	0.13(-0.04,0.29)	1.39E-01	0.967		52630003	18	S_Shelf	chr18:52626517-52626849
cg16587616	-0.52 (-0.68, -0.37)	1.81E-08	0.001	-0.07(-0.3,0.15)	5.25E-01	0.989	KHDRBS2;KHDRBS2	62996022	6	Island	chr6:62995855-62996228
cg10197432	-0.55 (-0.71, -0.39)	1.86E-08	0.001	0.16(0,0.31)	4.66E-02	0.776	C6orf103	1.47E+08	6	Island	chr6:146920034-146920245
cg03462714	0.7 (0.49, 0.91)	2.65E-08	0.001	.			ADARB2	1547894	10		
cg06889951	0.34 (0.23, 0.44)	5.43E-08	0.002	0(-0.15,0.15)	9.87E-01	0.997	FGF3	69625218	11	Island	chr11:69625108-69625473
cg26284982	-0.58 (-0.77, -0.4)	6.13E-08	0.002	0(-0.1,0.1)	9.89E-01	0.997	C17orf57	45401312	17	Island	chr17:45400874-45401440
cg27256528	-0.4 (-0.52, -0.27)	9.62E-08	0.003	0.1(-0.04,0.24)	1.49E-01	0.967	A2BP1;A2BP1	6533667	16	Island	chr16:6533019-6533983
cg25249539	-0.59 (-0.77, -0.4)	1.07E-07	0.003	-0.06(-0.41,0.28)	7.14E-01	0.992		56803358	20	Island	chr20:56803282-56804112
cg27367142	-0.6 (-0.8, -0.41)	1.13E-07	0.003	-0.03(-0.31,0.25)	8.12E-01	0.992		77127199	10		
cg13011003	1.26 (0.85, 1.67)	1.63E-07	0.004	-0.83(-1.36,-0.29)	2.67E-03	0.493		1.26E+08	12		
cg20677901	-0.64 (-0.85, -0.43)	1.91E-07	0.004	0.19(0.02,0.36)	3.13E-02	0.719	TP73	3568210	1	Island	chr1:3566445-3569636
cg11801110	0.38 (0.26, 0.51)	1.96E-07	0.004	-0.01(-0.1,0.08)	8.38E-01	0.992	PIM3	50356763	22	Island	chr22:50353596-50357215
cg14232870	0.61 (0.4, 0.81)	2.90E-07	0.006	0.05(-0.1,0.2)	5.16E-01	0.989	SHANK2;SHANK2	70458782	11		
cg00345443	-0.45 (-0.6, -0.3)	3.09E-07	0.006	0.15(-0.09,0.4)	2.23E-01	0.967		54121778	3	Island	chr3:54121624-54122147
cg01615818	-1.27 (-1.69, -0.84)	3.19E-07	0.006	0.02(-0.51,0.55)	9.38E-01	0.992	ZNF681	23941438	19	Island	chr19:23941240-23941538
cg03865648	-0.46 (-0.61, -0.3)	4.07E-07	0.007	.				1.73E+08	3	N_Shore	chr3:173115404-173115775
cg00768741	0.37 (0.24, 0.49)	4.46E-07	0.008	-0.02(-0.13,0.1)	7.77E-01	0.992	HSPG2	22210791	1	N_Shore	chr1:22210992-22211318
cg26278987	0.78 (0.52, 1.05)	4.56E-07	0.008	-0.03(-0.81,0.75)	9.40E-01	0.992	C4orf50	5960351	4		
cg19846314	-0.56 (-0.76, -0.37)	4.84E-07	0.008	0.16(-0.15,0.47)	3.13E-01	0.967	GAD1;GAD1	1.72E+08	2	Island	chr2:171678546-171680358
cg00630892	0.74 (0.49, 1)	5.08E-07	0.008	-0.07(-0.17,0.04)	2.18E-01	0.967	NRBP2	1.45E+08	8	Island	chr8:144921644-144921914
cg13496244	0.38 (0.25, 0.51)	6.43E-07	0.009	0.1(-0.1,0.3)	3.14E-01	0.967		4925830	7	S_Shelf	chr7:4922707-4923576
cg00947796	0.36 (0.23, 0.48)	6.63E-07	0.009	-0.02(-0.17,0.12)	7.55E-01	0.992		1316185	16		
cg06946213	0.32 (0.2, 0.43)	6.95E-07	0.01	0.02(-0.12,0.16)	7.74E-01	0.992		1.03E+08	6		
cg08522676	-0.32 (-0.43, -0.21)	7.31E-07	0.01	0.08(-0.03,0.2)	1.69E-01	0.967		54343702	12	Island	chr12:54343622-54343848
cg07583091	-0.48 (-0.65, -0.31)	7.83E-07	0.01	0.04(-0.08,0.17)	5.00E-01	0.989	GTF2H1;HPS5;HPS5;GT	18343626	11	Island	chr11:18343625-18344238
cg25137787	0.73 (0.47, 0.99)	8.05E-07	0.01	-0.03(-0.3,0.25)	8.52E-01	0.992		2863815	1		
cg13718729	-0.53 (-0.72, -0.34)	8.85E-07	0.01	0.12(-0.16,0.39)	4.03E-01	0.968	GRIN1;GRIN1;GRIN1	1.40E+08	9	Island	chr9:140056283-140057837
cg06526522	0.43 (0.28, 0.58)	9.01E-07	0.01	0.01(-0.09,0.12)	8.36E-01	0.992	PRDM16;PRDM16	3129005	1		
cg09419102	0.35 (0.22, 0.47)	9.01E-07	0.01	0.04(-0.02,0.1)	2.41E-01	0.967		65550444	11	S_Shore	chr11:65547499-65549261
cg11569407	0.37 (0.24, 0.5)	9.12E-07	0.01	.				1.30E+08	10	Island	chr10:130008558-130009620
cg06458094	0.32 (0.2, 0.43)	9.39E-07	0.01	0.03(-0.09,0.16)	5.83E-01	0.992		20448558	17		
cg04069951	-0.37 (-0.5, -0.24)	9.46E-07	0.01	-0.17(-0.44,0.11)	2.28E-01	0.967	CD81	2398336	11	Island	chr11:2398223-2399598
cg07969609	0.44 (0.29, 0.6)	9.58E-07	0.01	.				98438319	8		
cg20646995	-0.39 (-0.53, -0.25)	1.03E-06	0.01	0.01(-0.19,0.22)	9.01E-01	0.992	LRWD1;ALKBH4	1.02E+08	7	Island	chr7:102105123-102105782
cg17032590	0.49 (0.31, 0.66)	1.09E-06	0.011	0.12(-0.03,0.27)	1.04E-01	0.967	SLC2A7	9074847	1		
cg06043190	0.52 (0.33, 0.71)	1.14E-06	0.011	-0.01(-0.12,0.1)	8.84E-01	0.992	EML4;EML4	42396882	2	Island	chr2:42396301-42396933
cg20263686	0.32 (0.21, 0.44)	1.21E-06	0.011	-0.03(-0.13,0.08)	6.12E-01	0.992	ANK1;ANK1;ANK1;AN	41643429	8		
cg03430846	-0.43 (-0.59, -0.28)	1.28E-06	0.011	.			NRG1	31497082	8	Island	chr8:31496525-31498346
cg13324337	-0.55 (-0.75, -0.35)	1.30E-06	0.011	-0.01(-0.15,0.13)	8.85E-01	0.992		1.00E+08	15		
cg04986899*	-0.42 (-0.58, -0.27)	1.37E-06	0.011	-0.02(-0.15,0.12)	8.22E-01	0.992	XYLT1	17553784	16		
cg15530351	0.77 (0.49, 1.06)	1.40E-06	0.011	-0.07(-0.23,0.09)	3.83E-01	0.968	ANGPT2;ANGPT2;ANG	6407676	8		
cg18226049	0.38 (0.24, 0.52)	1.41E-06	0.011	-0.03(-0.15,0.1)	6.96E-01	0.992	TSSC4	2424791	11	S_Shelf	chr11:2421040-2422474
cg21696208	-0.48 (-0.65, -0.3)	1.44E-06	0.011	-0.03(-0.22,0.17)	7.97E-01	0.992	MTFR1;MTFR1	66556075	8	N_Shore	chr8:66556764-66557513
cg24288527	-0.45 (-0.62, -0.29)	1.44E-06	0.011	0.06(-0.06,0.18)	3.54E-01	0.968	PPP1R3C	93390347	10	N_Shelf	chr10:93392667-93393147
cg20063462	-0.61 (-0.84, -0.39)	1.49E-06	0.011	0.02(-0.16,0.21)	8.03E-01	0.992	TMEM87B	1.13E+08	2	N_Shore	chr2:112812827-112813614
cg04293085	-0.37 (-0.5, -0.23)	1.56E-06	0.012	0.12(0,0.24)	5.98E-02	0.843	ADRA1D	4202378	20	Island	chr20:4202148-4202765
cg14449633	0.71 (0.45, 0.97)	1.58E-06	0.012	0.1(-0.12,0.31)	3.68E-01	0.968		51532432	10		
cg17841267	0.36 (0.23, 0.5)	1.63E-06	0.012	-0.08(-0.24,0.07)	2.92E-01	0.967		1.12E+08	10		
cg01238672	0.4 (0.25, 0.54)	1.65E-06	0.012	0.09(-0.06,0.25)	2.41E-01	0.967		1.70E+08	6		
cg04438098	0.81 (0.51, 1.11)	1.70E-06	0.012	0.58(0.22,0.95)	1.79E-03	0.493		1.12E+08	13	S_Shore	chr13:112212062-112212297
cg12439423	0.34 (0.21, 0.46)	1.77E-06	0.012	0.04(-0.06,0.15)	4.19E-01	0.968	ANK1;ANK1;ANK1;AN	41522721	8		
cg07243405	0.35 (0.22, 0.47)	1.81E-06	0.012	0.07(-0.05,0.19)	2.52E-01	0.967	C7orf20	915203	7	N_Shore	chr7:915753-916644
cg23217126	-0.49 (-0.67, -0.31)	1.90E-06	0.013	0.16(0.01,0.31)	3.77E-02	0.719	DOK6	67068268	18	Island	chr18:67067509-67069168
cg21515956	0.45 (0.28, 0.61)	1.93E-06	0.013	0.03(-0.06,0.12)	5.26E-01	0.989	KIF26A	1.05E+08	14	S_Shelf	chr14:104601959-104605672
cg17696468	-0.35 (-0.48, -0.22)	1.95E-06	0.013	-0.1(-0.23,0.03)	1.19E-01	0.967	ELP4	31742528	11		
cg06970076	0.5 (0.32, 0.69)	2.03E-06	0.013	-0.07(-0.32,0.17)	5.55E-01	0.99	IFT140	1560791	16	S_Shore	chr16:1559965-1560194
cg15268244	0.34 (0.21, 0.46)	2.15E-06	0.013	0.19(0.04,0.34)	1.26E-02	0.719		77196840	15	N_Shore	chr15:77197258-77197895
cg15311933	-0.62 (-0.85, -0.39)	2.24E-06	0.014	-0.01(-0.17,0.14)	8.62E-01	0.992	PQLC2;AKR7A2;PQLC2	19638897	1	Island	chr1:19638207-19639309
cg06419180*	-0.39 (-0.54, -0.25)	2.29E-06	0.014	-0.06(-0.15,0.04)	2.38E-01	0.967	CSGALNACT1;CSGALI	19262414	8		
cg26297950	-0.35 (-0.47, -0.22)	2.39E-06	0.014	0.03(-0.07,0.13)	6.20E-01	0.992	FLCN;FLCN	17124828	17	Island	chr17:17124698-17124931
cg09115646	-0.76 (-1.04, -0.47)	2.43E-06	0.014	-0.11(-0.26,0.05)	1.84E-01	0.967		2978687	10		
cg26789064	0.52 (0.32, 0.71)	2.43E-06	0.014	0.1(-0.09,0.28)	2.93E-01	0.967		4866130	4	N_Shore	chr4:4866438-4866813
cg20074159	0.59 (0.37, 0.81)	2.45E-06	0.014	0.16(-0.07,0.4)	1.69E-01	0.967		1.10E+08	3		
cg06962275	-0.49 (-0.68, -0.31)	2.52E-06	0.014	0.01(-0.11,0.12)	9.09E-01	0.992	NCRNA00171	30010203	6		
cg01616682	-0.41 (-0.56, -0.26)	2.55E-06	0.014	0.13(-0.09,0.34)	2.39E-01	0.967	CALCB	15095017	11	Island	chr11:15094957-15095872
cg12266551	0.46 (0.28, 0.63)	2.77E-06	0.015	-0.04(-0.15,0.08)	5.25E-01	0.989	TOMM40;TOMM40;TO	45394624	19	Island	chr19:45393833-45394992
cg12417362	-0.48 (-0.66, -0.3)	2.77E-06	0.015	0.1(-0.04,0.25)	1.55E-01	0.967	NAV1	2.02E+08	1	Island	chr1:201617041-201619788
cg23127291	0.46 (0.29, 0.64)	2.83E-06	0.015	-0.08(-0.24,0.08)	3.04E-01	0.967	CCDC102A	57563324	16	Island	chr16:57562451-57563325
cg18403361	-0.42 (-0.58, -0.26)	2.89E-06	0.015	0.04(-0.11,0.2)	5.94E-01	0.992	CLEC14A	38725750	14	S_Shore	chr14:38724254-38725537
cg21216268	0.32 (0.2, 0.44)	2.95E-06	0.015	-0.02(-0.1,0.06)	5.91E-01	0.992	PHC3	1.70E+08	3		
cg18875631	0.38 (0.23, 0.52)	2.99E-06	0.015	-0.05(-0.16,0.07)	4.22E-01	0.968	MIR543	1.01E+08	14		
cg07965335	0.44 (0.27, 0.61)	3.07E-06	0.015	-0.01(-0.11,0.09)	8.51E-01	0.992	SORBS2;SORBS2;SORI	1.87E+08	4		
cg05638165	0.45 (0.28, 0.62)	3.08E-06	0.015	0(-0.12,0.12)	9.40E-01	0.992	KIAA1614	1.81E+08	1		
cg04521224	0.5 (0.31, 0.68)	3.10E-06	0.015	0.03(-0.09,0.15)	6.22E-01	0.992		37332584	2		
cg22119716	-0.58 (-0.8, -0.36)	3.17E-06	0.015	.				3497214	2		
cg01078197	-0.47 (-0.65, -0.29)	3.34E-06	0.016	0.06(-0.18,0.31)	6.12E-01	0.992	MSL3L2	2.35E+08	2	Island	chr2:234776882-234777098
cg16993936	0.41 (0.25, 0.57)	3.43E-06	0.016	0(-0.17,0.18)	9.65E-01	0.994	OPTN;OPTN;OPTN;OP	13146903	10		
cg03351580	0.36 (0.22, 0.5)	3.46E-06	0.016	-0.02(-0.16,0.13)	8.27E-01	0.992		1.67E+08	6	N_Shore	chr6:166666837-166667541
cg21814615	0.25 (0.16, 0.35)	3.53E-06	0.016	-0.03(-0.13,0.08)	6.12E-01	0.992	KNTC1	1.23E+08	12		
cg23928234	0.43 (0.26, 0.59)	3.62E-06	0.016	0(-0.12,0.12)	9.71E-01	0.995	WWC2	1.84E+08	4		
cg22470309	-0.36 (-0.49, -0.22)	3.80E-06	0.017	0.07(-0.04,0.18)	2.25E-01	0.967	RGL3;RGL3	11529252	19	Island	chr19:11529213-11529588
cg27079740	0.45 (0.28, 0.62)	3.89E-06	0.017	0.09(-0.06,0.25)	2.35E-01	0.967	GPM6A;GPM6A;GPM6A	1.77E+08	4		
cg08529049	-0.34 (-0.47, -0.21)	3.94E-06	0.017	0.25(0.02,0.47)	3.38E-02	0.719	CWH43	48988038	4	Island	chr4:48987790-48988808

	TAIWAN COHORT			NHBCS			CpG Information				
Name	Coef (CI)	P-value	FDR P-value	Coef (CI)	P-value	FDR P-value	Gene	MAPINFO	CHR	Relation to	CpG Islands Name
cg23767840	1.61 (0.99, 2.24)	3.99E-06	0.017	-0.12(-0.31,0.07)	2.23E-01	0.967	EPN2;EPN2;EPN2	19174122	17		
cg10363926	0.38 (0.24, 0.53)	4.04E-06	0.017	-0.01(-0.11,0.08)	7.67E-01	0.992	SH3RF3	1.10E+08	2		
cg12497914	-0.51 (-0.71, -0.32)	4.12E-06	0.017	0.08(-0.06,0.22)	2.83E-01	0.967	EBF4	2674285	20	Island	chr20:2672904-2674698
cg25408698	0.36 (0.22, 0.49)	4.15E-06	0.017	0.01(-0.14,0.15)	8.99E-01	0.992	HRNBP3	77099293	17	N_Shore	chr17:77099995-77100203
cg11170796	0.37 (0.23, 0.51)	4.29E-06	0.018	0(-0.09,0.08)	9.15E-01	0.992	TCF3;TCF3	1650224	19	S_Shelf	chr19:1646117-1646713
cg12635602	0.5 (0.31, 0.7)	4.35E-06	0.018	-0.01(-0.22,0.19)	9.04E-01	0.992		58731357	11	Island	chr11:58731311-58731535
cg05494740	0.36 (0.22, 0.5)	4.45E-06	0.018	-0.03(-0.11,0.04)	4.09E-01	0.968	C16orf7	89775683	16	N_Shore	chr16:89777018-89778789
cg03709938	-0.3 (-0.42, -0.19)	4.50E-06	0.018	0.03(-0.07,0.13)	5.42E-01	0.99	MMEL1	2560950	1	Island	chr1:2559892-2561055
cg04705068	0.3 (0.18, 0.42)	4.58E-06	0.018	0.05(-0.11,0.21)	5.12E-01	0.989	FAM20B	1.79E+08	1		
cg14199148	-0.61 (-0.85, -0.37)	4.59E-06	0.018	0.06(-0.07,0.2)	3.75E-01	0.968	WDFY2;WDFY2	52158669	13	Island	chr13:52158056-52159235
cg15028160	-0.66 (-0.92, -0.41)	4.62E-06	0.018	-0.23(-0.59,0.13)	2.08E-01	0.967	PPFIA3;C19orf73;PPFIA	49622717	19	Island	chr19:49622023-49623635
cg15152301	0.39 (0.24, 0.54)	4.74E-06	0.018	0(-0.13,0.13)	9.79E-01	0.997	CACNA2D2;CACNA2D	50425602	3		
cg05876635	-0.42 (-0.59, -0.26)	4.83E-06	0.018	0.07(-0.06,0.2)	2.99E-01	0.967		55955383	7	Island	chr7:55954935-55955387
cg10985810*	0.36 (0.22, 0.5)	4.95E-06	0.018	0.02(-0.07,0.11)	6.61E-01	0.992	PTPRN2;PTPRN2;PTPR	1.58E+08	7	Island	chr7:158282214-158283382
cg22311507	-0.58 (-0.81, -0.35)	5.00E-06	0.018	-0.02(-0.15,0.1)	7.06E-01	0.992	MYCN	16079921	2	Island	chr2:16079801-16083032
cg17652792	0.26 (0.16, 0.37)	5.03E-06	0.018	-0.05(-0.22,0.12)	5.68E-01	0.991	PCDHB16;PCDHB16	1.41E+08	5	N_Shore	chr5:140563510-140564361
cg17416383	0.5 (0.31, 0.7)	5.06E-06	0.018	0(-0.13,0.12)	9.50E-01	0.993	HEXIM2	43247150	17	Island	chr17:43246839-43247167
cg10250651	-0.38 (-0.53, -0.23)	5.07E-06	0.018	0.03(-0.09,0.15)	6.10E-01	0.992	PNN	39644019	14	N_Shore	chr14:39644344-39645022
cg23052757	-0.45 (-0.63, -0.27)	5.12E-06	0.018	0.02(-0.13,0.17)	7.75E-01	0.992	C4orf17	1.00E+08	4		
cg02959759	-0.77 (-1.06, -0.47)	5.20E-06	0.019	.			CACNA1C;CACNA1C;C	2801584	12	S_Shore	chr12:2800139-2801062
cg00530298	0.67 (0.41, 0.93)	5.25E-06	0.019	0.09(-0.07,0.25)	2.62E-01	0.967	POLS	6749408	5		
cg25282083	0.43 (0.26, 0.6)	5.35E-06	0.019	0.04(-0.07,0.14)	5.04E-01	0.989	BAZ2A	56992955	12		
cg16906204	-0.37 (-0.51, -0.22)	5.61E-06	0.019	0.04(-0.09,0.17)	5.58E-01	0.99	HNRNPM;HNRNPM	8509784	19	Island	chr19:8509635-8510765
cg13688262	0.34 (0.21, 0.48)	5.62E-06	0.019	0.01(-0.13,0.15)	9.15E-01	0.992	TESK1	35609485	9	S_Shelf	chr9:35603968-35605991
cg23824713	0.49 (0.3, 0.68)	5.79E-06	0.02	-0.06(-0.26,0.15)	5.86E-01	0.992	SERPINB5	61157738	18		
cg11696986	-0.39 (-0.54, -0.23)	5.89E-06	0.02	0.08(-0.03,0.19)	1.72E-01	0.967		20837778	7	Island	chr7:20837551-20838369
cg08210297	0.35 (0.21, 0.49)	5.91E-06	0.02	0.08(-0.02,0.17)	1.18E-01	0.967	ALPI;ALPI	2.33E+08	2	N_Shelf	chr2:233323380-233323936
cg26974035	-0.51 (-0.71, -0.31)	6.37E-06	0.021	0.01(-0.1,0.12)	8.52E-01	0.992	LOC727677	1.28E+08	8		
cg25550629	0.35 (0.21, 0.48)	6.46E-06	0.021	0.04(-0.06,0.14)	4.52E-01	0.988		53546895	12		
cg13656504*	0.29 (0.17, 0.4)	6.50E-06	0.021	0.04(-0.08,0.17)	4.77E-01	0.988	TNXB	32030160	6		
cg18581445	-0.33 (-0.46, -0.2)	6.52E-06	0.021	0.04(-0.04,0.13)	3.07E-01	0.967	SHD;SHD	4280054	19	S_Shore	chr19:4279042-4279809
cg26618039	0.34 (0.2, 0.47)	6.53E-06	0.021	-0.05(-0.14,0.04)	2.68E-01	0.967	CPNE7;CPNE7	89650531	16	Island	chr16:89650417-89650697
cg21364719	0.29 (0.18, 0.41)	6.78E-06	0.021	0.02(-0.13,0.17)	8.12E-01	0.992	ZADH2	72919061	18	N_Shore	chr18:72919914-72921785
cg18207109	-0.41 (-0.58, -0.25)	6.81E-06	0.021	-0.06(-0.27,0.14)	5.44E-01	0.99	AIFM2	71889019	10	N_Shelf	chr10:71892092-71892802
cg01755562	-0.43 (-0.6, -0.26)	6.84E-06	0.021	0.05(-0.07,0.16)	3.98E-01	0.968	CLSTN2	1.40E+08	3	Island	chr3:139653430-139655115
cg13313430*	-0.61 (-0.85, -0.37)	6.84E-06	0.021	-0.04(-0.17,0.1)	5.74E-01	0.992	ICA1;ICA1;ICA1	8157910	7		
cg12522049	-0.51 (-0.72, -0.31)	6.94E-06	0.021	-0.04(-0.15,0.08)	5.39E-01	0.99		85798898	15	Island	chr15:85798738-85798964
cg17884327	-0.39 (-0.54, -0.23)	7.23E-06	0.022	-0.08(-0.34,0.18)	5.44E-01	0.99		1.98E+08	3	Island	chr3:197840490-197840783
cg18812904	0.45 (0.27, 0.63)	7.35E-06	0.022	-0.05(-0.16,0.05)	2.97E-01	0.967		77609087	18	Island	chr18:77608383-77609118
cg14221748	0.42 (0.25, 0.58)	7.36E-06	0.022	0.07(-0.06,0.21)	2.62E-01	0.967	BAT2L1	1.34E+08	9		
cg13365436*	-0.39 (-0.55, -0.24)	7.39E-06	0.022	-0.14(-0.33,0.05)	1.37E-01	0.967	AP2A2	968650	11		
cg19512961	0.31 (0.19, 0.43)	7.49E-06	0.022	-0.05(-0.16,0.05)	3.30E-01	0.967	GOLGA3	1.33E+08	12	Island	chr12:133363184-133363403
cg13853761	0.45 (0.27, 0.63)	7.51E-06	0.022	0(-0.17,0.16)	9.95E-01	0.997	C11orf2	64863232	11	N_Shore	chr11:64863698-64863980
cg25291978	0.46 (0.28, 0.64)	7.65E-06	0.022	0.03(-0.06,0.12)	5.19E-01	0.989	PLAGL1;PLAGL1	1.44E+08	6	S_Shore	chr6:144384894-144385888
cg05350411	-0.41 (-0.58, -0.25)	7.84E-06	0.023	0.06(-0.05,0.18)	2.78E-01	0.967		569174	4	Island	chr4:568513-570488
cg01994750	0.54 (0.33, 0.76)	8.02E-06	0.023	.				95500778	11		
cg05282260	0.62 (0.37, 0.87)	8.02E-06	0.023	-0.01(-0.28,0.26)	9.24E-01	0.992		1.57E+08	6		
ch.10.756792F	-0.51 (-0.72, -0.31)	8.05E-06	0.023	.				29649153	10		
cg25910443	-0.51 (-0.71, -0.31)	8.09E-06	0.023	0.04(-0.07,0.16)	4.40E-01	0.983	C9orf3	97488953	9	Island	chr9:97488818-97489172
cg08724753	0.36 (0.22, 0.51)	8.16E-06	0.023	-0.02(-0.15,0.1)	7.38E-01	0.992	C6orf103	1.47E+08	6		
cg22806341	-0.3 (-0.42, -0.18)	8.23E-06	0.023	0.07(-0.1,0.24)	4.41E-01	0.983		1.30E+08	10	N_Shore	chr10:130339526-130339777
cg24615831	-0.33 (-0.46, -0.2)	8.24E-06	0.023	0.12(0,0.24)	5.93E-02	0.843	PTS	1.12E+08	11	Island	chr11:112097022-112097547
cg15553408	0.39 (0.23, 0.55)	8.38E-06	0.023	0.07(-0.08,0.21)	3.70E-01	0.968	LTB;LTB	31549010	6	Island	chr6:31548436-31549277
cg08214267	0.48 (0.29, 0.67)	8.43E-06	0.023	-0.12(-0.23,-0.01)	3.09E-02	0.719	GPR77	47843887	19		
cg05026840	0.72 (0.43, 1.01)	8.46E-06	0.023	-0.07(-0.17,0.03)	1.46E-01	0.967	SCARF1;SCARF1;SCAF	1538656	17	Island	chr17:1538415-1538851
cg15998406*	-0.55 (-0.78, -0.33)	8.62E-06	0.023	-0.01(-0.29,0.27)	9.42E-01	0.992	EFNA2	1287832	19	Island	chr19:1287416-1287885
cg05028274	0.34 (0.2, 0.48)	8.63E-06	0.023	-0.09(-0.2,0.02)	1.08E-01	0.967		71755733	16	N_Shore	chr16:71756949-71758166
cg17210546	0.3 (0.18, 0.43)	8.79E-06	0.023	0.06(-0.04,0.16)	2.33E-01	0.967		3122422	19	Island	chr19:3122333-3122716
cg12818596	-0.28 (-0.39, -0.17)	8.87E-06	0.023	0.17(0,0.34)	4.77E-02	0.776		61119419	14	S_Shelf	chr14:61114102-61116552
cg15863322	0.44 (0.26, 0.62)	8.88E-06	0.023	-0.06(-0.18,0.05)	2.81E-01	0.967	MTUS2	29599410	13		
cg07473454	0.3 (0.18, 0.43)	8.93E-06	0.023	0.04(-0.04,0.11)	3.20E-01	0.967	PRAM1	8555526	19	Island	chr19:8555436-8555650
cg07826246	0.29 (0.17, 0.4)	9.05E-06	0.023	0.03(-0.06,0.12)	5.03E-01	0.989		1717111	7		
cg19594433	0.46 (0.27, 0.65)	9.16E-06	0.023	0.1(-0.02,0.22)	9.85E-02	0.967	SCAMP2	75147064	15		
cg05089296	-0.4 (-0.56, -0.24)	9.56E-06	0.024	0.03(-0.11,0.16)	7.11E-01	0.992		81636615	14		
cg19973338	0.35 (0.21, 0.49)	9.57E-06	0.024	0.01(-0.07,0.09)	7.30E-01	0.992	C17orf99	76141249	17		
cg15940508	-0.37 (-0.52, -0.22)	9.58E-06	0.024	0.14(-0.11,0.39)	2.70E-01	0.967		1.57E+08	6		
cg07105117	0.31 (0.18, 0.43)	9.87E-06	0.024	-0.03(-0.17,0.12)	7.07E-01	0.992	UGT1A10;UGT1A6;UG	2.35E+08	2	Island	chr2:234652246-234652720
cg24056567	-0.27 (-0.38, -0.16)	9.91E-06	0.024	0.09(-0.05,0.22)	1.92E-01	0.967	PSD3	18871909	8	Island	chr8:18871072-18872046
cg12413918*	0.41 (0.24, 0.58)	9.97E-06	0.024	0.07(-0.15,0.29)	5.19E-01	0.989	ERBB2;ERBB2	37855819	17	N_Shore	chr17:37856448-37856891
cg12879143	0.39 (0.23, 0.55)	1.00E-05	0.024	0.14(-0.06,0.34)	1.62E-01	0.967		1.39E+08	3	Island	chr3:138679244-138679579
cg00725356	-0.33 (-0.46, -0.2)	1.02E-05	0.025	0.17(0.03,0.31)	1.90E-02	0.719	KLF15	1.26E+08	3	Island	chr3:126074783-126076865
cg02580045	0.39 (0.23, 0.54)	1.02E-05	0.025	-0.01(-0.12,0.09)	7.85E-01	0.992	ANKRD13A	1.10E+08	12		
cg04136781	-0.31 (-0.44, -0.19)	1.04E-05	0.025	0.01(-0.11,0.13)	8.72E-01	0.992	PPT2;PPT2	32129786	6		
cg13968717	0.43 (0.26, 0.61)	1.04E-05	0.025	-0.04(-0.14,0.06)	4.46E-01	0.987	CDYL;CDYL;CDYL;CD	4954232	6		
cg21424664	0.38 (0.23, 0.54)	1.05E-05	0.025	0.01(-0.1,0.11)	9.00E-01	0.992	CHCHD2	56174299	7	Island	chr7:56173912-56174398
cg24138528	-0.48 (-0.67, -0.28)	1.07E-05	0.025	-0.02(-0.17,0.13)	8.13E-01	0.992		1.42E+08	6		
cg00782839	-0.5 (-0.7, -0.29)	1.08E-05	0.025	0.02(-0.17,0.22)	8.27E-01	0.992		8816130	2	N_Shore	chr2:8816314-8817264
cg05948856	0.42 (0.25, 0.59)	1.08E-05	0.025	-0.08(-0.24,0.07)	2.95E-01	0.967	EEF2;SNORD37	3982825	19	Island	chr19:3982824-3983190
cg14181982	-0.89 (-1.25, -0.52)	1.08E-05	0.025	0(-0.14,0.14)	9.84E-01	0.997	KLRAQ1;KLRAQ1;KLR	48668251	2	Island	chr2:48667559-48668355
cg09896346	-0.37 (-0.53, -0.22)	1.12E-05	0.026	0.14(-0.01,0.29)	6.71E-02	0.893	TCEA3;TCEA3	23751195	1	Island	chr1:23750508-23751663
cg16435686	1.72 (1.01, 2.42)	1.14E-05	0.026	0.05(-1.03,1.14)	9.21E-01	0.992		47964834	10	N_Shelf	chr10:47967314-47967636
cg18033436	-0.56 (-0.79, -0.33)	1.14E-05	0.026	-0.05(-0.17,0.08)	4.76E-01	0.988		1.03E+08	10	Island	chr10:102982302-102983600
cg26139606	0.32 (0.19, 0.45)	1.14E-05	0.026	-0.12(-0.25,0)	5.89E-02	0.843		44990002	13		
cg18793198	-0.41 (-0.58, -0.24)	1.15E-05	0.026	0.05(-0.07,0.18)	4.11E-01	0.968		1.31E+08	12	S_Shore	chr12:131389263-131389783
cg04881214*	-0.57 (-0.81, -0.34)	1.16E-05	0.026	-0.03(-0.18,0.11)	6.47E-01	0.992	PPM1B;PPM1B;PPM1B	44395864	2	Island	chr2:44394983-44396629
cg19768241	-0.34 (-0.47, -0.2)	1.17E-05	0.026	0.05(-0.12,0.23)	5.58E-01	0.99	CALCB	15095070	11	Island	chr11:15094957-15095872
cg09343421	0.3 (0.17, 0.42)	1.20E-05	0.026	0.03(-0.05,0.1)	4.77E-01	0.988	SRRT;SRRT;SRRT;SRR	1.00E+08	7	N_Shore	chr7:100484425-100484651
cg15355967	0.29 (0.17, 0.41)	1.21E-05	0.026	0(-0.08,0.08)	9.47E-01	0.993		3697542	2	N_Shore	chr2:3698659-3698875

	TAIWAN COHORT			NHBCS			CpG Information					
Name	Coef (CI)	P-value	FDR P-value	Coef (CI)	P-value	FDR P-value	Gene	MAPINFO	CHR	Relation to	CpG Islands	Name
cg17048169*	0.39 (0.23, 0.54)	1.21E-05	0.026	0.13(0,0.26)	5.31E-02	0.815	NGEF;NGEF	2.34E+08	2	S_Shelf	chr2:233740668-233741879	
cg25793785	0.27 (0.16, 0.38)	1.21E-05	0.026	0.16(-0.08,0.41)	1.93E-01	0.967		38281423	19	Island	chr19:38281090-38281424	
cg09239064	0.31 (0.19, 0.44)	1.22E-05	0.026	-0.09(-0.24,0.05)	1.93E-01	0.967	C16orf81	89229439	16	N_Shelf	chr16:89232410-89232683	
cg14700965	0.55 (0.33, 0.78)	1.24E-05	0.026	.				1.09E+08	8	N_Shelf	chr8:109455600-109456263	
cg13024668	0.4 (0.23, 0.56)	1.25E-05	0.026	0.01(-0.11,0.13)	8.76E-01	0.992		1.89E+08	2			
cg15136797	-0.41 (-0.57, -0.24)	1.25E-05	0.026	0.06(-0.08,0.19)	4.14E-01	0.968	RNF103	86851071	2	Island	chr2:86850042-86851178	
cg15179472	0.32 (0.19, 0.46)	1.25E-05	0.026	0.09(-0.04,0.21)	1.66E-01	0.967		70736858	15			
cg27183188	0.54 (0.32, 0.76)	1.25E-05	0.026	0.06(-0.06,0.18)	3.53E-01	0.968	ALDOC	26904121	17			
cg06167730	-0.37 (-0.52, -0.22)	1.26E-05	0.026	-0.08(-0.24,0.08)	3.11E-01	0.967		991533	8	Island	chr8:991532-991862	
cg20995172	0.31 (0.18, 0.43)	1.26E-05	0.026	-0.04(-0.13,0.05)	3.67E-01	0.968	C6orf47	31627225	6	N_Shore	chr6:31628338-31628658	
cg08582720	0.31 (0.19, 0.44)	1.27E-05	0.026	0.1(0,0.2)	5.11E-02	0.808	PLTP;PLTP	44538611	20	N_Shore	chr20:44539729-44540099	
cg20548564	0.47 (0.28, 0.67)	1.30E-05	0.026	-0.04(-0.2,0.12)	6.50E-01	0.992	POLR1D	28240073	13			
cg25463742	0.51 (0.3, 0.73)	1.30E-05	0.026	0.01(-0.12,0.14)	8.54E-01	0.992	CASZ1;CASZ1	10839574	1			
cg22938488	0.33 (0.19, 0.46)	1.31E-05	0.026	-0.01(-0.11,0.08)	7.73E-01	0.992	LPCAT1	1501670	5			
cg24120847	-0.44 (-0.62, -0.26)	1.31E-05	0.026	0.07(-0.05,0.19)	2.49E-01	0.967	DERL3;DERL3;DERL3	24181562	22	Island	chr22:24180491-24181666	
cg27150870	0.38 (0.23, 0.54)	1.31E-05	0.026	-0.09(-0.29,0.11)	3.72E-01	0.968	ANKRD11	89390609	16			
cg14894369	0.3 (0.18, 0.43)	1.32E-05	0.026	0.01(-0.19,0.2)	9.55E-01	0.993	DIABLO;DIABLO;DIAI	1.23E+08	12	N_Shore	chr12:122710054-122711371	
cg14933246	-0.43 (-0.61, -0.25)	1.33E-05	0.026	0(-0.14,0.14)	9.89E-01	0.997	PHF11;PHF11	50091767	13			
cg21218636	-0.37 (-0.52, -0.22)	1.33E-05	0.026	0.07(-0.08,0.22)	3.41E-01	0.967	PCTP;PCTP;PCTP	53828515	17	Island	chr17:53828422-53828738	
cg10439725	0.34 (0.2, 0.48)	1.35E-05	0.026	0.09(0,0.17)	3.91E-02	0.72	ZDHHC1	67431954	16	N_Shore	chr16:67433128-67433554	
cg04086977	0.34 (0.2, 0.48)	1.36E-05	0.027	0.02(-0.09,0.12)	7.53E-01	0.992		1.35E+08	10	N_Shore	chr10:134830984-134832209	
cg13713537*	-0.44 (-0.62, -0.26)	1.36E-05	0.027	-0.07(-0.19,0.06)	2.79E-01	0.967	NOS1	1.18E+08	12			
cg07283896	-0.41 (-0.57, -0.24)	1.37E-05	0.027	-0.06(-0.26,0.14)	5.28E-01	0.989		11855283	6			
cg21959334	0.37 (0.21, 0.52)	1.39E-05	0.027	-0.06(-0.17,0.04)	2.50E-01	0.967		2.29E+08	1	N_Shelf	chr1:228582428-228582894	
cg10847390	0.34 (0.2, 0.48)	1.41E-05	0.027	-0.01(-0.1,0.09)	9.09E-01	0.992		1.35E+08	10			
cg05701403	-0.37 (-0.52, -0.22)	1.44E-05	0.027	0.11(-0.1,0.31)	3.15E-01	0.967	PLEKHA7	16947551	11	S_Shore	chr11:16946392-16947398	
cg08943361	0.48 (0.28, 0.67)	1.44E-05	0.027	0.02(-0.11,0.16)	7.66E-01	0.992		56462172	16	S_Shelf	chr16:56458735-56459748	
cg24120960	0.44 (0.26, 0.62)	1.48E-05	0.028	.			STAG1	1.36E+08	3			
cg13078140*	-0.48 (-0.68, -0.28)	1.49E-05	0.028	-0.07(-0.2,0.06)	3.19E-01	0.967	SLIT2	20255082	4	Island	chr4:20253276-20256868	
cg13467672	-0.36 (-0.51, -0.21)	1.49E-05	0.028	0.09(-0.03,0.2)	1.26E-01	0.967		40464625	17	Island	chr17:40464293-40464947	
cg25692835	0.49 (0.28, 0.69)	1.52E-05	0.028	0.1(-0.01,0.22)	7.21E-02	0.893	C2orf64;UNC50;C2orf64	99224894	2	Island	chr2:99224681-99225322	
cg27023871	0.7 (0.41, 0.99)	1.55E-05	0.029	-0.02(-0.19,0.15)	8.13E-01	0.992	TBCD	80848259	17			
cg04446345	-0.32 (-0.45, -0.19)	1.56E-05	0.029	0.05(-0.06,0.16)	3.52E-01	0.968	UPK3A;UPK3A	45680844	22	Island	chr22:45680644-45681016	
cg17619347	0.47 (0.27, 0.66)	1.56E-05	0.029	-0.01(-0.11,0.08)	7.95E-01	0.992	SRRT;SRRT;SRRT;SRR	1.00E+08	7			
cg04903360*	-0.35 (-0.49, -0.2)	1.58E-05	0.029	-0.04(-0.18,0.09)	5.17E-01	0.989	PARD3	35103065	10	N_Shore	chr10:35103124-35105243	
cg18514573	-0.47 (-0.67, -0.27)	1.58E-05	0.029	0.1(-0.11,0.31)	3.54E-01	0.968		21610423	6			
cg21412973	-0.26 (-0.36, -0.15)	1.58E-05	0.029	0.04(-0.12,0.21)	6.03E-01	0.992		1408557	4	Island	chr4:1408521-1409224	
cg11852843	0.29 (0.17, 0.4)	1.59E-05	0.029	0(-0.12,0.12)	9.63E-01	0.993	OBSCN;OBSCN	2.28E+08	1	N_Shore	chr1:228473835-228474043	
cg14582248	0.39 (0.23, 0.56)	1.59E-05	0.029	0(-0.1,0.1)	9.71E-01	0.995		5066984	6			
cg21801445	-0.35 (-0.49, -0.2)	1.62E-05	0.029	-0.09(-0.19,0.02)	9.84E-02	0.967		13213606	18	N_Shelf	chr18:13217073-13219044	
cg05658707	0.44 (0.25, 0.62)	1.63E-05	0.029	0.03(-0.18,0.23)	7.97E-01	0.992	VGLL4	11722129	3			
cg09230014	0.39 (0.23, 0.55)	1.65E-05	0.029	-0.01(-0.18,0.16)	8.97E-01	0.992	PTPRN2;PTPRN2;PTPR	1.58E+08	7			
cg19774624	-0.67 (-0.95, -0.39)	1.66E-05	0.029	.			HDAC5;HDAC5	42201019	17	Island	chr17:42200521-42201543	
cg22510637	0.36 (0.21, 0.51)	1.66E-05	0.029	-0.02(-0.13,0.09)	7.02E-01	0.992	R3HCC1	23147025	8	S_Shore	chr8:23145528-23146006	
cg02503850	-0.26 (-0.37, -0.15)	1.67E-05	0.029	0.13(-0.02,0.27)	9.15E-02	0.967	ADAMTS14;ADAMTS1	72432552	10	Island	chr10:72432242-72432738	
cg15192041	0.33 (0.19, 0.47)	1.69E-05	0.029	-0.02(-0.16,0.11)	7.27E-01	0.992		69414923	11			
cg01538575	0.29 (0.17, 0.41)	1.71E-05	0.029	0(-0.08,0.09)	9.16E-01	0.992	RFX2;RFX2	6039991	19	Island	chr19:6039990-6040228	
cg00961792	0.26 (0.15, 0.37)	1.72E-05	0.029	0.01(-0.05,0.07)	7.62E-01	0.992	PIGR	2.07E+08	1			
cg02808220	0.41 (0.24, 0.58)	1.72E-05	0.029	-0.03(-0.14,0.08)	5.94E-01	0.992	MIR483;INS-IGF2;IGF2	2156282	11	S_Shore	chr11:2154033-2154387	
cg07247103	0.31 (0.18, 0.44)	1.72E-05	0.029	0.04(-0.05,0.12)	4.06E-01	0.968	BANP;BANP	88059817	16	N_Shore	chr16:88061122-88061595	
cg10473311*	-0.64 (-0.91, -0.37)	1.72E-05	0.029	-0.34(-0.6,-0.08)	1.19E-02	0.719	PTPRN2;PTPRN2;PTPR	1.58E+08	7			
cg16256065	0.35 (0.2, 0.5)	1.72E-05	0.029	0(-0.11,0.11)	9.99E-01	0.999	LRP5	68216328	11	Island	chr11:68216289-68216547	
cg01563959	0.4 (0.23, 0.57)	1.73E-05	0.029	-0.03(-0.14,0.08)	5.91E-01	0.992	CYTSB;CYTSB	20014025	17			
cg01791874	-0.58 (-0.82, -0.33)	1.74E-05	0.029	-0.02(-0.11,0.07)	6.88E-01	0.992	11-Mar	16180055	5	Island	chr5:16179064-16180420	
cg25941154	0.33 (0.19, 0.47)	1.74E-05	0.029	0(-0.14,0.14)	9.95E-01	0.997	JAZF1	28208557	7			
cg01646289	0.43 (0.25, 0.61)	1.76E-05	0.029	0.06(-0.06,0.18)	3.35E-01	0.967		1.35E+08	10	Island	chr10:134732601-134733777	
cg14757492	0.29 (0.17, 0.41)	1.76E-05	0.029	-0.05(-0.15,0.05)	3.19E-01	0.967	COPE;DDX49;COPE;CC	19029194	19	N_Shore	chr19:19029924-19030653	
cg20897072	0.54 (0.32, 0.77)	1.76E-05	0.029	0.1(-0.05,0.24)	1.88E-01	0.967		1495082	6			
cg14256983	0.38 (0.22, 0.53)	1.77E-05	0.029	0(-0.17,0.16)	9.59E-01	0.993	C9orf102	98639872	9	S_Shore	chr9:98637245-98638445	
cg24309215	0.31 (0.18, 0.45)	1.77E-05	0.029	0.04(-0.14,0.21)	6.83E-01	0.992		1.21E+08	12	N_Shore	chr12:121022419-121022704	
cg26286098	0.56 (0.32, 0.8)	1.77E-05	0.029	.			B3GALTL	31905755	13			
cg16750440	-0.58 (-0.82, -0.33)	1.80E-05	0.029	0.06(-0.14,0.26)	5.68E-01	0.991	C10orf79	1.06E+08	10	S_Shore	chr10:105991899-105992248	
cg26876637	0.35 (0.2, 0.5)	1.82E-05	0.029	-0.14(-0.25,-0.02)	2.13E-02	0.719	HRNR	1.52E+08	1			
cg02607130*	0.47 (0.27, 0.67)	1.83E-05	0.029	0.05(-0.1,0.21)	5.02E-01	0.989	GRIN3B	1008643	19	Island	chr19:1007551-1012222	
cg04756597	0.31 (0.18, 0.44)	1.84E-05	0.029	0.04(-0.05,0.14)	3.67E-01	0.968		44730607	22	S_Shelf	chr22:44726724-44727590	
cg23911291*	0.31 (0.18, 0.45)	1.87E-05	0.03	0.05(-0.07,0.18)	3.85E-01	0.968	DAXX;DAXX;DAXX;D	33288572	6	N_Shore	chr6:33288733-33289008	
cg26266708	-0.57 (-0.81, -0.33)	1.89E-05	0.03	.			ISL2	76630962	15	Island	chr15:76630029-76630970	
cg05567435	0.37 (0.21, 0.53)	1.92E-05	0.03	-0.02(-0.18,0.13)	7.75E-01	0.992	METTL7B	56074329	12			
cg18594551	0.31 (0.18, 0.44)	1.95E-05	0.031	0.06(-0.03,0.15)	2.11E-01	0.967	KRTAP23-1	31720835	21			
cg12384572	0.4 (0.23, 0.57)	1.98E-05	0.031	0.03(-0.08,0.14)	5.61E-01	0.99	USP4;USP4	49377951	3	S_Shore	chr3:49377340-49377854	
cg13116816	0.41 (0.24, 0.59)	2.00E-05	0.031	0.13(-0.25,0.51)	4.88E-01	0.988	C7orf23	86839071	7			
cg02305765	-0.6 (-0.85, -0.35)	2.01E-05	0.031	-0.05(-0.24,0.13)	5.59E-01	0.99		53107910	12	N_Shore	chr12:53107912-53108471	
cg25994616	-0.54 (-0.76, -0.31)	2.01E-05	0.031	-0.03(-0.15,0.09)	6.18E-01	0.992	NOTCH4	32166739	6	S_Shelf	chr6:32163292-32164383	
cg18523886	0.31 (0.18, 0.45)	2.02E-05	0.031	-0.09(-0.26,0.08)	3.02E-01	0.967		39126551	20			
cg02399294	0.28 (0.16, 0.4)	2.03E-05	0.031	0.01(-0.14,0.15)	8.98E-01	0.992	CCDC40	78063763	17			
cg04432454	-0.45 (-0.64, -0.26)	2.03E-05	0.031	-0.05(-0.19,0.09)	5.00E-01	0.989	LMF1	949229	16	N_Shore	chr16:949658-949960	
cg26721877	-0.35 (-0.49, -0.2)	2.03E-05	0.031	0.06(-0.11,0.22)	5.09E-01	0.989	STC2	1.73E+08	5	Island	chr5:172754056-172757098	
cg04299274	0.35 (0.2, 0.5)	2.05E-05	0.031	0.01(-0.07,0.1)	7.36E-01	0.992	KSR1	25798741	17			
cg08854799	0.33 (0.19, 0.47)	2.08E-05	0.031	0.02(-0.11,0.15)	7.54E-01	0.992	FAM83E	49107139	19	Island	chr19:49106882-49107178	
cg06505394	-0.72 (-1.02, -0.41)	2.09E-05	0.031	-0.04(-0.26,0.18)	7.04E-01	0.992	HNRNPUL1;HNRNPUL	41771095	19	Island	chr19:41770013-41771113	
cg09244071	0.32 (0.18, 0.46)	2.10E-05	0.031	0.05(-0.05,0.15)	2.96E-01	0.967	CUX1;CUX1;CUX1	1.02E+08	7			
cg23077029	-0.23 (-0.32, -0.13)	2.11E-05	0.031	0.09(-0.01,0.19)	7.75E-02	0.893		1.08E+08	6	N_Shore	chr6:107810066-107812733	
cg12640770	0.42 (0.24, 0.6)	2.13E-05	0.031	0.13(-0.04,0.29)	1.46E-01	0.967	IRAK1BP1	79577691	6	Island	chr6:79577158-79577756	
cg06988897	-0.57 (-0.81, -0.33)	2.14E-05	0.031	-0.06(-0.44,0.32)	7.39E-01	0.992	FAM38A	88803803	16	Island	chr16:88803802-88804112	
cg07281948	0.47 (0.27, 0.67)	2.14E-05	0.031	-0.08(-0.21,0.05)	2.14E-01	0.967	LRP1	57601935	12			
cg10905310	-0.52 (-0.74, -0.3)	2.14E-05	0.031	0.03(-0.09,0.15)	6.15E-01	0.992	GDPD5	75236908	11	Island	chr11:75236189-75237781	
cg00513411	-0.5 (-0.72, -0.29)	2.17E-05	0.032	0.01(-0.09,0.11)	8.21E-01	0.992		21586262	6	N_Shore	chr6:21587421-21589175	
cg08495545	0.36 (0.21, 0.51)	2.18E-05	0.032	-0.01(-0.11,0.08)	7.69E-01	0.992		76965309	17			

	TAIWAN COHORT			NHBCS			CpG Information					
Name	Coef (CI)	P-value	FDR P-value	Coef (CI)	P-value	FDR P-value	Gene	MAPINFO	CHR	Relation to	CpG Islands	Name
cg16034168	-0.62 (-0.88, -0.35)	2.18E-05	0.032	0.06(-0.07,0.19)	3.82E-01	0.968	ACOT7;ACOT7;ACOT7	6336711	1	N_Shelf	chr1:6339948-6340268	
cg13437525	-0.5 (-0.71, -0.29)	2.20E-05	0.032	0.21(0.08,0.35)	2.47E-03	0.493	METTL7A	51318487	12			
cg04635504	0.26 (0.15, 0.37)	2.21E-05	0.032	0.05(-0.08,0.19)	4.21E-01	0.968	KCNQ1;KCNQ1	2829241	11	S_Shore	chr11:2828445-2828884	
cg15202954	-0.39 (-0.56, -0.23)	2.21E-05	0.032	.			NALCN	1.02E+08	13	Island	chr13:102068116-102069258	
cg05533124	-0.5 (-0.71, -0.28)	2.22E-05	0.032	0.03(-0.13,0.2)	6.90E-01	0.992	AQP4;AQP4	24443298	18	Island	chr18:24443200-24443458	
cg24194674	0.89 (0.51, 1.27)	2.22E-05	0.032	-0.04(-0.22,0.14)	6.63E-01	0.992	ADAM12;ADAM12	1.28E+08	10			
cg08839808	0.77 (0.44, 1.1)	2.25E-05	0.032	-0.08(-0.43,0.26)	6.28E-01	0.992		1.57E+08	6			
cg14948795	0.3 (0.17, 0.43)	2.25E-05	0.032	-0.12(-0.21,-0.02)	1.41E-02	0.719	CDK2AP1	1.24E+08	12	S_Shore	chr12:123755246-123756408	
cg05655247	0.41 (0.24, 0.59)	2.26E-05	0.032	0.01(-0.1,0.12)	8.72E-01	0.992		1065028	1			
cg14641600	0.27 (0.16, 0.39)	2.26E-05	0.032	-0.04(-0.14,0.05)	3.72E-01	0.968	TCF7;TCF7;TCF7;TCF7	1.33E+08	5			
cg26773771	-0.72 (-1.03, -0.41)	2.26E-05	0.032	-0.01(-0.14,0.12)	9.16E-01	0.992		21830857	17			
cg07138793	0.52 (0.3, 0.75)	2.28E-05	0.032	-0.12(-0.34,0.11)	2.99E-01	0.967	OCA2	28148330	15	Island	chr15:28147925-28148465	
cg11841231	0.35 (0.2, 0.5)	2.29E-05	0.032	-0.2(-0.39,-0.01)	3.67E-02	0.719	PARD3B;PARD3B;PAR	2.06E+08	2			
cg26577201	-0.43 (-0.61, -0.25)	2.29E-05	0.032	-0.01(-0.15,0.14)	9.24E-01	0.992	CRYBA2;CRYBA2;CRY	2.20E+08	2	Island	chr2:219857682-219858917	
cg10729426	-0.57 (-0.81, -0.33)	2.37E-05	0.033	-0.14(-0.3,0.02)	7.75E-02	0.893	ZNF549	58038585	19	Island	chr19:58038572-58039208	
cg18527919	0.42 (0.24, 0.61)	2.37E-05	0.033	0.05(-0.12,0.22)	5.63E-01	0.99	C2orf84	24398170	2	Island	chr2:24397645-24398194	
cg15334916	0.38 (0.22, 0.54)	2.40E-05	0.033	0.06(-0.09,0.2)	4.35E-01	0.982	COL13A1;COL13A1;CO	71563776	10	S_Shore	chr10:71561302-71562876	
cg27583604	0.32 (0.19, 0.46)	2.41E-05	0.033	0.07(-0.03,0.17)	1.58E-01	0.967	PLK3	45267345	1	S_Shore	chr1:45265534-45266783	
cg15191798	-0.59 (-0.85, -0.34)	2.42E-05	0.033	-0.01(-0.22,0.19)	8.88E-01	0.992	LRRC49;THAP10;THAI	71184618	15	Island	chr15:71184276-71184689	
cg15979150	-0.37 (-0.52, -0.21)	2.43E-05	0.033	.			FAM164C;FAM164C	75535757	14	N_Shore	chr14:75535825-75536376	
cg07894883	0.3 (0.17, 0.43)	2.44E-05	0.033	0.1(-0.03,0.22)	1.30E-01	0.967	ZCCHC24	81146877	10	S_Shore	chr10:81145958-81146197	
cg22551163	0.54 (0.31, 0.77)	2.45E-05	0.033	-0.01(-0.17,0.15)	9.12E-01	0.992	SFRS8	1.32E+08	12			
cg05071046	-0.3 (-0.43, -0.17)	2.46E-05	0.033	0.08(-0.01,0.18)	8.95E-02	0.967		54325583	16	Island	chr16:54325040-54325703	
cg21849812	0.34 (0.19, 0.49)	2.48E-05	0.033	-0.05(-0.18,0.08)	4.43E-01	0.984		17848522	1			
cg05890620	-0.43 (-0.61, -0.24)	2.49E-05	0.033	0.04(-0.06,0.14)	4.14E-01	0.968	ICAM4;ICAM4;ICAM4	10397612	19	Island	chr19:10397603-10398703	
cg10255761	0.35 (0.2, 0.5)	2.49E-05	0.033	-0.03(-0.15,0.08)	5.39E-01	0.99	KLHDC8B	49210029	3	S_Shore	chr3:49208629-49209196	
cg06962768	0.37 (0.21, 0.52)	2.51E-05	0.033	0.12(-0.02,0.27)	1.03E-01	0.967	EPDR1	37959782	7	N_Shore	chr7:37960316-37961046	
cg17702455	1.22 (0.69, 1.74)	2.53E-05	0.033	-0.04(-0.23,0.15)	6.73E-01	0.992	JAZF1	28058989	7			
cg11160944	-0.67 (-0.95, -0.38)	2.54E-05	0.033	0.14(0.02,0.26)	2.05E-02	0.719		45570048	17	Island	chr17:45569833-45570134	
cg12424468	0.39 (0.22, 0.56)	2.55E-05	0.033	-0.05(-0.19,0.09)	4.66E-01	0.988	C11orf2	64863373	11	N_Shore	chr11:64863698-64863980	
cg10419849	-0.35 (-0.5, -0.2)	2.57E-05	0.033	0.04(-0.08,0.17)	4.81E-01	0.988	TNKS	9537322	8			
cg21578322	-0.52 (-0.75, -0.3)	2.57E-05	0.033	0.03(-0.08,0.14)	5.55E-01	0.99	TTLL9;DUSP15;DUSP1	30458044	20	Island	chr20:30457907-30458712	
cg21211873*	0.35 (0.2, 0.5)	2.58E-05	0.033	0.03(-0.06,0.13)	4.87E-01	0.988	CYFIP1;CYFIP1	22960638	15			
cg25373579	0.27 (0.15, 0.38)	2.58E-05	0.033	-0.01(-0.08,0.06)	8.10E-01	0.992	VEGFA;VEGFA;VEGFA	43752302	6			
cg07015927	0.3 (0.17, 0.42)	2.59E-05	0.033	0.01(-0.14,0.17)	8.57E-01	0.992	DSCR8;DSCR8;DSCR8;	39493537	21			
cg19490868	-0.28 (-0.4, -0.16)	2.59E-05	0.033	0.05(-0.09,0.2)	4.86E-01	0.988	RTN1;RTN1;RTN1	60097124	14	N_Shore	chr14:60097208-60097553	
cg05422836	-0.42 (-0.61, -0.24)	2.61E-05	0.033	0.06(-0.14,0.26)	5.48E-01	0.99		81806286	8	Island	chr8:81805955-81806327	
cg09544549*	0.32 (0.18, 0.45)	2.63E-05	0.033	0.07(-0.05,0.2)	2.31E-01	0.967	MYH9	36697013	22			
cg10927727	0.34 (0.19, 0.48)	2.63E-05	0.033	0.03(-0.1,0.16)	6.06E-01	0.992		1.31E+08	2	Island	chr2:130986131-130987183	
cg17542795	-0.53 (-0.75, -0.3)	2.65E-05	0.034	.				1.79E+08	4			
cg07329149	-0.55 (-0.78, -0.31)	2.69E-05	0.034	0.04(-0.06,0.15)	4.31E-01	0.98	NDUFB3;FAM126B	2.02E+08	2	Island	chr2:201936024-201936376	
cg22488462	0.3 (0.17, 0.42)	2.70E-05	0.034	-0.11(-0.27,0.05)	1.85E-01	0.967	CARS2	1.11E+08	13			
cg14106345	0.43 (0.24, 0.61)	2.71E-05	0.034	0.07(-0.07,0.2)	3.24E-01	0.967	RNF112	19317240	17			
cg04496485	-0.37 (-0.53, -0.21)	2.74E-05	0.034	-0.05(-0.21,0.12)	5.64E-01	0.99		714565	1	S_Shore	chr1:713984-714547	
cg16184346	0.35 (0.2, 0.5)	2.74E-05	0.034	-0.08(-0.26,0.1)	3.95E-01	0.968	PLEKHB1;PLEKHB1;PI	73367968	11	N_Shelf	chr11:73371800-73372632	
cg08458852	0.63 (0.36, 0.9)	2.79E-05	0.035	0.11(-0.05,0.26)	1.68E-01	0.967	PARN;PARN	14529837	16			
cg00280177	0.45 (0.26, 0.65)	2.81E-05	0.035	.				24799870	8	Island	chr8:24799703-24800147	
cg06700226	0.36 (0.21, 0.52)	2.81E-05	0.035	-0.02(-0.13,0.09)	7.55E-01	0.992		5198019	12			
cg14295458	0.39 (0.22, 0.55)	2.81E-05	0.035	0.08(-0.09,0.25)	3.38E-01	0.967	OSBPL7	45891164	17	Island	chr17:45890956-45891165	
cg26748794	-1 (-1.43, -0.57)	2.83E-05	0.035	-0.4(-1.34,0.53)	3.97E-01	0.968	FAM38A	88804051	16	Island	chr16:88803802-88804112	
cg05055490	0.37 (0.21, 0.53)	2.85E-05	0.035	-0.02(-0.16,0.12)	7.45E-01	0.992		26284602	7			
cg06219206	-0.46 (-0.66, -0.26)	2.86E-05	0.035	0.05(-0.06,0.16)	3.99E-01	0.968	SOX5	24383177	12			
cg19149031	0.51 (0.29, 0.73)	2.86E-05	0.035	-0.15(-0.51,0.21)	4.02E-01	0.968	SLC22A3	1.61E+08	6	Island	chr6:160783666-160784135	
cg20908131	-0.6 (-0.87, -0.34)	2.86E-05	0.035	0.08(-0.07,0.24)	2.79E-01	0.967	C14orf64	98419775	14			
cg24994360	0.3 (0.17, 0.44)	2.90E-05	0.035	0.01(-0.11,0.14)	8.06E-01	0.992		2.42E+08	2			
cg21391023	0.48 (0.27, 0.69)	2.91E-05	0.035	0.11(-0.05,0.28)	1.68E-01	0.967	BTBD8	92545580	1	N_Shore	chr1:92545819-92546480	
cg08800878	-0.28 (-0.41, -0.16)	2.92E-05	0.035	0.01(-0.12,0.15)	8.76E-01	0.992	EMX1	73152809	2	S_Shore	chr2:73151200-73152060	
cg24724630	-0.32 (-0.45, -0.18)	2.92E-05	0.035	0.01(-0.11,0.14)	8.33E-01	0.992	NRG2;NRG2;NRG2;NR	1.39E+08	5			
cg09173861	0.35 (0.2, 0.5)	2.93E-05	0.035	0.01(-0.12,0.14)	8.65E-01	0.992	SYT17	19180212	16	Island	chr16:19179002-19180217	
cg18333694	0.27 (0.15, 0.39)	2.95E-05	0.035	-0.02(-0.1,0.05)	5.39E-01	0.99		357882	11	S_Shore	chr11:355947-356447	
cg06187356	-0.37 (-0.53, -0.21)	2.96E-05	0.035	0(-0.13,0.12)	9.59E-01	0.993	NPBWR2	62738210	20	S_Shore	chr20:62737382-62738071	
cg08845333	-0.3 (-0.43, -0.17)	2.96E-05	0.035	0.09(-0.02,0.19)	1.08E-01	0.967	ICAM5	10403403	19	Island	chr19:10399844-10405375	
cg11058229	0.32 (0.18, 0.45)	3.00E-05	0.035	-0.01(-0.08,0.07)	8.63E-01	0.992		97089082	14			
cg21514227	-0.33 (-0.47, -0.18)	3.00E-05	0.035	0.04(-0.11,0.19)	5.94E-01	0.992	NOTCH3	15288315	19	Island	chr19:15288314-15288911	
cg24206697	-0.49 (-0.71, -0.28)	3.00E-05	0.035	0.01(-0.09,0.1)	8.96E-01	0.992	ZNF514	95822125	2	N_Shelf	chr2:95824802-95825721	
cg05041061	0.5 (0.28, 0.72)	3.04E-05	0.035	-0.02(-0.23,0.19)	8.70E-01	0.992	BAHCC1	79426049	17	Island	chr17:79425847-79426169	
cg13045351	0.69 (0.39, 0.99)	3.05E-05	0.035	.			ZSCAN16	28092048	6			
cg16120811	-0.28 (-0.4, -0.16)	3.07E-05	0.036	0.05(-0.11,0.22)	5.31E-01	0.989	WEE1	9594484	11	Island	chr11:9594346-9596536	
cg04594483	0.35 (0.2, 0.51)	3.10E-05	0.036	0.05(-0.09,0.2)	4.52E-01	0.988	RADIL	4862020	7			
cg21885134	0.3 (0.17, 0.43)	3.10E-05	0.036	-0.06(-0.15,0.04)	2.22E-01	0.967		1794994	5			
cg26076233	2.71 (1.53, 3.9)	3.12E-05	0.036	.			DYNLL1	1.21E+08	12	N_Shelf	chr12:120933649-120934713	
cg14431547	0.32 (0.18, 0.46)	3.14E-05	0.036	0.11(-0.1,0.32)	2.89E-01	0.967	MELK	36572421	9	N_Shore	chr9:36572703-36573073	
cg07759052	-0.54 (-0.77, -0.3)	3.16E-05	0.036	-0.02(-0.16,0.13)	8.37E-01	0.992	HMBOX1;HMBOX1	28752150	8	S_Shelf	chr8:28748070-28748431	
cg05229340	0.38 (0.22, 0.55)	3.21E-05	0.037	.			CCDC46	63833529	17			
cg13923018	-0.33 (-0.48, -0.19)	3.22E-05	0.037	0.09(-0.03,0.21)	1.24E-01	0.967	PDZD2	31855380	5	Island	chr5:31855003-31855426	
cg22454769	-0.41 (-0.59, -0.23)	3.24E-05	0.037	-0.18(-0.37,0.01)	6.10E-02	0.843	FHL2;FHL2;FHL2;FHL2	1.06E+08	2	Island	chr2:106014878-106015884	
cg22647900	0.36 (0.2, 0.51)	3.24E-05	0.037	-0.06(-0.18,0.05)	2.68E-01	0.967	EIF3M	32604844	11	N_Shore	chr11:32605183-32605729	
cg14122138	-0.58 (-0.83, -0.33)	3.25E-05	0.037	0(-0.16,0.16)	9.80E-01	0.997	SRCIN1	36719737	17	Island	chr17:36719544-36719938	
cg03097134	0.33 (0.18, 0.47)	3.26E-05	0.037	-0.08(-0.18,0.03)	1.47E-01	0.967		37598264	6			
cg01602153	0.53 (0.3, 0.77)	3.27E-05	0.037	0.11(-0.04,0.26)	1.55E-01	0.967	FAM110B	59058660	8	Island	chr8:59058167-59059414	
cg12837905	0.36 (0.2, 0.52)	3.32E-05	0.037	-0.03(-0.15,0.09)	5.84E-01	0.992		1.32E+08	10	N_Shore	chr10:131813042-131813277	
cg21204804	-0.41 (-0.59, -0.23)	3.32E-05	0.037	0.15(0.03,0.28)	1.92E-02	0.719		8290321	11	Island	chr11:8289532-8290322	
cg17377797	0.35 (0.2, 0.51)	3.33E-05	0.037	0.09(-0.1,0.27)	3.42E-01	0.967	OR6C1	55714289	12			
cg06954677	-0.37 (-0.53, -0.21)	3.35E-05	0.037	-0.03(-0.15,0.1)	6.74E-01	0.992	IDO2	39873708	8			
cg06467181	-0.42 (-0.6, -0.23)	3.37E-05	0.037	0.03(-0.12,0.17)	7.00E-01	0.992	ZCCHC2	60191671	18	Island	chr18:60189509-60192109	
cg11860412	-0.25 (-0.36, -0.14)	3.37E-05	0.037	0.02(-0.09,0.14)	6.77E-01	0.992	ONECUT3	1753217	19	Island	chr19:1753216-1755606	
cg24592500	0.72 (0.4, 1.04)	3.39E-05	0.037	0.02(-0.13,0.17)	7.88E-01	0.992		2863801	1			
cg25142228	0.22 (0.12, 0.31)	3.42E-05	0.037	-0.04(-0.12,0.05)	3.85E-01	0.968	SETD1B	1.22E+08	12	N_Shore	chr12:122247781-122248424	

	TAIWAN COHORT			NHBCS			CpG Information				
Name	Coef (CI)	P-value	FDR P-value	Coef (CI)	P-value	FDR P-value	Gene	MAPINFO	CHR	Relation to	CpG Islands Name
cg09822423	-0.42 (-0.61, -0.24)	3.43E-05	0.037	.				1.93E+08	1		
cg24349819	-0.34 (-0.49, -0.19)	3.43E-05	0.037	-0.05(-0.19,0.09)	4.73E-01	0.988	ARHGAP27;ARHGAP27	43476547	17	S_Shelf	chr17:43472527-43474343
cg20989855	-0.37 (-0.53, -0.21)	3.45E-05	0.037	-0.09(-0.31,0.12)	3.86E-01	0.968	KCNIP4;KCNIP4;KCNIP	20985927	4		
cg11776115*	0.32 (0.18, 0.46)	3.47E-05	0.037	0.04(-0.05,0.14)	3.36E-01	0.967	DOT1L	2189267	19	N_Shore	chr19:2190989-2191262
cg05479662	0.35 (0.19, 0.5)	3.48E-05	0.037	0.05(-0.08,0.18)	4.39E-01	0.983		47347811	8		
cg05502283	0.32 (0.18, 0.46)	3.48E-05	0.037	-0.06(-0.14,0.03)	1.77E-01	0.967	NFIX	13203715	19	N_Shelf	chr19:13207375-13207621
cg16325482	0.36 (0.2, 0.52)	3.48E-05	0.037	0(-0.11,0.11)	9.39E-01	0.992	CANT1;CANT1;CANT1	76989499	17	N_Shore	chr17:76989642-76990039
cg07152869	0.42 (0.23, 0.6)	3.51E-05	0.038	-0.25(-0.76,0.26)	3.38E-01	0.967	KIAA0556	27741555	16		
cg05590569	0.34 (0.19, 0.49)	3.52E-05	0.038	0.03(-0.13,0.18)	7.38E-01	0.992		10144009	2		
cg10140240	-0.37 (-0.54, -0.21)	3.53E-05	0.038	0.01(-0.09,0.11)	7.97E-01	0.992	BIRC5;BIRC5;BIRC5;EI	76220672	17		
cg01841415	0.33 (0.18, 0.47)	3.56E-05	0.038	0.03(-0.09,0.15)	6.15E-01	0.992	KCNT1	1.39E+08	9	N_Shore	chr9:138680881-138681159
cg18372548	0.35 (0.19, 0.5)	3.57E-05	0.038	0(-0.11,0.12)	9.36E-01	0.992	STIM2;STIM2;STIM2	26913852	4		
cg00766914	-0.33 (-0.47, -0.18)	3.60E-05	0.038	0.02(-0.1,0.14)	7.43E-01	0.992	PDSS2	1.08E+08	6	Island	chr6:107780774-107781409
cg20491488	0.34 (0.19, 0.49)	3.62E-05	0.038	0.04(-0.12,0.19)	6.48E-01	0.992	ITPRIP	1.06E+08	10	Island	chr10:106074965-106075368
cg22343181	0.41 (0.23, 0.59)	3.64E-05	0.038	-0.06(-0.15,0.04)	2.45E-01	0.967	NOC4L	1.33E+08	12	N_Shore	chr12:132633217-132635301
cg07058086	-0.56 (-0.8, -0.31)	3.65E-05	0.038	0.03(-0.09,0.15)	6.74E-01	0.992	KIF13B	29120186	8	Island	chr8:29119957-29120787
cg14646974	-1.1 (-1.58, -0.61)	3.66E-05	0.038	0.06(-0.53,0.65)	8.42E-01	0.992	HSD17B7P2	38645378	10	Island	chr10:38645112-38645513
cg11556164	-0.3 (-0.43, -0.17)	3.69E-05	0.038	-0.11(-0.25,0.03)	1.17E-01	0.967	LRRN3;IMMP2L;LRRN	1.11E+08	7		
cg04025701	-0.4 (-0.58, -0.23)	3.70E-05	0.038	0.04(-0.09,0.17)	5.31E-01	0.989	RPS14;RPS14;RPS14;RI	1.50E+08	5	Island	chr5:149828881-149829631
cg16725642	0.29 (0.16, 0.41)	3.70E-05	0.038	-0.02(-0.12,0.09)	7.37E-01	0.992	EPHB4	1.00E+08	7	Island	chr7:100421295-100421515
cg15959529	0.38 (0.21, 0.55)	3.71E-05	0.038	0.04(-0.09,0.16)	5.55E-01	0.99	TRIM39;TRIM39	30309908	6	N_Shelf	chr6:30312837-30313419
cg00888154	0.36 (0.2, 0.51)	3.74E-05	0.038	0.07(-0.02,0.16)	1.35E-01	0.967	PIGG;PIGG	503399	4		
cg14422315	0.25 (0.14, 0.35)	3.74E-05	0.038	0.02(-0.06,0.1)	6.42E-01	0.992	CRTAC1	99664004	10		
cg16237262	0.41 (0.23, 0.59)	3.74E-05	0.038	0.07(-0.06,0.2)	2.71E-01	0.967	SLC9A4	1.03E+08	2		
cg03971344	0.31 (0.17, 0.44)	3.77E-05	0.038	-0.11(-0.21,-0.01)	3.32E-02	0.719	ANKRD11	89349124	16	S_Shore	chr16:89345463-89348521
cg20335293	-0.33 (-0.48, -0.19)	3.77E-05	0.038	0(-0.16,0.15)	9.76E-01	0.996		1.57E+08	2		
cg25599950	-0.41 (-0.6, -0.23)	3.77E-05	0.038	-0.02(-0.18,0.15)	8.43E-01	0.992	COBL	51259763	7		
cg00137629	0.27 (0.15, 0.39)	3.81E-05	0.039	0.02(-0.06,0.1)	6.65E-01	0.992	GUSB	65445384	7	N_Shore	chr7:65446771-65447340
cg01619490	0.46 (0.26, 0.67)	3.82E-05	0.039	0.03(-0.14,0.21)	7.03E-01	0.992		40724242	21	S_Shelf	chr21:40720185-40721625
cg17170190	0.29 (0.16, 0.42)	3.87E-05	0.039	0.01(-0.07,0.09)	8.56E-01	0.992	B4GALT2;B4GALT2	44445979	1	S_Shore	chr1:44444635-44445622
cg26011692	0.36 (0.2, 0.52)	3.87E-05	0.039	0.12(-0.03,0.27)	1.16E-01	0.967	LTBP2	75052601	14	Island	chr14:75052567-75052820
cg25527494	-0.38 (-0.55, -0.21)	3.88E-05	0.039	0.02(-0.14,0.18)	8.11E-01	0.992		1.01E+08	13	N_Shore	chr13:100641334-100642188
cg15504662	-0.44 (-0.64, -0.25)	3.90E-05	0.039	0.15(0.01,0.29)	3.76E-02	0.719	LIMK2;LIMK2;LIMK2	31644160	22		
cg18454685	-0.35 (-0.5, -0.19)	3.91E-05	0.039	0.13(0.02,0.24)	2.28E-02	0.719	CACNA1G;CACNA1G;(	48639239	17	Island	chr17:48636103-48639279
cg07147863	-0.31 (-0.44, -0.17)	3.95E-05	0.039	0.07(-0.08,0.21)	3.68E-01	0.968	FBXL18	5534548	7	Island	chr7:5534384-5534766
cg02823625	0.37 (0.21, 0.53)	3.96E-05	0.039	-0.03(-0.52,0.46)	9.08E-01	0.992	KLHL30	2.39E+08	2	N_Shore	chr2:239048071-239048454
cg14002714	0.37 (0.21, 0.53)	3.97E-05	0.039	0.03(-0.12,0.18)	6.75E-01	0.992	KTN1;KTN1;KTN1;KTN	56095810	14		
cg02369024	0.3 (0.17, 0.44)	4.06E-05	0.04	0.06(-0.02,0.13)	1.53E-01	0.967		63204500	7		
cg24892571	0.36 (0.2, 0.51)	4.06E-05	0.04	0.01(-0.06,0.09)	7.02E-01	0.992	HMGA2;HMGA2	66284828	12		
cg06980341	0.32 (0.18, 0.46)	4.07E-05	0.04	0.04(-0.08,0.17)	4.92E-01	0.989	ARHGEF10L	17874602	1		
cg21442876	0.47 (0.26, 0.68)	4.09E-05	0.04	0.02(-0.12,0.17)	7.51E-01	0.992	PET112L	1.53E+08	4		
cg05394800	-0.31 (-0.44, -0.17)	4.10E-05	0.04	-0.01(-0.2,0.18)	9.55E-01	0.993		50707050	13	N_Shore	chr13:50707585-50708019
cg13897122	0.41 (0.23, 0.6)	4.14E-05	0.04	0.02(-0.13,0.16)	8.12E-01	0.992	IL18RAP	1.03E+08	2		
cg07643314	-0.28 (-0.41, -0.16)	4.15E-05	0.04	0.06(-0.06,0.18)	3.00E-01	0.967	DAZL	16647343	3	S_Shore	chr3:16645948-16646259
cg02849103	-0.35 (-0.5, -0.19)	4.16E-05	0.04	0.05(-0.07,0.17)	4.14E-01	0.968		10350874	16		
cg11860777	-0.42 (-0.61, -0.23)	4.16E-05	0.04	-0.06(-0.17,0.05)	2.58E-01	0.967	IGFBP4	38609504	17		
cg16426715*	0.23 (0.13, 0.34)	4.19E-05	0.04	0.06(-0.02,0.14)	1.58E-01	0.967	GRK6;GRK6;GRK6	1.77E+08	5		
cg17240454	0.32 (0.18, 0.46)	4.21E-05	0.04	-0.06(-0.2,0.08)	4.02E-01	0.968	SPDEF	34524278	6		
cg13053653	-0.32 (-0.47, -0.18)	4.22E-05	0.04	0.03(-0.09,0.16)	6.10E-01	0.992	HOXD3	1.77E+08	2	N_Shore	chr2:177039551-177039951
ch.4.104561249F	0.31 (0.17, 0.45)	4.22E-05	0.04	-0.02(-0.14,0.1)	7.73E-01	0.992		1.04E+08	4		
cg23119977	0.39 (0.22, 0.57)	4.23E-05	0.04	0(-0.09,0.09)	9.72E-01	0.995	FRMD5	44213307	15		
cg06049367	0.38 (0.21, 0.55)	4.31E-05	0.041	-0.07(-0.22,0.09)	3.95E-01	0.968	INSC;INSC	15139573	11	S_Shelf	chr11:15136058-15136545
cg07059167	-0.46 (-0.67, -0.26)	4.33E-05	0.041	0.02(-0.1,0.14)	7.64E-01	0.992	PTOV1	50354224	19	Island	chr19:50353790-50355483
cg11696388	0.31 (0.17, 0.45)	4.33E-05	0.041	0(-0.08,0.08)	9.41E-01	0.992	ANK1;ANK1;ANK1;AN	41557540	8	N_Shore	chr8:41559255-41559609
cg07808859	-0.44 (-0.64, -0.24)	4.35E-05	0.041	.				1.35E+08	5		
cg02493644	-0.44 (-0.64, -0.24)	4.36E-05	0.041	0.06(-0.06,0.18)	3.32E-01	0.967		84816314	7	S_Shore	chr7:84814839-84816242
cg25971741	0.82 (0.45, 1.18)	4.36E-05	0.041	-0.03(-0.16,0.1)	6.06E-01	0.992	JARID2	15343116	6		
cg16121206	-0.51 (-0.74, -0.28)	4.38E-05	0.041	-0.12(-0.46,0.22)	4.82E-01	0.988	APOL2;APOL2	36636055	22		
cg04946387	0.91 (0.5, 1.31)	4.40E-05	0.041	-0.14(-0.52,0.24)	4.78E-01	0.988	B3GAT1;B3GAT1	1.34E+08	11	Island	chr11:134253449-134254151
cg07862554	-0.32 (-0.46, -0.18)	4.42E-05	0.041	0(-0.13,0.12)	9.39E-01	0.992	MME;MME;MME;MME	1.55E+08	3	N_Shore	chr3:154797398-154797988
cg27301231	-0.28 (-0.41, -0.16)	4.44E-05	0.041	0.02(-0.15,0.19)	7.95E-01	0.992	C3orf59	1.93E+08	3		
cg26222498	0.38 (0.21, 0.55)	4.47E-05	0.042	0.06(-0.06,0.18)	2.94E-01	0.967	INPP5A	1.34E+08	10	S_Shelf	chr10:134460274-134460479
cg04986304	0.35 (0.2, 0.51)	4.48E-05	0.042	-0.04(-0.15,0.06)	4.16E-01	0.968		54590530	1	S_Shelf	chr1:54587364-54587578
cg06810011	-0.3 (-0.44, -0.17)	4.51E-05	0.042	0.01(-0.15,0.16)	9.24E-01	0.992	IMMP2L	1.11E+08	7		
cg06884262	-0.25 (-0.36, -0.14)	4.54E-05	0.042	0.05(-0.09,0.19)	4.88E-01	0.988	KHSRP	6425638	19	Island	chr19:6425485-6425843
cg12612213*	0.28 (0.16, 0.41)	4.54E-05	0.042	0.1(-0.03,0.23)	1.22E-01	0.967	EPHB3	1.84E+08	3	N_Shore	chr3:184286935-184287329
cg02048317	0.29 (0.16, 0.42)	4.55E-05	0.042	-0.04(-0.21,0.13)	6.63E-01	0.992		80004967	17	N_Shelf	chr17:80008514-80010567
cg25463779	0.27 (0.15, 0.4)	4.55E-05	0.042	-0.01(-0.13,0.11)	8.98E-01	0.992	FAM101A	1.25E+08	12		
cg00378730	-0.27 (-0.4, -0.15)	4.57E-05	0.042	0.04(-0.08,0.16)	4.85E-01	0.988	CHAT;CHAT;CHAT;CH	50822650	10	Island	chr10:50822350-50822666
cg13357821	0.4 (0.22, 0.58)	4.58E-05	0.042	-0.11(-0.21,-0.02)	2.20E-02	0.719	CUX1;CUX1;CUX1	1.02E+08	7		
cg04641165	0.38 (0.21, 0.55)	4.62E-05	0.042	0.02(-0.1,0.14)	7.60E-01	0.992	MYLIP	16139117	6		
cg03490881	0.33 (0.18, 0.47)	4.63E-05	0.042	0.11(0,0.22)	4.28E-02	0.763	GIMAP6;GIMAP6;GIM/	1.50E+08	7		
cg06214983	0.31 (0.17, 0.44)	4.63E-05	0.042	0.08(-0.04,0.2)	1.74E-01	0.967	MANEAL;MANEAL	38259501	1	Island	chr1:38259095-38260427
cg13556548	-0.34 (-0.5, -0.19)	4.63E-05	0.042	0.09(-0.03,0.22)	1.39E-01	0.967	C1orf55	2.26E+08	1	S_Shore	chr1:226186803-226187336
cg19752565	0.27 (0.15, 0.39)	4.63E-05	0.042	0(-0.09,0.1)	9.38E-01	0.992	SCN7A	1.67E+08	2		
cg25760416	0.35 (0.19, 0.51)	4.63E-05	0.042	0.01(-0.15,0.17)	9.14E-01	0.992	HFM1	91871748	1	S_Shore	chr1:91869876-91870390
cg25226092	0.32 (0.18, 0.46)	4.67E-05	0.042	0.01(-0.12,0.13)	9.18E-01	0.992	MOBP;MOBP	39508863	3		
cg16196384	-0.31 (-0.45, -0.17)	4.69E-05	0.042	0.07(-0.05,0.19)	2.38E-01	0.967	PMPCB	1.03E+08	7	Island	chr7:102937857-102938141
cg09929564	-0.52 (-0.75, -0.29)	4.72E-05	0.042	0.19(0.02,0.36)	2.85E-02	0.719	ILF2	1.54E+08	1	S_Shore	chr1:153643318-153643529
cg27366964	-0.52 (-0.75, -0.29)	4.74E-05	0.042	.			C10orf11	77946906	10		
cg21509457	0.32 (0.18, 0.46)	4.75E-05	0.042	-0.01(-0.19,0.18)	9.40E-01	0.992	UBE3C	1.57E+08	7		
cg27431596	-0.45 (-0.66, -0.25)	4.76E-05	0.042	0.05(-0.06,0.16)	4.17E-01	0.968	ZNF697	1.20E+08	1	Island	chr1:120190377-120191251
cg02697721	0.33 (0.18, 0.48)	4.78E-05	0.042	-0.12(-0.3,0.06)	2.04E-01	0.967	FAM110A;FAM110A	819800	20	N_Shore	chr20:821674-822500
cg22795239*	0.38 (0.21, 0.55)	4.81E-05	0.043	0.01(-0.15,0.16)	9.25E-01	0.992	WNT16	1.21E+08	7		
cg16403344	-0.33 (-0.48, -0.18)	4.83E-05	0.043	0.07(-0.06,0.21)	2.87E-01	0.967	RPH3A;RPH3A	1.13E+08	12		
cg14510947	-0.47 (-0.68, -0.26)	4.87E-05	0.043	0.09(-0.04,0.22)	1.60E-01	0.967	SNORA76	62223546	17	Island	chr17:62223224-62224124
cg05624577	0.73 (0.4, 1.06)	4.91E-05	0.043	-0.23(-0.65,0.18)	2.73E-01	0.967		81411055	15	Island	chr15:81410715-81411067
cg07058988	0.37 (0.21, 0.54)	4.98E-05	0.043	0.02(-0.1,0.15)	7.24E-01	0.992		80297159	17		

	TAIWAN COHORT			NHBCS			CpG Information					
Name	Coef (CI)	P-value	FDR P-value	Coef (CI)	P-value	FDR P-value	Gene	MAPINFO	CHR	Relation to	CpG Islands	Name
cg23637124	-0.32 (-0.47, -0.18)	4.98E-05	0.043	0.16(0.04,0.27)	8.42E-03	0.719	SHC4;SHC4	49255455	15	Island	chr15:49254984-49255564	
cg10105110	0.32 (0.18, 0.47)	5.00E-05	0.043	0.02(-0.1,0.13)	7.94E-01	0.992	C10orf99	85940649	10			
cg15674514	0.33 (0.18, 0.48)	5.00E-05	0.043	-0.18(-0.35,-0.02)	3.21E-02	0.719		76674273	18	Island	chr18:76674209-76674797	
cg19750824	2.45 (1.35, 3.56)	5.03E-05	0.044	-0.8(-1.98,0.39)	1.84E-01	0.967		1.79E+08	1			
cg25189764*	0.48 (0.26, 0.69)	5.04E-05	0.044	0.01(-0.11,0.13)	8.81E-01	0.992	FYN;FYN	1.12E+08	6			
cg11233468	0.28 (0.15, 0.41)	5.11E-05	0.044	-0.05(-0.15,0.05)	3.40E-01	0.967	FUT6;FUT6	5831171	19	N_Shore	chr19:5831595-5832297	
cg13716760	0.31 (0.17, 0.45)	5.11E-05	0.044	0.01(-0.1,0.11)	9.08E-01	0.992		15371248	9			
cg26655764	0.33 (0.18, 0.47)	5.11E-05	0.044	-0.02(-0.17,0.12)	7.48E-01	0.992	CPNE7;CPNE7	89657805	16	N_Shelf	chr16:89661799-89662043	
cg09681360	0.43 (0.23, 0.62)	5.14E-05	0.044	0.09(-0.06,0.24)	2.20E-01	0.967	PHC3	1.70E+08	3	N_Shore	chr3:169898946-169899626	
cg15297799	-0.23 (-0.33, -0.13)	5.16E-05	0.044	0.02(-0.12,0.16)	8.09E-01	0.992	C18orf22	77794239	18	N_Shore	chr18:77794360-77794761	
cg10210510	-0.44 (-0.63, -0.24)	5.19E-05	0.044	0(-0.13,0.13)	9.62E-01	0.993	COL9A2	40771135	1	S_Shore	chr1:40769186-40769871	
cg15474407	-0.42 (-0.61, -0.23)	5.19E-05	0.044	0.07(-0.08,0.21)	3.57E-01	0.968	ARNTL2	27500572	12			
cg16047663	-0.39 (-0.56, -0.21)	5.21E-05	0.044	-0.08(-0.2,0.05)	2.17E-01	0.967	RNPEP	2.02E+08	1			
cg05137466	-0.48 (-0.7, -0.27)	5.22E-05	0.044	0.15(-0.01,0.31)	7.21E-02	0.893		34809811	9	Island	chr9:34809590-34810229	
cg02478369	1.72 (0.95, 2.5)	5.23E-05	0.044	0.06(-0.07,0.19)	3.55E-01	0.968		15083645	17			
cg07523741	0.35 (0.19, 0.51)	5.24E-05	0.044	-0.09(-0.29,0.11)	3.90E-01	0.968	PKD2L2	1.37E+08	5	Island	chr5:137224986-137225477	
cg23261846	-0.53 (-0.77, -0.29)	5.24E-05	0.044	0.02(-0.1,0.14)	7.53E-01	0.992		30591384	12			
cg09676630	0.4 (0.22, 0.58)	5.25E-05	0.044	0.02(-0.06,0.09)	6.61E-01	0.992	FBRSL1	1.33E+08	12	N_Shore	chr12:133102168-133102604	
cg22285621	-0.28 (-0.4, -0.15)	5.25E-05	0.044	0.02(-0.14,0.18)	8.08E-01	0.992	SSH3	67071322	11	Island	chr11:67070807-67071801	
cg26330076	0.32 (0.18, 0.47)	5.25E-05	0.044	-0.11(-0.28,0.05)	1.70E-01	0.967	FICD	1.09E+08	12	S_Shelf	chr12:108908702-108909475	
cg10991454	0.29 (0.16, 0.42)	5.30E-05	0.044	-0.03(-0.11,0.05)	4.34E-01	0.982	LCLAT1;LCLAT1	30834674	2			
cg05155965	-0.35 (-0.5, -0.19)	5.32E-05	0.044	0.07(-0.04,0.19)	2.08E-01	0.967	NR2F2;NR2F2	96873885	15	Island	chr15:96873408-96877721	
cg24863642*	0.31 (0.17, 0.45)	5.38E-05	0.045	0.1(-0.06,0.25)	2.25E-01	0.967	NDUFS2;NDUFS2	1.61E+08	1	Island	chr1:161171809-161172256	
cg24870483	0.35 (0.19, 0.51)	5.41E-05	0.045	-0.01(-0.11,0.09)	8.47E-01	0.992		2.43E+08	2	N_Shore	chr2:242805752-242806034	
cg26965718*	0.59 (0.32, 0.86)	5.41E-05	0.045	0.12(-0.03,0.27)	1.12E-01	0.967	HGS	79658957	17	N_Shelf	chr17:79662850-79663055	
cg03655389	0.31 (0.17, 0.44)	5.43E-05	0.045	-0.02(-0.13,0.09)	7.30E-01	0.992	TPPP	665188	5	Island	chr5:665187-665401	
cg21169267	0.37 (0.21, 0.54)	5.44E-05	0.045	-0.01(-0.18,0.16)	9.25E-01	0.992	FAM129B;FAM129B	1.30E+08	9			
cg05940691	-0.4 (-0.58, -0.22)	5.46E-05	0.045	0.07(-0.06,0.21)	2.64E-01	0.967	WDR64	2.42E+08	1			
cg24354818	0.81 (0.44, 1.17)	5.47E-05	0.045	0.57(0.11,1.03)	1.61E-02	0.719	TNFRSF6B;TNFRSF6B;	62328094	20	Island	chr20:62328013-62328558	
cg21610090	-0.26 (-0.37, -0.14)	5.52E-05	0.045	0(-0.13,0.13)	9.92E-01	0.997	CDKL4	39455859	2			
cg01501009	-0.41 (-0.6, -0.23)	5.53E-05	0.045	0.01(-0.13,0.15)	8.44E-01	0.992	ZNF835	57183268	19	Island	chr19:57182887-57183375	
cg00451102	-0.46 (-0.66, -0.25)	5.56E-05	0.045	-0.08(-0.22,0.05)	2.32E-01	0.967		21679284	10			
cg24158553	-0.38 (-0.55, -0.21)	5.57E-05	0.045	-0.03(-0.18,0.11)	6.46E-01	0.992	GABPB1;GABPB1;FLJ1	50647608	15	Island	chr15:50646437-50647742	
cg10929866	0.59 (0.32, 0.85)	5.58E-05	0.045	-0.02(-0.29,0.26)	9.13E-01	0.992	SNX9	1.58E+08	6			
cg12354377	-0.29 (-0.42, -0.16)	5.59E-05	0.045	-0.07(-0.22,0.08)	3.82E-01	0.968	ANK3	62149557	10			
cg22175345	0.41 (0.22, 0.59)	5.59E-05	0.045	-0.05(-0.26,0.15)	6.26E-01	0.992		17194180	4			
cg07830254	0.28 (0.15, 0.4)	5.60E-05	0.045	-0.14(-0.26,-0.02)	2.31E-02	0.719	C7orf34	1.43E+08	7			
cg16379910	-0.32 (-0.47, -0.18)	5.64E-05	0.045	0.08(-0.05,0.21)	2.39E-01	0.967		78636730	14			
cg02342415*	0.34 (0.19, 0.5)	5.68E-05	0.046	0.07(-0.15,0.3)	5.21E-01	0.989	IQSEC1;IQSEC1	12940876	3	Island	chr3:12940753-12941134	
cg00976381	0.42 (0.23, 0.6)	5.69E-05	0.046	0.11(-0.07,0.28)	2.27E-01	0.967	SLC23A1;SLC23A1	1.39E+08	5	Island	chr5:138713814-138714340	
cg02348830	0.39 (0.21, 0.56)	5.71E-05	0.046	0.06(-0.05,0.16)	3.02E-01	0.967	TREX1;ATRIP;ATRIP;T	48506168	3			
cg07592519	-0.58 (-0.84, -0.32)	5.74E-05	0.046	0.08(-0.05,0.21)	2.14E-01	0.967	KCNQ1;KCNQ1	2735101	11			
cg16453617	0.37 (0.2, 0.54)	5.76E-05	0.046	-0.02(-0.13,0.08)	6.48E-01	0.992	CACNA1H;CACNA1H	1257884	16	S_Shelf	chr16:1254108-1254375	
cg19547200	-0.71 (-1.03, -0.39)	5.76E-05	0.046	-0.02(-0.14,0.11)	7.96E-01	0.992	SERHL	42896249	22	N_Shore	chr22:42896636-42897041	
cg19307180	-0.43 (-0.62, -0.23)	5.77E-05	0.046	-0.09(-0.23,0.06)	2.45E-01	0.967	DGKH;DGKH	42803600	13			
cg07296387	0.5 (0.27, 0.72)	5.78E-05	0.046	-0.14(-0.33,0.05)	1.48E-01	0.967	CDH22	44838981	20	Island	chr20:44838887-44839204	
cg26530061	-0.33 (-0.47, -0.18)	5.79E-05	0.046	0.12(-0.01,0.24)	7.56E-02	0.893		27265522	7	S_Shore	chr7:27265158-27265493	
cg08744727	0.41 (0.22, 0.59)	5.80E-05	0.046	-0.01(-0.19,0.17)	9.11E-01	0.992	ZNF876P	206112	4	N_Shore	chr4:206377-206892	
cg12083893	0.37 (0.2, 0.54)	5.81E-05	0.046	0(-0.09,0.1)	9.76E-01	0.996		91111929	10			
cg00965154	0.42 (0.23, 0.62)	5.82E-05	0.046	-0.01(-0.17,0.15)	9.07E-01	0.992	SMTN;SMTN;SMTN	31485117	22	S_Shelf	chr22:31480774-31481373	
cg12072690	0.39 (0.21, 0.56)	5.82E-05	0.046	-0.06(-0.2,0.08)	4.10E-01	0.968	MCF2L;MCF2L	1.14E+08	13	Island	chr13:113750486-113751565	
cg08490349	0.26 (0.14, 0.38)	5.83E-05	0.046	0.03(-0.07,0.13)	5.18E-01	0.989	MPRIP;MPRIP	17086207	17			
cg16912910	0.32 (0.18, 0.47)	5.83E-05	0.046	0.01(-0.16,0.18)	8.94E-01	0.992		29497165	6			
cg02082252*	0.41 (0.22, 0.6)	5.85E-05	0.046	0.06(-0.06,0.19)	3.30E-01	0.967	EHMT2;EHMT2	31866286	6	N_Shore	chr6:31867691-31867957	
cg23806084	0.51 (0.28, 0.74)	5.86E-05	0.046	0.04(-0.05,0.14)	3.75E-01	0.968	PSMB9;TAP1;PSMB9;T	32821605	6	Island	chr6:32820849-32822370	
cg12084011	0.46 (0.25, 0.66)	5.89E-05	0.046	0.03(-0.08,0.13)	6.10E-01	0.992	BCAP29;BCAP29;BCAP	1.07E+08	7	Island	chr7:107220344-107221075	
cg20802509	0.35 (0.19, 0.5)	5.91E-05	0.046	-0.02(-0.12,0.08)	6.81E-01	0.992	URB2	2.30E+08	1	S_Shelf	chr1:229761128-229762299	
cg15513163	-0.29 (-0.42, -0.16)	5.95E-05	0.046	0.18(0.04,0.32)	1.40E-02	0.719		70034139	1	Island	chr1:70032967-70034495	
cg06624525	0.29 (0.16, 0.42)	5.96E-05	0.046	-0.06(-0.18,0.05)	2.72E-01	0.967		35001084	15			
cg08615818	0.37 (0.2, 0.54)	5.96E-05	0.046	-0.01(-0.23,0.22)	9.54E-01	0.993	HCCA2	1531515	11			
cg08166750	-0.35 (-0.52, -0.19)	5.99E-05	0.046	-0.05(-0.2,0.09)	4.64E-01	0.988	EPS8L1;EPS8L1	55598443	19	Island	chr19:55597977-55598887	
cg09962824	0.33 (0.18, 0.48)	5.99E-05	0.046	0.02(-0.06,0.11)	5.54E-01	0.99	CBS	44479417	21	N_Shore	chr21:44480559-44480772	
cg05908587	-0.31 (-0.46, -0.17)	6.00E-05	0.046	0.02(-0.11,0.15)	7.47E-01	0.992	GALNTL1;GALNTL1	69726251	14	Island	chr14:69726250-69728393	
cg10706989	-0.44 (-0.65, -0.24)	6.01E-05	0.046	0.12(-0.01,0.25)	7.26E-02	0.893	C13orf34;C13orf37;C13	73302078	13	Island	chr13:73301230-73302154	
cg12613382	0.31 (0.17, 0.45)	6.03E-05	0.046	0.01(-0.07,0.09)	8.28E-01	0.992		1.33E+08	12	S_Shore	chr12:132663322-132663748	
cg23947450	0.28 (0.15, 0.41)	6.03E-05	0.046	-0.03(-0.15,0.1)	6.67E-01	0.992	UNC84A;UNC84A	900037	7			
cg05055326	-0.36 (-0.53, -0.2)	6.11E-05	0.046	0(-0.13,0.14)	9.53E-01	0.993	HCG9	29945080	6	Island	chr6:29944402-29945169	
cg26990023	0.4 (0.22, 0.58)	6.13E-05	0.046	0.02(-0.13,0.17)	8.03E-01	0.992	SMOC2;SMOC2	1.69E+08	6	N_Shore	chr6:168972516-168974109	
cg14526039	0.34 (0.18, 0.49)	6.17E-05	0.047	0.01(-0.09,0.1)	8.97E-01	0.992	KLHDC8B	49210221	3	S_Shore	chr3:49208629-49209196	
cg21495568	0.37 (0.2, 0.54)	6.19E-05	0.047	-0.03(-0.16,0.1)	6.64E-01	0.992	SAP18	21715145	13	S_Shore	chr13:21714289-21715115	
cg01707127	0.37 (0.2, 0.54)	6.22E-05	0.047	0(-0.12,0.11)	9.56E-01	0.993	LMF1	947051	16	S_Shore	chr16:945437-946420	
cg12934466	-0.31 (-0.45, -0.17)	6.24E-05	0.047	-0.02(-0.16,0.12)	7.45E-01	0.992	ZNF397;ZNF397	32820979	18	Island	chr18:32820785-32821203	
cg23155333	0.37 (0.2, 0.54)	6.24E-05	0.047	-0.09(-0.19,0.01)	7.41E-02	0.893		35323673	19	Island	chr19:35323531-35323952	
cg03914237	0.38 (0.21, 0.56)	6.25E-05	0.047	0.07(-0.07,0.22)	3.29E-01	0.967	THSD7A	11558886	7			
cg08180572	-0.51 (-0.74, -0.28)	6.29E-05	0.047	0.04(-0.17,0.26)	6.94E-01	0.992	SPNS1;SPNS1;SPNS1;S	28985143	16	N_Shore	chr16:28985881-28986661	
cg12384932	0.22 (0.12, 0.33)	6.31E-05	0.047	-0.01(-0.16,0.14)	9.30E-01	0.992	NALCN	1.02E+08	13			
cg07008591	0.39 (0.21, 0.56)	6.34E-05	0.047	0.01(-0.11,0.13)	8.75E-01	0.992	TEAD1	12809661	11			
cg07818422	-0.3 (-0.43, -0.16)	6.40E-05	0.048	-0.02(-0.12,0.09)	7.44E-01	0.992	WDFY2	52301946	13			
cg26272907	0.71 (0.38, 1.03)	6.43E-05	0.048	0.17(-0.06,0.41)	1.50E-01	0.967	TUBGCP3	1.13E+08	13	N_Shelf	chr13:113201706-113201927	
cg16041798	0.34 (0.19, 0.5)	6.44E-05	0.048	0.12(-0.03,0.28)	1.25E-01	0.967	MTHFD1	64905375	14			
cg05946856	-0.31 (-0.45, -0.17)	6.48E-05	0.048	0.1(-0.13,0.33)	3.82E-01	0.968		1.57E+08	6			
cg07042546	0.28 (0.15, 0.41)	6.50E-05	0.048	-0.05(-0.15,0.05)	3.43E-01	0.967	RNF214;RNF214	1.17E+08	11	S_Shore	chr11:117102742-117103452	
cg03955354	0.35 (0.19, 0.51)	6.51E-05	0.048	0.08(-0.04,0.2)	1.98E-01	0.967		2925039	16			
cg03904876	-0.25 (-0.37, -0.14)	6.63E-05	0.049	0.08(-0.03,0.19)	1.44E-01	0.967		54322092	12	S_Shore	chr12:54321301-54321721	
ch.7.109930408F	-0.33 (-0.49, -0.18)	6.64E-05	0.049	-0.04(-0.27,0.19)	7.42E-01	0.992		1.10E+08	7			
cg05341199	0.34 (0.18, 0.49)	6.67E-05	0.049	0.05(-0.09,0.19)	4.81E-01	0.988	GPR133	1.32E+08	12	N_Shelf	chr12:131605558-131605767	
cg02286715	-0.24 (-0.35, -0.13)	6.69E-05	0.049	0.04(-0.08,0.16)	4.75E-01	0.988	PPP2R2C	6473881	4	Island	chr4:6472175-6474534	
cg00791851	-0.31 (-0.45, -0.16)	6.71E-05	0.049	0(-0.46,0.46)	9.95E-01	0.997		76518896	1			

	TAIWAN COHORT			NHBCS			CpG Information				
Name	**Coef (CI)**	**P-value**	**FDR P-value**	**Coef (CI)**	**P-value**	**FDR P-value**	**Gene**	**MAPINFO**	**CHR**	**Relation to**	**CpG Islands Name**
cg08198265*	0.57 (0.31, 0.83)	6.77E-05	0.049	0.01(-0.09,0.15)	8.75E-01	0.992	BST1	15708451	4	S_Shelf	chr4:15704640-15705000
cg19850333	-0.36 (-0.53, -0.2)	6.77E-05	0.049	0.04(-0.06,0.15)	4.28E-01	0.977	CCRL2;CCRL2	46448579	3		
cg00049047	-0.37 (-0.54, -0.2)	6.78E-05	0.049	0.06(-0.06,0.19)	3.34E-01	0.967	GDNF	37838425	5	Island	chr5:37836747-37840726
cg25588480	0.41 (0.22, 0.6)	6.78E-05	0.049	-0.05(-0.19,0.09)	4.82E-01	0.988	MINK1;MINK1;MINK1;	4763240	17		
cg18132007	-0.21 (-0.3, -0.11)	6.79E-05	0.049	0.11(0.01,0.2)	2.86E-02	0.719	TP53I11;TP53I11	44972684	11	Island	chr11:44971048-44972685
cg16364152	-0.43 (-0.63, -0.23)	6.80E-05	0.049	-0.01(-0.13,0.1)	8.27E-01	0.992	RP5-1022P6.2	5591818	20	Island	chr20:5591490-5591875
cg14794023	0.27 (0.15, 0.4)	6.81E-05	0.049	0.1(-0.03,0.23)	1.18E-01	0.967	POM121L12	53103576	7	Island	chr7:53103275-53103801
cg21678813	-0.32 (-0.47, -0.17)	6.81E-05	0.049	0.13(0.02,0.24)	2.64E-02	0.719		20229879	11		
cg23319696	0.45 (0.24, 0.65)	6.82E-05	0.049	0.09(-0.1,0.27)	3.46E-01	0.968		2111870	5	Island	chr5:2111836-2112484
cg06935052	0.29 (0.16, 0.42)	6.84E-05	0.049	-0.04(-0.15,0.07)	4.74E-01	0.988	SMARCB1;SMARCB1	24176449	22		
cg19254118	0.44 (0.24, 0.64)	6.85E-05	0.049	-0.01(-0.15,0.14)	9.26E-01	0.992	ADAT3;SCAMP4	1907972	19	N_Shore	chr19:1908118-1908509
cg19724043	0.27 (0.14, 0.39)	6.86E-05	0.049	0.02(-0.12,0.17)	7.39E-01	0.992		15380395	21	N_Shelf	chr21:15383530-15383813
cg27640763	0.46 (0.25, 0.68)	6.89E-05	0.049	0.17(-0.02,0.37)	7.92E-02	0.893	LUM	91503109	12		
cg00217795	0.3 (0.16, 0.44)	6.90E-05	0.049	-0.03(-0.18,0.11)	6.47E-01	0.992	DIO2;DIO2;DIO2	80677688	14		
cg09568216	-0.52 (-0.76, -0.28)	6.90E-05	0.049	-0.04(-0.13,0.06)	4.48E-01	0.988	NPR3	32779925	5		
cg12523924	-0.33 (-0.49, -0.18)	6.90E-05	0.049	0.13(0.01,0.25)	3.39E-02	0.719		63255359	5	Island	chr5:63255044-63255407
cg12241963	0.27 (0.14, 0.39)	6.91E-05	0.049	0.02(-0.14,0.17)	8.19E-01	0.992		33807279	6		
cg16275882	-0.26 (-0.38, -0.14)	6.91E-05	0.049	0.13(0.02,0.24)	1.69E-02	0.719		1288184	7	S_Shore	chr7:1286022-1287658
cg16703934	0.3 (0.16, 0.44)	6.91E-05	0.049	-0.07(-0.2,0.06)	2.71E-01	0.967	TSPAN33	1.29E+08	7	S_Shore	chr7:128784695-128785096
cg08622675	0.43 (0.23, 0.63)	6.92E-05	0.049	0.02(-0.11,0.16)	7.36E-01	0.992	KDELR2;KDELR2	6524998	7	S_Shore	chr7:6523064-6523897
cg07990873	0.27 (0.15, 0.4)	6.93E-05	0.049	-0.03(-0.14,0.08)	6.11E-01	0.992	ZNF671	58235167	19	N_Shelf	chr19:58238585-58239028
cg02816732	0.38 (0.2, 0.55)	6.94E-05	0.049	-0.12(-0.27,0.03)	1.13E-01	0.967	TNS3	47384337	7		
cg02333281	0.46 (0.25, 0.67)	6.96E-05	0.049	-0.06(-0.27,0.15)	5.93E-01	0.992		6636970	10		
cg26581228	0.51 (0.27, 0.74)	6.98E-05	0.049	-0.03(-0.13,0.06)	5.19E-01	0.989	TRERF1	42326264	6		
cg05310249	-0.43 (-0.62, -0.23)	6.99E-05	0.049	-0.02(-0.12,0.09)	7.34E-01	0.992	NKX2-6	23560590	8	Island	chr8:23559838-23560591
cg07158747	-0.4 (-0.58, -0.21)	6.99E-05	0.049	.				91196488	1	S_Shelf	chr1:91190489-91192804
cg02487202	0.31 (0.17, 0.45)	7.02E-05	0.049	-0.04(-0.12,0.05)	4.02E-01	0.968	ANKRD11	89358232	16		
cg06896857	0.34 (0.18, 0.5)	7.03E-05	0.049	0.04(-0.07,0.15)	4.74E-01	0.988	PPP1R15A;PPP1R15A	49375797	19	Island	chr19:49375484-49375928
cg00079219	-0.29 (-0.43, -0.16)	7.04E-05	0.049	0.08(-0.05,0.22)	2.03E-01	0.967	HOTAIR	54360131	12	N_Shore	chr12:54360374-54360660
cg08391419	0.31 (0.17, 0.46)	7.04E-05	0.049	0.01(-0.09,0.1)	9.14E-01	0.992	STK32C	1.34E+08	10	S_Shelf	chr10:134071971-134072193
cg05819837	0.35 (0.19, 0.5)	7.07E-05	0.049	0.07(-0.01,0.14)	8.29E-02	0.917	CUX1;CUX1;CUX1	1.02E+08	7		
cg09187936	0.23 (0.12, 0.33)	7.07E-05	0.049	-0.02(-0.12,0.09)	7.54E-01	0.992	SETD1B	1.22E+08	12	S_Shelf	chr12:122265374-122265954
cg04406115	0.32 (0.17, 0.46)	7.09E-05	0.049	-0.02(-0.16,0.13)	8.28E-01	0.992	KDM4B	5065640	19	Island	chr19:5065639-5065919
cg11175091	0.45 (0.24, 0.65)	7.09E-05	0.049	-0.02(-0.18,0.15)	8.47E-01	0.992	MIR1243;ANK2;ANK2;/	1.14E+08	4		
cg15266969	0.29 (0.16, 0.42)	7.10E-05	0.049	0.07(-0.05,0.2)	2.23E-01	0.967	SLC22A12;SLC22A12	64369352	11		
cg00704664	1.15 (0.62, 1.68)	7.13E-05	0.049	1.21(0.02,2.4)	4.55E-02	0.776	CDH4	60500578	20	N_Shore	chr20:60501966-60502173
cg00269725	0.73 (0.39, 1.07)	7.23E-05	0.049	-0.03(-0.22,0.16)	7.31E-01	0.992		1.57E+08	6		
cg20918219	-0.22 (-0.32, -0.12)	7.23E-05	0.049	0.06(-0.02,0.15)	1.55E-01	0.967	SCARA3;SCARA3	27493854	8	S_Shelf	chr8:27490959-27491775
cg08557393	0.33 (0.18, 0.48)	7.24E-05	0.049	-0.06(-0.19,0.07)	3.31E-01	0.967	DOK4	57521521	16		
cg25338036	0.47 (0.26, 0.69)	7.24E-05	0.049	-0.02(-0.15,0.1)	6.93E-01	0.992	CSMD1	3047536	8		
cg03131730	0.27 (0.14, 0.39)	7.29E-05	0.049	0.06(-0.03,0.15)	2.01E-01	0.967	CCDC42;CCDC42	8638810	17		
cg05129295	0.39 (0.21, 0.56)	7.29E-05	0.049	0.02(-0.1,0.13)	7.87E-01	0.992		1316294	8		
cg01592801	-0.34 (-0.5, -0.18)	7.30E-05	0.049	-0.14(-0.37,0.1)	2.55E-01	0.967	KCNS2	99438942	8	Island	chr8:99438692-99440425
cg10130718	0.54 (0.29, 0.79)	7.35E-05	0.049	0.08(-0.34,0.5)	7.00E-01	0.992	DMRTB1	53925368	1	Island	chr1:53925191-53926228
cg03897139	0.31 (0.17, 0.45)	7.39E-05	0.05	0.03(-0.08,0.14)	5.50E-01	0.99	DPYSL4	1.34E+08	10	N_Shore	chr10:134019500-134019776
cg05965745	-0.41 (-0.59, -0.22)	7.39E-05	0.05	-0.05(-0.3,0.21)	7.18E-01	0.992	PRDM16;PRDM16	3077798	1	N_Shelf	chr1:3080934-3081292
cg13205528	-0.49 (-0.72, -0.26)	7.44E-05	0.05	-0.03(-0.25,0.19)	7.70E-01	0.992		2705849	1	N_Shore	chr1:2706025-2706961
cg10221365	0.28 (0.15, 0.4)	7.45E-05	0.05	0.06(-0.11,0.22)	5.02E-01	0.989	JMJD5;JMJD5	27214422	16	N_Shore	chr16:27214772-27215678
cg02493798	1.11 (0.6, 1.63)	7.50E-05	0.05	0.19(-0.37,0.76)	5.01E-01	0.989	ALOX12	6899577	17	Island	chr17:6898820-6900427
cg12401679	0.27 (0.14, 0.39)	7.51E-05	0.05	0.01(-0.15,0.17)	9.08E-01	0.992	LOC619207	1.35E+08	10	N_Shore	chr10:135270783-135271061
cg17943647	-0.55 (-0.81, -0.3)	7.51E-05	0.05	0.05(-0.09,0.18)	5.04E-01	0.989	TRIM2;TRIM2	1.54E+08	4		

Footnote: This results were obtained by fitting Robust regression with the M-values of DNA-methylation as response and log10 of maternal urinary arsenic level adjusting for Urinary creatinine, child's gender, batch effect, mother's age, mother's pre-pregnancy body mass index (BMI), mother's education level, and cell proportions of 6 cell types. M-values are logit of DNA methylation, defined as log2 [β/(1-β)]. CpG sites in bold showed consistent association with Total urinary arsenic level of mother in NHBCS, and CpGs marked with "*" are the CpGs such that their corresponding genes are in the identified KEGG pathways.

Table A1.2

	Lymphocytes				Myeloid Cells	
Subject	**CD8T**	**CD4T**	**NK**	**Bcell**	**Mono**	**Gran**
1	0.086339	0.161789	0.03705	0.125422	0.108936	0.535165
2	0.071657	0.128531	0.025776	0.120749	0.130483	0.579767
3	0.055667	0.117509	0.112656	0.106542	0.120631	0.523271
4	0.094488	0.105196	0.035431	0.098882	0.114146	0.585546
5	0.127137	0.144023	0.0091	0.077574	0.080921	0.568545
6	0.114757	0.171282	0	0.243679	0.127776	0.405114
7	0.221126	0.353146	0	0.191755	0.044692	0.219486
8	0.106903	0.159796	0.021962	0.125591	0.098902	0.526875
9	0.141326	0.176845	0.044873	0.108733	0.126921	0.440781
10	0.110878	0.195382	0	0.23698	0.128623	0.343273
11	0.113305	0.185754	0	0.207857	0.103421	0.4232
12	0.103108	0.306985	1.07E-19	0.046686	0	0.558618
13	0.056519	0.128894	0.064005	0.071728	0.109834	0.58592
14	0.039974	0.109659	0.130263	0.117196	0.080217	0.583316
15	0.131113	0.11647	0.020356	0.058528	0.07092	0.625321
16	0.093583	0.165071	0	0.093822	0.087372	0.581508
17	0.143037	0.104781	0	0.243772	0.1317	0.4711
18	0.091523	0.278061	0.116745	0.146964	0.100536	0.276378
19	0.10395	0.219228	0.014414	0.130644	0.104624	0.467834
20	0.042257	0.123753	0.082375	0.133762	0.142382	0.499199
21	0.124195	0.157698	0	0.131555	0.086831	0.531584
22	0.090689	0.155113	0.12881	0.177518	0.087908	0.419279
23	0.034937	0.096715	0.043709	0.124229	0.1107	0.63069
24	0.070149	0.180137	0.077627	0.126873	0.139379	0.410679
25	0.034482	0.135348	0.060564	0.141314	0.123158	0.560727
26	0.064729	0.038857	0.0107	0.061138	0.110737	0.73377
27	0.030939	0.145214	0.042139	0.103386	0.075217	0.63486
28	0.069561	0.176162	0.012068	0.261309	0.129053	0.39821
29	0.077367	0.200833	0	0.153033	0.084993	0.516621
30	0.046674	0.074382	0.060829	0.086505	0.098737	0.669492
31	0.061626	0.061423	0.024778	0.126057	0.113859	0.648363
32	0.124063	0.137529	0.070476	0.151346	0.218414	0.361695
33	0.115761	0.206974	0	0.178757	0.111402	0.453671
34	0.138564	0.128998	0.031599	0.140466	0.092372	0.519693
35	0.100378	0.128192	0.015141	0.175882	0.154002	0.461204
36	0.083233	0.119729	0.021163	0.19733	0.104917	0.531799
37	0.092963	0.07327	0.058152	0.058765	0.081584	0.686868
38	0.166116	0.326477	-1.39E-17	0.143829	0.073455	0.324551
39	0.11892	0.153317	0	0.254617	0.08274	0.421659
40	0.02767	0.085303	0.082031	0.08386	0.155102	0.5797
41	0.070268	0.05107	0.026707	0.08257	0.106381	0.70062
42	0.072235	0.153685	0.063122	0.143498	0.130354	0.495626
43	0.020794	0.118555	0.159978	0.148758	0.17757	0.401548
44	0.07095	0.161602	0	0.086025	0.115794	0.590266
45	0.016839	0.112772	0.045162	0.222292	0.222783	0.414852

	Lymphocytes				Myeloid Cells	
Subject	**CD8T**	**CD4T**	**NK**	**Bcell**	**Mono**	**Gran**
46	0.131383	0.22382	-6.94E-18	0.154252	0.121875	0.404513
47	0.02962	0.021543	0.020384	0.069937	0.104432	0.771136
48	0.155763	0.175066	0.117977	0.173808	0.076379	0.394373
49	0.090426	0.190173	0.057454	0.152837	0.073335	0.508091
50	0.048113	0.115364	0.029918	0.118548	0.17001	0.538238
51	0.120098	0.269878	0.018139	0.141287	0.126483	0.349225
52	0.057611	0.18277	0.101527	0.152387	0.11435	0.423623
53	0.120721	0.165444	0.042014	0.135022	0.11225	0.471869
54	0.103501	0.131154	0.053606	0.115437	0.14038	0.519567
55	0.075224	0.153331	3.47E-18	0.095834	0.111252	0.587624
56	0.068445	0.11561	0.123447	0.144862	0.208628	0.377559
57	0.086244	0.326938	0	0.168179	0.051681	0.393962
58	0.042673	0.091076	0.033382	0.097413	0.115793	0.641253
59	0.045027	0.060593	0.032034	0.087649	0.096683	0.711222
60	0.034927	0.03991	0.022319	0.105901	0.128001	0.696177
61	0.066205	0.124148	0	0.129001	0.169022	0.53968
62	0.091891	0.139212	0.011755	0.147737	0.129403	0.536163
63	0.087555	0.099459	0.099699	0.18809	0.108163	0.507678
64	0	0.121129	0.212002	0.121968	0.083253	0.507372

Footnote: Cell proportions for 6 cell types, calculated using *estimatecellcounts* in R-package *minfi* using the DNA methylation dataset for 64 subjects in this study

Table A2.1

	Taiwanese				IoW				CpG Information				
CpG	Estimate	StdErr	RAW_P	ahoc_p	Estimate	StdErr	RAW_P	ahoc_p	UCSC_Re	UCSC_Re	CHR	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island
cg05712903	-2.3861	0.3144	3.63E-07	0.000117	1.2484	0.2754	0.000258	0.044319	RUFY1	TSS200	5	chr5:178977298-178978291	Island
cg07523741	-2.1622	0.4112	4.49E-05	0.014399	1.9706	0.4511	0.00037	0.063655	PKD2L2	Body	5	chr5:137224986-137225477	Island
cg26624744*	-2.65	0.5523	0.000125	0.040107	-2.179	0.5091	0.000451	0.077528	CYFIP2;C	TSS1500;'	5	chr5:156692812-156693749	N_Shore
cg08392484*	-2.5938	0.455	1.70E-05	0.005467	-1.4976	0.3767	0.000887	0.152492	ELL2	Body	5		
cg13984351*	-1.4511	0.2642	2.68E-05	0.008604	-1.6858	0.4342	0.00109	0.187548	BMI1	TSS200	10		
cg07945323	-2.143	0.4003	3.64E-05	0.011671	1.7933	0.4647	0.001149	0.197644	KRTAP6-:	TSS1500;'	21		
cg17207545*	-1.8271	0.3376	3.19E-05	0.010242	-0.7231	0.191	0.001351	0.232393			7	chr7:102134871-102135575	N_Shore
cg26015416*	1.2691	0.2662	0.000134	0.042938	0.5849	0.158	0.001629	0.280265	NTNG1;N	3'UTR;3'U	1	chr1:108023232-108023442	S_Shore
cg13593553	-1.362	0.2633	5.43E-05	0.017415	1.6389	0.4717	0.002706	0.465396	LOC2833:	TSS200;B	12		
cg25316853	-1.0846	0.2134	6.62E-05	0.021241	0.8909	0.2578	0.002822	0.485306	SLC1A3;S	TSS200;T	5		
cg04521724*	-2.4301	0.4792	6.78E-05	0.021754	-1.1768	0.3431	0.002987	0.513709			2		
cg08535184	-4.0804	0.8616	0.000144	0.046143	0.9926	0.3033	0.004233	0.72815	ZDHHC1	TSS1500	1	chr1:27153033-27153243	N_Shore
cg13778815	-2.1845	0.4165	4.61E-05	0.014811	0.9933	0.3263	0.006983	0.996133	FRG2	TSS1500	4		
cg22017777	-2.0933	0.386	3.12E-05	0.010023	1.1193	0.3766	0.008159	0.996133	EPB41L4/	Body	5		
cg26785949	-1.433	0.3041	0.000152	0.04876	1.0951	0.3799	0.009909	0.996133	CTU1	TSS1500	19	chr19:51611423-51612176	S_Shore
cg09856295	-3.48	0.6146	1.85E-05	0.005943	0.6376	0.2217	0.010065	0.996133	VPS39	TSS1500	15	chr15:42500160-42500561	S_Shore
cg16658180*	-1.9881	0.4214	0.00015	0.048141	-1.8384	0.6561	0.011787	0.996133			1		
cg16119772*	-2.3622	0.4373	3.26E-05	0.010475	-0.9321	0.3327	0.011793	0.996133			5		
cg14820221	-1.1354	0.22	5.57E-05	0.017878	0.1101	0.03958	0.012323	0.996133	SAFB2	Body	19	chr19:5593709-5594194	Island
cg24349819	0.996	0.1852	3.44E-05	0.011029	-0.7218	0.2614	0.012844	0.996133	ARHGAP:	Body;Bod	17	chr17:43472527-43474343	S_Shelf
cg17818792	-1.6083	0.3368	0.000131	0.042189	0.9657	0.3555	0.014142	0.996133			17		
cg01750170*	-1.6064	0.2863	2.07E-05	0.006632	-0.5702	0.2123	0.015091	0.996133			7	chr7:155595692-155599414	N_Shore
cg02830936*	-1.621	0.299	3.13E-05	0.010053	-0.5639	0.211	0.015533	0.996133	MTMR9L	TSS1500	1	chr1:32705524-32707141	S_Shore
cg25471923*	-2.0163	0.3834	4.47E-05	0.014358	-1.2814	0.4809	0.015796	0.996133			7	chr7:27238690-27240311	S_Shore
cg11506157*	-2.3942	0.4878	9.76E-05	0.031315	-1.2242	0.4613	0.016153	0.996133			5		
cg00565679	-1.9808	0.3723	3.91E-05	0.012541	0.8423	0.3183	0.01642	0.996133			12		
cg22590761*	-2.1274	0.427	8.28E-05	0.026565	-0.5393	0.2048	0.01686	0.996133	LOXL1;L(	5'UTR;1st	15	chr15:74218696-74220373	Island
cg05671241	-1.8796	0.391	0.000123	0.039343	1.3449	0.5119	0.017081	0.996133			3	chr3:126373571-126374142	Island
cg07708818*	-1.1902	0.2364	7.34E-05	0.023565	-1.1338	0.4412	0.019272	0.996133	GALNS	Body	16	chr16:88884168-88884523	N_Shore
cg01394833	-1.4305	0.3015	0.000141	0.045312	0.8815	0.3451	0.019916	0.996133			3	chr3:193587352-193587889	N_Shelf
ch_7_73585	-0.4311	0.08591	7.64E-05	0.024538	0.4027	0.1586	0.020539	0.996133					
cg08935238	-1.5733	0.2975	4.20E-05	0.013478	0.9927	0.4023	0.023858	0.996133	PCDHGA4;	Body;Body	5	chr5:140798757-140799359	N_Shore
cg04596842	-1.8717	0.3865	0.000113	0.036286	0.6421	0.2608	0.024126	0.996133			14	chr14:19685063-19687329	Island
cg10343024*	-1.5232	0.2483	6.74E-06	0.002164	-1.4319	0.587	0.02529	0.996133	PJA2	5'UTR	5	chr5:108744926-108745640	Island
cg26931526*	-2.2768	0.4607	9.06E-05	0.029069	-1.1845	0.4972	0.028445	0.996133			1	chr1:8277195-8277822	N_Shore
cg20411075*	1.2781	0.2118	8.33E-06	0.002673	0.6121	0.2603	0.030282	0.996133			13	chr13:21678416-21678648	N_Shore
cg17535691	1.2657	0.2645	0.000129	0.041364	-0.4991	0.213	0.030805	0.996133	TLX1NB	5'UTR	10	chr10:102882977-102883551	N_Shore
cg01443832*	-1.2298	0.2321	4.09E-05	0.013142	-0.429	0.1865	0.033596	0.996133	LOC64676	TSS1500	7	chr7:29724188-29725436	Island
cg05132568*	-1.6437	0.3166	5.19E-05	0.016651	-1.037	0.4517	0.033898	0.996133	KIAA2013	Body	1	chr1:11985643-11986634	N_Shelf
cg05955566*	-2.1435	0.4098	4.76E-05	0.015267	-1.3319	0.618	0.044932	0.996133	IKBIP;APAI	Body;TSS1	12	chr12:99038272-99039483	Island
cg00787013	-2.852	0.5844	0.000104	0.033345	1.0435	0.4845	0.045054	0.996133			10		
cg04614823*	-1.4216	0.286	8.50E-05	0.027299	-0.674	0.3136	0.045457	0.996133	TMEM5	Body	12	chr12:64173448-64174238	Island
cg26305042	2.158	0.3664	1.13E-05	0.00364	-0.4515	0.2114	0.046646	0.996133	WNT7B	Body	22	chr22:46326912-46327254	Island
cg04293526*	2.1306	0.4499	0.000144	0.046141	0.6957	0.3308	0.049815	0.996133	ITM2C;ITM	Body;Body	2	chr2:231729631-231730821	Island
cg07723251	-2.0924	0.4095	6.22E-05	0.019962	1.2251	0.5841	0.050357	0.996133	CA10;CA1(	TSS1500;T	17	chr17:50235175-50236466	S_Shore
cg04385523	-1.8465	0.3875	0.000135	0.043253	-1.0475	0.4999	0.050565	0.996133	C3orf23;C:	5'UTR;5'UT	3	chr3:44379532-44380310	S_Shore
cg26268565	-2.5699	0.4517	1.75E-05	0.005602	0.5693	0.2751	0.053189	0.996133	LHFPL3	Body	7		
cg15367212	-2.3685	0.432	2.74E-05	0.008781	-0.9944	0.4863	0.055789	0.996133			3		
cg07805777	0.9145	0.1541	1.04E-05	0.003327	0.4737	0.233	0.057081	0.996133			11	chr11:15962882-15963223	Island
cg06976485	-2.7173	0.5396	7.34E-05	0.023571	1.562	0.7792	0.060274	0.996133	ATF6B;ATF	Body;Body	6		
cg15323840	-1.5241	0.3184	0.000128	0.041122	-0.663	0.3335	0.062219	0.996133	SYNJ2	Body	6	chr6:158402072-158403444	S_Shore
cg06499213	-1.5944	0.3379	0.00015	0.048024	-0.7132	0.3638	0.065631	0.996133	TRAF3IP2;'	Body;Body	6		
cg26875958	-1.3584	0.276	9.46E-05	0.030368	-0.9417	0.4917	0.071513	0.996133	GRM1;GRI	3'UTR;Bod	6	chr6:146755475-146755901	N_Shelf
cg06998210	-1.3859	0.2941	0.000152	0.048759	-0.8177	0.4374	0.077927	0.996133	CA7;CA7	Body;TSS1	16	chr16:66878172-66879072	Island
ch_6_27918	-0.9032	0.1883	0.000126	0.040329	0.7404	0.3961	0.077959	0.996133					
cg24699146	-1.1967	0.2224	3.43E-05	0.010996	-0.6063	0.3255	0.078871	0.996133	HMGCL;HI	TSS1500;T	1		
cg03239925	-2.9066	0.6006	0.000114	0.036566	-0.5395	0.2905	0.079702	0.996133	NFATC1;NI	Body;Body	18	chr18:77230597-77230803	Island
cg08234372	-2.6627	0.4634	1.55E-05	0.004967	1.0774	0.5837	0.081428	0.996133			16		
cg09803177	-1.7389	0.2868	7.85E-06	0.00252	0.7065	0.3892	0.086183	0.996133	CPPED1;CI	TSS1500;T	16	chr16:12897365-12897889	S_Shore
cg23631062	-1.437	0.2579	2.26E-05	0.00724	0.4461	0.2468	0.087433	0.996133	OLFM3	TSS200	1		
cg11380128	-2.4902	0.5146	0.000114	0.036598	-1.0204	0.5684	0.089442	0.996133	PRLH	TSS1500	2		
cg03869928	-1.6397	0.3372	0.000108	0.034664	-1.1262	0.6403	0.095581	0.996133			2	chr2:8818292-8818503	N_Shore
cg01055579	-2.0959	0.3085	1.74E-06	0.000558	0.6608	0.3796	0.098781	0.996133	NLE1;NLE1	TSS200;TS	17	chr17:33468896-33469418	S_Shore
cg17746819	-1.1603	0.2237	5.24E-05	0.016818	0.4266	0.2458	0.099661	0.996133	DBT	TSS200	1		
cg13054881	-2.0628	0.4171	8.97E-05	0.028807	0.4601	0.2672	0.102171	0.996133	ADAM6	TSS1500	14		
cg25911551	-1.2727	0.2406	4.19E-05	0.013434	0.3079	0.1802	0.104618	0.996133	GAB1;GAB	Body;Body	4		
cg18194535	-2.6931	0.5629	0.000129	0.041384	1.0325	0.6071	0.106188	0.996133	LOC15482	TSS1500	7	chr7:158798591-158799434	S_Shore
cg05110962	-2.0719	0.4229	9.96E-05	0.031981	-0.7926	0.4933	0.12552	0.996133			15		
cg19787366	1.0729	0.2281	0.000155	0.049684	0.4598	0.2876	0.127305	0.996133	MPPE1	TSS1500	18	chr18:11908222-11909082	Island
cg23660197	-1.6824	0.3376	8.24E-05	0.02645	0.5687	0.357	0.128574	0.996133	MICB	3'UTR	6		
cg25637226	-2.2269	0.468	0.000137	0.043842	-1.5932	1.0008	0.128825	0.996133	CARS2	Body	13	chr13:111358080-111358894	N_Shelf
cg11970982	-2.2887	0.4768	0.000125	0.03998	0.8203	0.5209	0.132716	0.996133			2		
cg14788673	-2.2645	0.4731	0.000128	0.041176	0.8678	0.5581	0.137334	0.996133	FGGY;FGG	Body;Body	1		
cg15820400	-1.431	0.2634	3.05E-05	0.009787	0.5797	0.3755	0.139967	0.996133	FBXL17	Body	5		
cg02023402	-1.6257	0.2742	1.05E-05	0.00336	0.5494	0.3576	0.141925	0.996133			7		
cg13284152	-1.465	0.2962	8.98E-05	0.028841	0.7203	0.4694	0.142304	0.996133			6		
cg06218627	-1.2444	0.2592	0.000124	0.039812	0.6186	0.4062	0.145204	0.996133	LOC40237	TSS200;5'U	9	chr9:123555399-123555899	Island
cg00206271	-1.4157	0.2774	6.30E-05	0.020239	0.49	0.3259	0.149923	0.996133	SNTG2	Body	2	chr2:1163016-1163820	S_Shelf
cg23858558	-1.4833	0.2282	3.16E-06	0.001013	0.5074	0.3391	0.151883	0.996133	SDC2	TSS1500	8	chr8:97505747-97507607	N_Shore
cg14191955	-2.1271	0.4453	0.000131	0.042053	-0.5142	0.3451	0.153551	0.996133			17		
cg00883210	-2.1253	0.4121	5.61E-05	0.017999	0.8141	0.5479	0.154658	0.996133			4	chr4:3565311-3567185	Island
cg01880149	-2.8	0.5426	5.57E-05	0.017867	1.0764	0.7245	0.154663	0.996133	FLG	Body	1		
cg13951490	-1.7984	0.3559	7.06E-05	0.022649	0.7221	0.4954	0.162236	0.996133	PCDHGA4;	Body;Body	5	chr5:140798757-140799359	Island
cg25488567	-1.1795	0.2446	0.000119	0.038053	-0.6895	0.475	0.163767	0.996133	FAM26F	TSS1500	6	chr6:116783067-116783678	N_Shore
cg22673380	-0.8	0.1422	2.00E-05	0.006435	0.5949	0.4192	0.17289	0.996133	SLC12A7	Body	5	chr5:1054619-1054880	Island
	Taiwanese				IoW				CpG Information				

CpG	Taiwanese Estimate	Taiwanese StdErr	Taiwanese RAW_P	Taiwanese ahoc_p	IoW Estimate	IoW StdErr	IoW RAW_P	IoW ahoc_p	CpG Information UCSC_Re	CpG Information UCSC_Re	CpG Information CHR	CpG Information UCSC_CpG_Islands_Name	CpG Information Relation_to_UCSC_CpG_Island
cg00573504	-2.088	0.3525	1.06E-05	0.003396	-0.6346	0.4581	0.18287	0.996133			5		
cg10144474	-1.9852	0.412	0.00012	0.038364	-0.4391	0.319	0.185514	0.996133			15		
cg12143439	-1.304	0.2774	0.000156	0.049966	-0.3307	0.2409	0.186656	0.996133	CD14;CD1⁄	1stExon;5'	5	chr5:140011482-140012739	Island
cg13313214	-2.1982	0.4173	4.38E-05	0.014074	0.5719	0.422	0.192142	0.996133	STK31;STK	1stExon;5'	7		
cg02448934	0.4537	0.07789	1.31E-05	0.00419	0.3459	0.2675	0.21243	0.996133	BTNL2	Body	6		
cg00795205	-0.9408	0.1735	3.13E-05	0.010048	0.2277	0.1764	0.213052	0.996133	PRKCB;PRI	Body;Body	16		
cg07467338	-1.9048	0.3933	0.000113	0.036223	0.491	0.3819	0.214862	0.996133	FAM66D	Body	8		
cg18673825	-1.4222	0.3021	0.000153	0.049135	-0.633	0.4935	0.215879	0.996133	P2RX4	TSS1500	12	chr12:121647705-121648216	N_Shore
cg11273834	-2.0397	0.4173	0.000102	0.032775	0.7235	0.568	0.21897	0.996133	PPIL4	Body	6	chr6:149867031-149867574	N_Shelf
cg27500856	-1.9524	0.3833	6.46E-05	0.020736	0.4588	0.3659	0.225896	0.996133	C3orf71;AI	TSS1500;5	3	chr3:48955810-48956938	S_Shore
cg21219851	-1.5089	0.2878	4.64E-05	0.014884	0.4436	0.3573	0.230303	0.996133	RPTOR;RP	Body;Body	17	chr17:78896436-78896803	S_Shore
cg15352186	-2.1879	0.3995	2.77E-05	0.008896	0.8248	0.673	0.236139	0.996133	ADAM2	TSS200	8		
cg03460239	-1.5884	0.2718	1.25E-05	0.004013	0.538	0.4401	0.237332	0.996133	OR8B8	TSS200	11		
cg25601967	-1.5027	0.3188	0.000151	0.048486	0.4162	0.3462	0.24477	0.996133			12		
cg23372128	1.867	0.3926	0.000138	0.044164	0.6365	0.5319	0.246995	0.996133			12		
cg06535993	-1.167	0.2466	0.000145	0.046477	0.5613	0.4711	0.248974	0.996133	APBB1IP	TSS200	10	chr10:26727061-26727992	Island
cg20431441	-1.2909	0.2519	6.02E-05	0.019313	-0.2992	0.2538	0.253819	0.996133			8	chr8:82542798-82543475	N_Shelf
cg02584802	-1.6224	0.2741	1.07E-05	0.003431	0.3386	0.2874	0.254091	0.996133			11	chr11:115630398-115631117	S_Shelf
cg18773129	0.5703	0.1208	0.000149	0.04782	0.3749	0.3212	0.258254	0.996133	COMT;COI	5'UTR;5'U	22		
cg04267998	-1.4522	0.3059	0.00014	0.045017	-0.4506	0.3872	0.259696	0.996133	TRIP12	Body	2		
cg20026178	-1.0856	0.1838	1.09E-05	0.00351	0.422	0.3646	0.262209	0.996133	WRNIP1;W	Body;Body	6	chr6:2765203-2766775	S_Shore
cg03529662	-1.9868	0.3747	4.06E-05	0.013035	0.3722	0.3222	0.263242	0.996133	OR1A1	TSS1500	17		
cg19482843	-3.2159	0.673	0.000131	0.041947	0.7677	0.6654	0.2637	0.996133	EIF2C2;EIF	Body;Body	8	chr8:141559164-141559368	S_Shore
cg26335281	-1.2494	0.2512	8.42E-05	0.027039	-0.5922	0.5154	0.265631	0.996133	REEP2;REE	1stExon;5'	5		
cg06664085	1.2619	0.2683	0.000155	0.049629	-0.3362	0.2932	0.266637	0.996133	ARMC4;AR	5'UTR;1stE	10	chr10:28287447-28288057	Island
cg02500201	-2.3835	0.4496	4.07E-05	0.013069	-0.58	0.5078	0.268321	0.996133	CLCN7;CLC	Body;Body	16	chr16:1502628-1502858	Island
cg01541219	-2.8113	0.5436	5.43E-05	0.017421	-0.5643	0.5014	0.275204	0.996133	TAF4	Body	20	chr20:60630222-60630425	N_Shore
cg20420603	-3.3808	0.7143	0.000145	0.046446	-1.0915	0.9733	0.276823	0.996133	MYH10	Body	17	chr17:8380172-8380539	S_Shore
cg21981500	-1.7576	0.2721	3.44E-06	0.001103	-0.3844	0.3434	0.277659	0.996133	TMEM102	TSS200	17	chr17:7339808-7340896	N_Shore
cg19099213	-1.679	0.3438	0.000103	0.033155	-0.4512	0.4055	0.280556	0.996133	SPP2	TSS1500	2		
cg25196975	-2.0108	0.3405	1.10E-05	0.003525	0.6626	0.5968	0.281522	0.996133	ROBO3	Body	11	chr11:124738712-124739011	N_Shore
cg19647111	-1.1563	0.226	6.12E-05	0.019658	-0.4488	0.4159	0.294758	0.996133	TNXB	5'UTR	6		
cg25499537	-1.58	0.2783	1.79E-05	0.005755	-0.3305	0.3081	0.297554	0.996133	IGSF10	TSS1500	3	chr3:151178623-151178984	N_Shore
cg18759743	-1.1436	0.1737	2.65E-06	0.000852	0.3016	0.2837	0.301718	0.996133	ARHGEF10	Body;Body	1		
cg00931181	-1.8858	0.2966	4.23E-06	0.001358	-0.2317	0.2185	0.302814	0.996133			2		
cg02332769	-2.312	0.4069	1.77E-05	0.00569	-0.5073	0.4825	0.306988	0.996133	CCDC122	Body	13		
cg07156814	-2.3027	0.4453	5.43E-05	0.017442	-0.3082	0.2937	0.307776	0.996133	KIT;KIT	Body;Body	4		
cg19865002	-1.1992	0.2464	0.000107	0.034353	-0.3106	0.3049	0.321708	0.996133	RNF144B	5'UTR	6		
cg24117104	-1.8338	0.355	5.50E-05	0.017657	0.3042	0.3017	0.326662	0.996133			4	chr4:149363212-149367056	Island
cg25987171	2.1286	0.4342	9.88E-05	0.031718	0.3915	0.3893	0.32783	0.996133	SPDYC	TSS200	11		
cg02299937	2.0703	0.4363	0.000141	0.045234	0.4266	0.4314	0.335829	0.996133	FCGBP	Body	19	chr19:40366026-40366608	Island
cg03382370	0.9605	0.191	7.47E-05	0.023973	-0.1277	0.1299	0.338723	0.996133	LOC14633	Body	16	chr16:1122629-1123009	N_Shore
cg02346442	-1.9306	0.3737	5.49E-05	0.017625	-0.4372	0.4466	0.340563	0.996133			3		
cg25288034	1.6488	0.3473	0.00014	0.045005	0.286	0.2931	0.342179	0.996133			10		
cg18888710	-2.2181	0.4378	6.86E-05	0.022009	0.4731	0.494	0.350911	0.996133	ATP6V1C2	Body;Body	2		
cg18688704	-1.8322	0.3365	2.97E-05	0.009528	0.3702	0.3899	0.355058	0.996133	PDGFC	Body	4	chr4:157892685-157893286	N_Shelf
cg13346967	-2.1154	0.4328	0.000102	0.032779	-0.7477	0.7915	0.357351	0.996133	CYTH2;CYT	Body;Body	19	chr19:48972225-48973157	S_Shelf
cg03400512	-2.2383	0.4503	8.49E-05	0.027244	0.6007	0.6463	0.364994	0.996133			1		
cg04550474	-1.8785	0.3757	7.95E-05	0.025516	0.3016	0.3279	0.369717	0.996133	LOC64200	Body	7	chr7:57247588-57247927	N_Shore
cg07500347	-1.46	0.2851	6.08E-05	0.019519	-0.3187	0.3466	0.369853	0.996133			16	chr16:874280-874895	S_Shore
cg14406401	-1.6762	0.3457	0.000112	0.035792	0.3469	0.3782	0.371129	0.996133	NAV2;NAV	Body;Body	11		
cg11735008	-1.6423	0.3255	7.18E-05	0.023038	-0.4282	0.4795	0.383641	0.996133	HIVEP3;HI	5'UTR;5'U	1	chr1:42383640-42384003	N_Shelf
cg17330505	-0.5972	0.1116	3.64E-05	0.01169	-0.5149	0.5821	0.388068	0.996133			17		
cg09838336	-2.6072	0.5234	8.29E-05	0.026612	0.543	0.6379	0.405797	0.996133	C7orf44	5'UTR	7		
cg02722657	1.2973	0.2604	8.27E-05	0.026559	-0.2178	0.2586	0.410707	0.996133	RAPGEFL1	1stExon;5'	17	chr17:38333605-38334795	Island
cg24399212	-3.0161	0.5986	7.30E-05	0.023419	0.7176	0.8552	0.412413	0.996133			10		
cg16998490	-1.6837	0.294	1.61E-05	0.005173	-0.3094	0.3724	0.416999	0.996133	AZIN1;AZI	TSS1500;T	8	chr8:103875222-103877084	Island
cg16383005	-2.1371	0.3967	3.37E-05	0.01083	-0.5137	0.6268	0.423161	0.996133			6	chr6:166711162-166711605	Island
cg10201735	-1.9505	0.3408	1.62E-05	0.005212	-0.3683	0.451	0.424764	0.996133			14	chr14:97499605-97500176	S_Shelf
cg19005233	-2.2846	0.4409	5.31E-05	0.017033	0.373	0.4603	0.428389	0.996133	PCDHB15	1stExon	5	chr5:140626444-140627373	Island
cg20458779	2.5572	0.4202	7.50E-06	0.002407	0.454	0.5604	0.428431	0.996133	FCGBP	Body	19	chr19:40366026-40366608	Island
cg03033688	-1.6543	0.333	8.55E-05	0.027443	0.2535	0.315	0.431359	0.996133	HS3ST2	TSS1500	16	chr16:22824616-22826459	N_Shore
cg21500456	-2.4948	0.5211	0.000128	0.041059	-0.3519	0.4411	0.435452	0.996133			14	chr14:21191657-21191860	N_Shore
cg18638330	-1.6928	0.3288	5.71E-05	0.018343	-0.2984	0.3826	0.445599	0.996133			5		
cg27209964	-3.5507	0.6569	3.24E-05	0.010412	0.4297	0.5558	0.449536	0.996133			13		
cg08209711	-0.8999	0.191	0.000152	0.048786	0.2281	0.297	0.452459	0.996133	PSMB9;PSI	Body;Body	6	chr6:32820849-32822370	S_Shelf
cg13336167	-1.5366	0.3132	9.83E-05	0.031541	-0.2898	0.3783	0.453456	0.996133	ZNF579	5'UTR	19	chr19:56089679-56090859	S_Shore
cg22197358	-2.8514	0.4512	4.59E-06	0.001472	0.367	0.4794	0.45388	0.996133	TCF3;TCF3	Body;Body	19	chr19:1615202-1615812	Island
cg19080320	-1.284	0.2666	0.00012	0.038594	-0.343	0.4498	0.455617	0.996133	FOXK1	Body	7		
cg20299254	-1.6378	0.3064	3.70E-05	0.011878	0.4104	0.5476	0.463261	0.996133	WFDC2	TSS1500	20	chr20:44098280-44099536	N_Shore
cg00209918	1.47	0.2915	7.23E-05	0.023217	-0.2116	0.2844	0.466532	0.996133	SETD1A	Body	16	chr16:30991828-30992129	N_Shore
cg26957602	-1.8876	0.3774	7.93E-05	0.025462	0.3409	0.4647	0.472608	0.996133			15		
cg26143719	-2.766	0.5756	0.000123	0.039436	0.4767	0.6604	0.479703	0.996133	C1QTNF6;(	1stExon;1s	22		
cg07780669	-0.9156	0.1825	7.67E-05	0.024606	0.2355	0.3324	0.487727	0.996133	TMCO6	TSS200	5	chr5:140019034-140019502	N_Shore
cg14323117	-1.2791	0.2689	0.000137	0.044016	0.1729	0.2449	0.489242	0.996133			3	chr3:14643718-14644830	N_Shore
cg11029358	-2.3969	0.4811	8.27E-05	0.026539	-0.1729	0.2475	0.493779	0.996133	COL18A1;(	5'UTR;1stE	21	chr21:46876168-46876752	N_Shore
cg27486692	-1.8179	0.2848	4.02E-06	0.00129	-0.3692	0.5421	0.504471	0.996133	GUCY2C	3'UTR	12		
cg20075700	-2.7422	0.4981	2.60E-05	0.008361	-0.2936	0.4375	0.510725	0.996133			22	chr22:49376395-49376660	Island
cg27134084	-2.5356	0.5351	0.000143	0.0459	0.2503	0.3788	0.517148	0.996133			7		
cg08482979	1.4031	0.2938	0.000132	0.042222	0.1989	0.3041	0.521402	0.996133			16		
cg06814191	-2.7113	0.5625	0.000119	0.038174	-0.3791	0.581	0.522302	0.996133			11		
cg06829645	-2.5013	0.5237	0.000131	0.042173	-0.4076	0.6541	0.541005	0.996133			13		
cg10439742	-0.9147	0.1836	8.28E-05	0.026594	0.1416	0.2428	0.567091	0.996133	HSBP1	3'UTR	16	chr16:83841412-83841988	S_Shelf
cg08581737	-1.3628	0.2601	4.68E-05	0.015014	-0.1689	0.3131	0.596259	0.996133	DLX2	1stExon	2	chr2:172964762-172967857	Island
cg18356276	-2.0972	0.4301	0.000105	0.033677	0.2351	0.4381	0.598092	0.996133	AHR	TSS1500	7	chr7:17338252-17339101	N_Shore
cg12765716	-2.2927	0.4416	5.18E-05	0.016643	-0.2886	0.5691	0.618268	0.996133	TRPV4;TRI	5'UTR;5'U	12		

CpG	Taiwanese				IoW				CpG Information				
	Estimate	StdErr	RAW_P	ahoc_p	Estimate	StdErr	RAW_P	ahoc_p	UCSC_Re	UCSC_Re	CHR	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island
cg10803309	-1.9821	0.371	3.72E-05	0.01195	0.2141	0.4241	0.619787	0.996133	HIPK3;HIPI	Body;Body	11		
cg15272312	-0.6987	0.1144	7.17E-06	0.002303	-0.1653	0.3315	0.624049	0.996133	HNRNPA1I	1stExon;1s	13	chr13:53191405-53191872	Island
cg14410733	-2.8774	0.5989	0.000123	0.03954	0.2261	0.454	0.624575	0.996133			15		
cg08727867	-1.0338	0.2183	0.000144	0.046223	0.1883	0.3887	0.633881	0.996133	RBM11	TSS200	21	chr21:15588439-15588908	N_Shore
cg17769009	-2.5466	0.5394	0.000149	0.047728	-0.2967	0.6205	0.638283	0.996133	RPL23AP5	Body	8	chr8:179001-179288	Island
cg01864127	1.1965	0.2384	7.62E-05	0.02445	-0.153	0.3304	0.648875	0.996133	CLPTM1L	Body	5		
cg11781622	-1.7654	0.3506	7.36E-05	0.023612	0.1661	0.3741	0.662408	0.996133	PRKCZ;PRK	5'UTR;Bod	1	chr1:2040997-2041302	Island
cg24290586	0.473	0.08281	1.66E-05	0.005332	0.09401	0.2184	0.671942	0.996133	BTNL2	Body	6		
cg16019434	-1.0071	0.2099	0.000125	0.040115	-0.1264	0.2962	0.674506	0.996133			18	chr18:76737005-76741244	N_Shore
cg18469813	-1.6067	0.3408	0.000151	0.048515	-0.1656	0.3883	0.674741	0.996133			6		
cg12349837	-1.8511	0.3851	0.000123	0.039357	-0.03183	0.0763	0.681457	0.996133	COL4A3;C(	Body;Body	2	chr2:228028679-228029733	S_Shelf
cg11619050	-1.3712	0.2869	0.00013	0.041856	-0.1829	0.4465	0.686911	0.996133	APOM	TSS1500	6	chr6:31619856-31620525	S_Shelf
cg26447413	1.8636	0.3632	5.93E-05	0.019045	-0.366	0.8966	0.68796	0.996133	GAS1	1stExon	9	chr9:89560584-89562647	Island
cg15372959	-2.0238	0.3841	4.38E-05	0.014047	-0.2055	0.5236	0.699349	0.996133	TACC2;TA(	TSS1500;B	10	chr10:123922850-123923542	N_Shore
cg09821400	-1.5844	0.3078	5.72E-05	0.018374	-0.1785	0.4869	0.718194	0.996133			10		
cg13979407	0.7638	0.1577	0.000113	0.036308	-0.1171	0.3364	0.731841	0.996133	BTNL2	Body	6		
cg03364683	-1.2194	0.2095	1.32E-05	0.004231	0.07429	0.2152	0.733863	0.996133	DDX60L	TSS1500	4	chr4:169401350-169401725	S_Shore
cg03610567	-2.1992	0.4401	8.00E-05	0.025676	0.2441	0.7291	0.741584	0.996133	TRIM26	Body	6		
cg25687358	-1.3988	0.2738	6.23E-05	0.019998	-0.09616	0.2929	0.746478	0.996133	DUOXA1;D	Body;Body	15	chr15:45408573-45409528	S_Shore
cg08107075	-2.4409	0.4792	6.45E-05	0.020691	-0.1655	0.5044	0.746542	0.996133			7		
cg13443976	-2.2106	0.4288	5.63E-05	0.018058	0.202	0.618	0.747589	0.996133	CLCA4;CLC	Body;Body	1	chr1:87018668-87019254	S_Shelf
cg17387838	-1.395	0.2686	5.16E-05	0.01657	0.1095	0.3401	0.751165	0.996133	CACNG6;C	3'UTR;3'U	19	chr19:54515094-54515342	S_Shore
cg03140624	-1.4736	0.3072	0.000125	0.040212	-0.1452	0.4772	0.764324	0.996133	LYST	Body	1		
cg10002668	-1.8301	0.3754	0.000105	0.033803	-0.1243	0.4157	0.768404	0.996133	LAMB4	Body	7		
cg08486903	-1.1752	0.2463	0.000133	0.042591	-0.09742	0.3289	0.770452	0.996133	SMURF1;S	Body;Body	7	chr7:98741205-98741990	N_Shore
cg04990210	-2.198	0.3503	5.04E-06	0.001617	-0.1718	0.5832	0.771638	0.996133	CTNS;CTN!	Body;Body	17		
cg04723493	-1.442	0.2852	7.01E-05	0.022493	-0.1805	0.6232	0.775358	0.996133	FZD6;FZD6	TSS1500;T	8	chr8:104310809-104311620	Island
cg11413778	-3.1727	0.6716	0.000148	0.047412	-0.1503	0.5337	0.781406	0.996133	ALLC	5'UTR	2		
ch_3_27419	-0.8377	0.171	9.96E-05	0.031974	-0.07461	0.2724	0.78728	0.996133					
cg26664634	-0.6542	0.1299	7.31E-05	0.023463	0.06808	0.252	0.790129	0.996133	ORC4L;OR(	TSS200;5'U	2	chr2:148778222-148778530	Island
cg16579158	-1.6705	0.3371	8.78E-05	0.028179	0.07568	0.2815	0.791112	0.996133	PCDHGA2;	1stExon;Bo	5	chr5:140719129-140719343	Island
cg10548968	-1.7066	0.3493	0.000103	0.032982	-0.1001	0.4216	0.814923	0.996133	CAPN5	Body	11		
cg16396866	-1.6647	0.3249	6.03E-05	0.01937	-0.09242	0.3948	0.817571	0.996133	MCTP1	Body	5	chr5:94619459-94621121	N_Shelf
cg14403594	-1.0403	0.2124	1.00E-04	0.032093	0.06389	0.3008	0.834174	0.996133	POMT1;PC	TSS1500;T	9	chr9:134378193-134378792	N_Shore
cg23075771	-1.9534	0.3998	0.000103	0.032977	-0.1006	0.5271	0.850844	0.996133			1		
cg01974027	-1.9063	0.2982	3.94E-06	0.001265	-0.07521	0.3946	0.850958	0.996133	SYT6	5'UTR	1	chr1:114695136-114696672	N_Shore
cg13404532	-2.5384	0.5083	8.07E-05	0.025894	0.06492	0.3432	0.852088	0.996133	RAB7A	5'UTR	3	chr3:128483727-128484050	N_Shore
cg10083824	-1.2876	0.2398	3.51E-05	0.011268	-0.05349	0.2922	0.85679	0.996133	GRM4	TSS1500	6		
cg17642145	-1.1978	0.2269	4.28E-05	0.013747	0.04347	0.2545	0.866298	0.996133	PPP3CC	TSS200	8	chr8:22298112-22299142	Island
cg24600706	-1.3994	0.2916	0.000125	0.040048	-0.07174	0.448	0.874552	0.996133			10	chr10:118899247-118900329	S_Shore
cg15256743	-1.1554	0.231	7.91E-05	0.0254	0.05468	0.3444	0.875637	0.996133			12	chr12:94533729-94533952	S_Shelf
cg02030454	-1.3222	0.2286	1.43E-05	0.004585	-0.04857	0.3229	0.882124	0.996133	GJB3;GJB3	Body;Body	1	chr1:35250418-35250736	Island
cg01886741	-1.7741	0.3139	1.90E-05	0.006092	-0.05547	0.3794	0.88539	0.996133	BIRC5;BIR(	1stExon;1s	17	chr17:76210129-76210627	Island
cg09176901	-2.1276	0.3947	3.35E-05	0.01075	0.07928	0.5621	0.8894	0.996133	TIMELESS	Body	12		
cg19224656	-0.6659	0.1261	4.26E-05	0.013662	-0.03712	0.2797	0.895917	0.996133	SHQ1;SHQ	5'UTR;1stE	3	chr3:72896940-72897793	Island
cg07278997	-1.1137	0.2099	4.03E-05	0.012943	0.03138	0.2786	0.911574	0.996133	UBE2W;UI	TSS1500;T	8	chr8:74790342-74791412	Island
cg06382167	-2.2587	0.4079	2.43E-05	0.007794	-0.06264	0.5646	0.912893	0.996133	PDE11A;PI	5'UTR;1stE	2		
cg15549502	-3.0733	0.5927	5.27E-05	0.016904	-0.08166	0.7973	0.919556	0.996133	MORN1	Body	1	chr1:2266007-2266432	S_Shelf
cg11823603	-1.5277	0.3116	9.90E-05	0.031774	-0.04049	0.3979	0.920063	0.996133			5		
cg02751327	-1.4673	0.2689	2.90E-05	0.009293	-0.03795	0.3983	0.925146	0.996133			12	chr12:64784010-64784664	S_Shelf
cg04446303	-1.4577	0.3092	0.000151	0.048434	-0.03164	0.342	0.927314	0.996133	TFAP4;TFA	1stExon;5'	16	chr16:4321639-4324073	Island
cg19304063	-1.5689	0.3086	6.59E-05	0.021162	0.04057	0.4468	0.928656	0.996133	FARP1	3'UTR	13		
cg24677222	-1.6605	0.3091	3.48E-05	0.011182	-0.03552	0.4187	0.933338	0.996133	CLSTN2	Body	3		
cg21945260	-2.0333	0.4288	0.000142	0.045525	0.02326	0.2851	0.935882	0.996133			2		
cg08300990	-1.6212	0.287	1.91E-05	0.006121	0.05063	0.6537	0.939117	0.996133	OR8G1;OR	Body;1stE)	11		
cg02704485	-0.6204	0.1074	1.45E-05	0.004654	-0.01693	0.2341	0.943158	0.996133			3		
cg01769501	-0.8809	0.1524	1.43E-05	0.004603	-0.01262	0.2106	0.952861	0.996133	SDAD1	TSS200	4	chr4:76911843-76912252	Island
cg05590053	-1.899	0.2748	1.37E-06	0.000439	-0.01985	0.3318	0.952964	0.996133			8		
cg00614641	1.8633	0.3947	0.000149	0.047697	-0.02875	0.5134	0.955958	0.996133	FADS2	Body	11	chr11:61594996-61596710	Island
cg00648955	-3.3206	0.54	6.55E-06	0.002103	0.02167	0.3958	0.956942	0.996133	EHD3	TSS1500	2	chr2:31456740-31457653	N_Shore
cg24698488	-1.9674	0.4035	0.000105	0.033666	0.02895	0.5819	0.960865	0.996133	MMP14	3'UTR	14	chr14:23315027-23315265	S_Shore
cg08219099	-1.2593	0.2627	0.000126	0.040563	-0.02513	0.6033	0.967235	0.996133	MAEA;MA	Body;Body	4	chr4:1322394-1322875	S_Shelf
cg05548672	-2.3812	0.4638	5.89E-05	0.018912	-0.02032	0.5684	0.971869	0.996133	C16orf72	Body	16		
cg20761853	-1.6782	0.3518	0.000133	0.04267	-0.00832	0.2859	0.977102	0.996133	TIMP2	3'UTR	17	chr17:76851759-76851973	N_Shore
cg08575537	2.0672	0.4369	0.000145	0.046599	0.008531	0.3454	0.98057	0.996133	EPO	Body	7	chr7:100318106-100318684	Island
cg02860394	-1.1966	0.2259	4.12E-05	0.013233	0.002926	0.339	0.993208	0.996133	RNF103	TSS200	2	chr2:86850042-86851178	Island
cg13165753	-1.2379	0.2609	0.000141	0.045257	0.00434	0.5944	0.994255	0.996133	FN1;FN1;F	TSS1500;T	2	chr2:216300267-216300997	S_Shore
cg21179305	-1.9475	0.3974	9.93E-05	0.03189	0.002758	0.4009	0.994587	0.996133	ZCCHC7	TSS1500	9	chr9:37120050-37120985	N_Shore
cg07313835	-1.8993	0.397	0.000129	0.041404	-0.00253	0.5151	0.996133	0.996133	POU2F1	Body	1		
cg00090648	-0.778	0.1565	8.50E-05	0.02729					C1orf55	TSS1500	1	chr1:226186803-226187336	Island
cg00095105	-1.9562	0.4155	0.000153	0.049175							10		
cg00173799	-2.0917	0.4389	0.000134	0.043106					PGR	TSS1500	11	chr11:100998030-100999774	S_Shore
cg00460793	-1.9521	0.3464	1.96E-05	0.006297							7		
cg00576474	-2.6198	0.4691	2.19E-05	0.00704							2	chr2:242929993-242930252	N_Shore
cg00624949	-1.0591	0.1896	2.18E-05	0.006992					EGFL8	Body	6	chr6:32134460-32135196	Island
cg00627621	-0.9188	0.1717	3.64E-05	0.011689					TRIB1	TSS200	8	chr8:126441471-126443552	Island
cg00647646	1.5304	0.2815	3.03E-05	0.009739					CPSF6	Body	12	chr12:69633272-69634095	Island
cg00712044	-1.0648	0.2003	3.95E-05	0.012689					KIAA1826;	1stExon;5'	11	chr11:105892725-105892952	Island
cg00804628	0.8863	0.1723	5.75E-05	0.018461					NXPH2	TSS200	2	chr2:139537692-139538650	Island
cg00810986	-1.4181	0.2853	8.48E-05	0.027231					PCGF3	Body	4	chr4:724335-726070	S_Shore
cg00853687	-1.7765	0.3102	1.61E-05	0.005163							1		
cg00866373	-0.9748	0.207	0.000152	0.048934					ZCCHC4;ZC	5'UTR;1stE	4	chr4:25314217-25314558	Island
cg00969787	-1.9334	0.2922	2.48E-06	0.000797					TESK2	5'UTR	1	chr1:45956345-45956893	N_Shore
cg01026661	-1.3889	0.2204	4.75E-06	0.001525					FAM173B	Body	5		
cg01132471	-1.9216	0.4078	0.000152	0.048697					SLC1A7	Body	1	chr1:53558388-53558642	N_Shore
cg01396112	1.6082	0.2927	2.67E-05	0.008565					NOS2	Body	17	chr17:26120369-26120944	Island

CpG	Estimate	StdErr	RAW_P	ahoc_p	Estimate	StdErr	RAW_P	ahoc_p	UCSC_Re	UCSC_Re	CHR	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island
cg01464748	2.1287	0.4464	0.000133	0.042837					ALOX5	Body	10	chr10:45922831-45923375	Island
cg01685923	-0.3916	0.07561	5.33E-05	0.017119							13		
cg01831337	0.6453	0.1279	7.17E-05	0.023					UST	Body	6	chr6:149198359-149199534	N_Shore
cg01871995	-1.3889	0.2824	9.54E-05	0.030612					FGL1;FGL1	TSS1500;T	8		
cg02112681	-1.0283	0.1761	1.27E-05	0.004069					C7orf50;C	Body;Body	7	chr7:1119980-1120248	Island
cg02159489	-0.2166	0.04332	7.97E-05	0.02557							17	chr17:79454734-79455823	S_Shelf
cg02820514	1.4777	0.3081	0.000126	0.040308					ZIC4	TSS1500	3	chr3:147126988-147128999	N_Shelf
cg02923571	0.8839	0.1695	4.94E-05	0.015867					PWP1	TSS200	12	chr12:108079442-108079893	Island
cg02982238	-2.9894	0.6307	0.000143	0.045758							7		
cg02987249	-2.5048	0.4527	2.45E-05	0.007868					TRIP10	Body	19	chr19:6752515-6753657	N_Shelf
cg03050687	0.9801	0.2035	0.00012	0.038543					ADAL;LCM	TSS1500;1	15	chr15:43621830-43622953	Island
cg03128025	-1.2206	0.2595	0.000155	0.049614							2		
cg03180302	-1.7696	0.3321	3.83E-05	0.012305					TTR	1stExon	18		
cg03202545	-0.5976	0.1127	4.06E-05	0.013039					HIPK2;HIPI	Body;Body	7		
cg03215005	-0.9427	0.192	9.74E-05	0.031281					MX1	TSS1500	21	chr21:42792293-42792704	Island
cg03237431	-1.121	0.2379	0.000152	0.048736					CACNG1	TSS1500	17	chr17:65040729-65041029	N_Shore
cg03242458	-3.4038	0.6794	7.78E-05	0.024963					TRIM15	Body	6		
cg03263979	-1.9703	0.4165	0.000146	0.046743							19	chr19:33864176-33865010	N_Shelf
cg03277925	0.8427	0.1607	4.61E-05	0.014807					ADARB2	Body	10		
cg03437605	-2.0149	0.4094	9.47E-05	0.030404					RHOBTB2	5'UTR	8		
cg03442350	-0.634	0.1041	7.46E-06	0.002395							6		
cg03651219	-2.302	0.4559	7.11E-05	0.022837							10		
cg03689195	2.3232	0.4817	0.000118	0.03793							13	chr13:58203586-58204322	Island
cg04134528	-2.6097	0.5529	0.000149	0.047887							5		
cg04193922	-1.0958	0.2176	7.35E-05	0.023596					LOC72933	TSS200;TS	4	chr4:123653585-123654000	Island
cg04245294	-2.145	0.4519	0.00014	0.045073							3	chr3:37901938-37904039	N_Shelf
cg04248157	1.3255	0.2686	9.19E-05	0.029502					KCNIP1;KC	TSS200;TS	5	chr5:169931065-169931471	N_Shore
cg04315226	-2.0414	0.3729	2.79E-05	0.008943					FAM164C;	3'UTR;3'UT	14		
cg04398972	0.7449	0.1485	7.67E-05	0.024635							4	chr4:75858300-75859931	N_Shelf
cg04407100	-2.9657	0.4143	8.38E-07	0.000269							10		
cg04435807	-1.8898	0.3871	0.000103	0.033196					ANKRD22	Body	10		
cg04461334	-1.7699	0.3376	4.64E-05	0.014889					MAP4K3	Body	2	chr2:39663897-39664667	N_Shelf
cg04903900	-1.9568	0.3015	3.22E-06	0.001033							16		
cg04913730	0.8166	0.1696	0.00012	0.038558					LOC14633	Body	16	chr16:1122629-1123009	N_Shore
cg04935278	-2.4073	0.5052	0.000135	0.043191					ITPRIP	TSS200	10	chr10:106097595-106098556	N_Shelf
cg04984709	-1.8085	0.3381	3.67E-05	0.011785					PDE8B;PDI	Body;Body	5		
cg04985652	-0.7892	0.1494	4.26E-05	0.013661					DTD1	TSS200	20	chr20:18568436-18569092	Island
cg05031521	1.5726	0.2913	3.29E-05	0.010552					SOX8	TSS1500	16	chr16:1029878-1035327	Island
cg05044694	-1.3366	0.217	6.42E-06	0.00206					PPP1CB;PF	Body;Body	2	chr2:28974212-28975388	Island
cg05092988	2.5226	0.4648	3.09E-05	0.00991					DSCAML1	Body	11		
cg05099145	1.0797	0.2145	7.39E-05	0.023722					MAPK8IP3	Body;Body	16	chr16:1796810-1797229	N_Shore
cg05426702	-0.7936	0.1589	8.03E-05	0.025777							15	chr15:40573628-40576118	Island
cg05590265	-2.0911	0.4445	0.000154	0.049564							6		
cg05918473	-2.2093	0.4093	3.29E-05	0.010572					ZAR1L	1stExon	13	chr13:32885424-32885662	Island
cg05939041	-2.9078	0.5946	0.000102	0.032637					ITPK1;ITPK	Body;Body	14		
cg06031526	-0.8396	0.1755	0.000129	0.041407							3		
cg06239037	-0.9344	0.1953	0.000129	0.041381					ACAA1;AC	TSS1500;T	3	chr3:38179857-38180689	Island
cg06363887	-0.6372	0.1158	2.61E-05	0.008379					UTP3	TSS200	4	chr4:71553854-71554473	Island
cg06666008	-1.5021	0.3169	0.000143	0.045788					PAX6;PAX6	Body;Body	11	chr11:31820060-31821416	N_Shore
cg06679777	-1.1748	0.2095	2.09E-05	0.006703					KATNAL1;I	TSS200;5'U	13	chr13:30880899-30881939	Island
cg06741896	-1.0739	0.2078	5.48E-05	0.017576					CCND1	3'UTR	11	chr11:69468810-69469152	N_Shore
cg07169712	0.9916	0.2095	0.000145	0.046427					GABBR1;G	Body;Body	6		
cg07175007	-0.8073	0.1548	4.93E-05	0.01583					UHMK1	TSS1500	1	chr1:162467599-162468027	N_Shore
cg07210335	-1.1128	0.2304	0.000116	0.037322					MAN2A1	Body	5	chr5:109024965-109026701	S_Shore
cg07297397	-1.1571	0.2329	8.55E-05	0.027459					IGF2BP3	TSS200	7	chr7:23508184-23509712	S_Shore
cg07317755	-0.804	0.1595	7.24E-05	0.023254					PDE12	TSS200	3	chr3:57541931-57543244	N_Shore
cg07360893	-0.9258	0.1866	8.66E-05	0.027812					IFT80;SMC	TSS200;TS	3	chr3:160117012-160118878	Island
cg07917502	-1.1209	0.2128	4.38E-05	0.01407					PREPL;PRE	TSS200;TS	2	chr2:44588954-44589358	Island
cg08207170	-1.8942	0.3916	0.000114	0.036723							16	chr16:89882392-89883458	S_Shelf
cg08479675	-2.1063	0.4262	9.06E-05	0.029093					CABP5	Body	19		
cg08643994	0.7543	0.1595	0.000146	0.046962							4		
cg08706545	-1.8729	0.3764	8.40E-05	0.026953					CHD5	TSS1500	1	chr1:6239237-6240814	Island
cg09008360	-1.8073	0.3736	0.000114	0.036714							8	chr8:1956926-1957177	N_Shore
cg09026722	-1.7494	0.3605	0.000111	0.035475					PEAR1	5'UTR	1	chr1:156863415-156863711	S_Shelf
cg09096528	-1.0102	0.2086	0.000113	0.03634					CDK4	TSS200	12	chr12:58145853-58146360	Island
cg09433665	-0.9908	0.1889	4.62E-05	0.014846					YPEL2	TSS200	17	chr17:57408787-57410145	Island
cg09456184	-1.2565	0.2308	2.98E-05	0.009559					HNRNPM;I	TSS1500;T	19	chr19:8509635-8510765	N_Shore
cg09830552	-0.8469	0.1796	0.000151	0.048351					CP110	TSS200	16	chr16:19535074-19535635	Island
cg10016575	1.632	0.2984	2.82E-05	0.00906					FCGBP	Body	19	chr19:40397965-40398505	Island
cg10087081	-1.39	0.287	0.000113	0.036319					KIAA0355	5'UTR	19		
cg10315347	-1.0825	0.2093	5.42E-05	0.017387					CGGBP1;C	TSS200;TS	3	chr3:88107550-88108509	Island
cg10409785	-1.39	0.2885	0.00012	0.038435					RG9MTD2	5'UTR;TSS	4	chr4:100484727-100485284	Island
cg10642957	-0.7595	0.1521	8.07E-05	0.025908					ZNF524	Body	19	chr19:56113827-56114232	Island
cg10707081	-1.2303	0.2604	0.000148	0.04739					PCDH9;PCI	Body;Body	13		
cg10858369	0.8162	0.1584	5.67E-05	0.018209					TRIM3;TRI	Body;Body	11	chr11:6477603-6478178	N_Shore
cg10934821	-0.9864	0.2095	0.000153	0.049201					TAOK3	5'UTR	12	chr12:118809324-118811021	Island
cg11015424	-1.8841	0.3839	9.78E-05	0.031391					RALGAPA1	Body;Body	14		
cg11759749	-3.2912	0.5251	5.11E-06	0.001641							6		
cg11982583	-1.0499	0.2038	5.66E-05	0.018175					GRID2	Body	4	chr4:93226348-93227007	N_Shore
cg12104477	-0.9588	0.1814	4.22E-05	0.013535					ARID3B	TSS1500	15	chr15:74833241-74833961	Island
cg12423278	-1.1831	0.2269	4.93E-05	0.015838							5		
cg12663426	-1.4778	0.2821	4.67E-05	0.015							1	chr1:155112976-155113179	S_Shelf
cg12687463	-1.3946	0.2946	0.000144	0.046315					CFHR1	TSS1500	1		
cg12689752	1.5006	0.2979	7.32E-05	0.023486					IQGAP3	3'UTR	1		
cg12855271	-2.4543	0.3973	6.18E-06	0.001982							8		
cg12985923	-2.5352	0.4862	4.94E-05	0.015855					KCNG2	TSS1500	18	chr18:77623265-77624280	N_Shore
		Taiwanese					**IoW**					**CpG Information**	

CpG	Taiwanese Estimate	Taiwanese StdErr	Taiwanese RAW_P	Taiwanese ahoc_p	IoW Estimate	IoW StdErr	IoW RAW_P	IoW ahoc_p	UCSC_Re	UCSC_Re	CHR	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island
cg13099261	-2.2995	0.4683	9.72E-05	0.031203							12		
cg13229782	-1.2782	0.2449	4.88E-05	0.015653							17	chr17:21178819-21179690	S_Shore
cg13332474	1.094	0.1895	1.46E-05	0.00469							7		
cg13487156	-1.9427	0.4111	0.000147	0.047272					RFTN1	Body	3		
cg13578229	-0.7263	0.1316	2.53E-05	0.008116					ZNF510	TSS200	9	chr9:99540025-99540472	Island
cg13612524	-1.1102	0.2239	8.74E-05	0.02806					ERLIN2;ER	TSS1500;T	8	chr8:37593834-37594703	Island
cg13656556	-2.0559	0.4136	8.49E-05	0.027268					GFPT2	Body	5		
cg13689204	-1.6799	0.3358	7.91E-05	0.025376							17		
cg13710556	-1.503	0.244	6.40E-06	0.002054							7	chr7:25891956-25892615	S_Shore
cg13884344	-2.0018	0.3386	1.08E-05	0.003475					RNF38;RN	3'UTR;3'U	9		
cg13918937	-1.1164	0.2328	0.000126	0.040306					C19orf10	1stExon	19	chr19:4669927-4670528	Island
cg14151682	-0.9587	0.1887	6.65E-05	0.021333					PINK1	TSS200	1	chr1:20960045-20960551	N_Shore
cg14354820	-2.0339	0.4079	8.19E-05	0.0263					ANK1	Body	8		
cg14398659	-1.1672	0.243	0.000124	0.039658					INTS4	TSS200	11	chr11:77705600-77706035	Island
cg14611830	-1.1771	0.2458	0.000128	0.04095					TTLL4	TSS200	2	chr2:219575448-219576080	Island
cg14631690	-2.4709	0.4646	3.92E-05	0.012594					CLPTM1L	Body	5	chr5:1335077-1335280	Island
cg14647515	-1.3148	0.2759	0.000134	0.043173					C10orf12	1stExon	10		
cg14753070	-2.2551	0.4392	5.89E-05	0.018918					IL27RA	3'UTR	19		
cg14894216	-1.619	0.3406	0.000138	0.044354					CATSPER1	TSS1500	11		
cg15147833	-1.8938	0.3627	4.86E-05	0.015612					OR1C1	1stExon	1		
cg15313617	-1.0822	0.2269	0.000134	0.042874					MAK16	Body	8	chr8:33342171-33343141	S_Shore
cg15400591	-0.8711	0.1832	0.000138	0.044175					PROCR	Body	20	chr20:33762403-33762774	S_Shore
cg15582138	-0.8252	0.1511	2.86E-05	0.009185							12		
cg15633603	-1.3646	0.2833	0.00012	0.038478							2		
cg15665081	-0.5232	0.09288	1.97E-05	0.006334					CASP4;CAS	5'UTR;Bod	11		
cg15712821	-0.9285	0.1827	6.62E-05	0.021252					PTK2B;PTK	5'UTR;5'U	8	chr8:27168205-27169082	S_Shelf
cg15897970	0.9015	0.1889	0.000133	0.042548					FLJ43390	TSS200	14	chr14:62583679-62584279	Island
cg16033376	-0.7046	0.1485	0.00014	0.0451					PRR24	TSS1500	19	chr19:47776370-47778740	Island
cg16078269	-1.6532	0.3306	7.94E-05	0.025497					ZBTB7C	TSS1500	18		
cg16248798	-1.4764	0.2923	7.11E-05	0.022821							1	chr1:247274585-247275757	N_Shelf
cg16328462	-2.6089	0.55	0.000141	0.045338							8		
cg16415931	-1.4967	0.295	6.74E-05	0.021633					ZBED5;ZBE	5'UTR;1stE	11	chr11:10879102-10880453	Island
cg16519192	1.0978	0.2116	5.24E-05	0.016813							12		
cg16575125	-0.5879	0.1193	9.33E-05	0.029938					PLEKHG4B	Body	5	chr5:169578-169798	S_Shelf
cg16848072	-2.927	0.6205	0.00015	0.048139							3		
cg16914277	-2.3503	0.4367	3.41E-05	0.010955					HLA-DOA	TSS1500	6	chr6:32975684-32975926	S_Shelf
cg17253785	-0.8776	0.1833	0.000128	0.041125					C8orf44	TSS1500	8		
cg17287974	-1.6002	0.3312	0.000116	0.037256							2		
cg17448109	-1.3295	0.2826	0.000154	0.049488					RNF115	Body	1		
cg17454086	-1.2814	0.2452	4.81E-05	0.01545					KIAA0907	Body	1	chr1:155904072-155904307	N_Shore
cg17470103	-2.4262	0.4784	6.78E-05	0.021772							1	chr1:151253658-151255769	N_Shore
cg17476389	-1.9708	0.365	3.29E-05	0.010545					TTC15	Body	2		
cg17529235	-2.3537	0.4614	6.34E-05	0.02036					C17orf51	Body	17	chr17:21454388-21454938	N_Shelf
cg17563032	-1.096	0.2318	0.000146	0.047005					SRRT;SRRT	Body;Body	7		
cg17606881	-1.2973	0.2735	0.000142	0.045445							3	chr3:172468372-172468845	S_Shore
cg17851795	-2.5593	0.5179	9.05E-05	0.02905					PBX2;GPSM	TSS1500;3	6		
cg17856830	-0.8911	0.1784	8.04E-05	0.025808							1		
cg17879648	-2.783	0.5249	4.07E-05	0.013051							18	chr18:14474587-14475008	N_Shore
cg17890283	-2.115	0.3903	3.15E-05	0.010113							8	chr8:80942117-80942593	S_Shelf
cg17993073	-1.3552	0.2657	6.35E-05	0.020381					IGF2R	3'UTR	6		
cg18002076	-2.3792	0.4362	2.91E-05	0.009349					KIAA0947	Body	5		
cg18308339	-1.3389	0.2233	9.06E-06	0.002907							5		
cg18477664	-2.2401	0.4535	9.10E-05	0.029219							7	chr7:155898140-155898386	S_Shelf
cg18481642	-2.6992	0.4812	2.08E-05	0.006668					STAT4	Body	2		
cg18821281	-0.8715	0.145	8.80E-06	0.002824					VGLL4;VGI	Body;1stE>	3	chr3:11684192-11684930	S_Shore
cg18983417	-2.1852	0.3709	1.13E-05	0.003631							15		
cg19194448	-1.8337	0.3611	6.68E-05	0.021439					RBM45	TSS1500	2	chr2:178977078-178977828	N_Shore
cg19236703	-2.155	0.3059	1.05E-06	0.000336					KCNQ2;KC	Body;Body	20	chr20:62079615-62079912	Island
cg19385365	-0.7872	0.1658	0.00014	0.045055					DGKI	Body	7		
cg19457770	-1.6858	0.3382	8.23E-05	0.026432					IL1F9	3'UTR	2		
cg19752083	-1.4714	0.2951	8.19E-05	0.0263					BAT2	Body	6	chr6:31599519-31599761	N_Shore
cg19891951	-1.3769	0.2624	4.58E-05	0.014698					DDN	TSS1500	12	chr12:49393139-49393361	Island
cg19913448	-1.0509	0.1806	1.32E-05	0.004239					GSR	TSS200	8	chr8:30584790-30585651	Island
cg19998289	-2.2705	0.4062	2.17E-05	0.006954					RNF34;RN	3'UTR;3'U	12		
cg20238308	-0.9452	0.1876	7.30E-05	0.023444					PTPRQ	TSS1500	12		
cg20268279	-1.2463	0.2384	4.79E-05	0.01536							14		
cg20278840	-1.7354	0.3449	7.41E-05	0.023798					TMTC1	Body	12		
cg20405508	-0.7329	0.1436	6.29E-05	0.020194					PDE6D	TSS1500	2	chr2:232645313-232646505	S_Shore
cg20848785	-1.4501	0.2728	3.95E-05	0.012668					VIPR2	Body	7	chr7:158886367-158886595	Island
cg20943039	0.9471	0.174	2.98E-05	0.009564							15		
cg21343777	-0.9616	0.188	6.15E-05	0.019747					RGS12;RG	Body;Body	4	chr4:3341783-3342021	N_Shore
cg21480464	-1.2647	0.2687	0.000153	0.049253					PEMT;PEM	Body;Body	17		
cg21700440	-0.4798	0.09635	8.31E-05	0.026677					HS3ST3B1	Body;TSS1	17	chr17:14204168-14207702	Island
cg21825397	-2.1304	0.4451	0.000129	0.041254							20		
cg22047910	-0.8689	0.1618	3.51E-05	0.011279					SLC39A10;	5'UTR;5'U	2	chr2:196521555-196522950	Island
cg22067069	-0.7032	0.1475	0.000134	0.042915					HSD17B4	TSS1500	5	chr5:118788125-118788428	N_Shore
cg22157525	-1.5517	0.306	6.78E-05	0.021749					NDST1	Body	5		
cg22287064	-1.3177	0.2662	8.89E-05	0.028553					MYO15B	TSS200	17	chr17:73583838-73586337	Island
cg22331108	-1.6337	0.2249	6.80E-07	0.000218							2	chr2:232478359-232479925	Island
cg22697572	0.6922	0.1431	0.000114	0.03662					LOC65434	TSS200	2		
cg22949256	-1.4867	0.2957	7.47E-05	0.023981					SGIP1	Body	1		
cg22966316	-1.6837	0.3442	0.000101	0.032548					TTC1	5'UTR	5		
cg23374847	-3.0824	0.6247	9.21E-05	0.02957							2		
cg23671600	1.4123	0.2618	3.32E-05	0.010668					FBXO6	5'UTR	1	chr1:11724113-11724885	Island
cg23761017	1.3371	0.2831	0.000148	0.047462							4	chr4:185937242-185937750	N_Shore
cg23978322	-1.9565	0.3777	5.32E-05	0.017072					FHL5;FHL5	1stExon;5'	6		

CpG	Estimate	StdErr	RAW_P	ahoc_p	Estimate	StdErr	RAW_P	ahoc_p	UCSC_Re	UCSC_Re	CHR	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island
cg24054668	-1.646	0.2802	1.17E-05	0.003759					IQCE;IQCE	Body;Body	7		
cg24128590	1.0482	0.2054	6.32E-05	0.020289					C14orf43;(	3'UTR;3'UT	14	chr14:74185482-74185994	N_Shore
cg24391991	-0.9326	0.1933	0.000118	0.037822					MAP3K8;N	1stExon;5'	10	chr10:30722378-30723707	Island
cg24823751	0.6716	0.1389	0.000115	0.036956					MIR1915;(	TSS1500;T	10	chr10:21783198-21786420	Island
cg24843246	-0.906	0.1903	0.000136	0.043565					PNPLA6;PI	Body;Body	19	chr19:7621639-7622262	Island
cg24868248	-0.8812	0.1619	2.99E-05	0.009604					SFRS1;SFR	TSS1500;T	17	chr17:56083750-56085373	Island
cg24883586	1.3478	0.2856	0.00015	0.04799					RAB11FIP1	TSS200;TS	8	chr8:37756454-37757339	Island
cg25008393	-0.9218	0.1888	0.000103	0.033153					PFDN2;NIT	Body;TSS1	1	chr1:161087722-161088112	N_Shore
cg25023095	1.6941	0.3569	0.00014	0.045061					RRP12;RRI	Body;Body	10		
cg25130710	-1.0804	0.2175	8.56E-05	0.027473					FBRSL1	Body	12	chr12:133159225-133160576	Island
cg25257018	-2.6533	0.5329	8.33E-05	0.02675					KDM4DL	TSS1500	11		
cg25306087	-1.1236	0.2013	2.21E-05	0.007086					OTOP2;US	TSS200;TS	17	chr17:72918995-72921019	Island
cg25312876	-0.9213	0.1928	0.000131	0.041959					KDM4B	3'UTR	19	chr19:5151332-5151730	Island
cg25321332	-3.1363	0.6635	0.000147	0.047125							18		
cg25404410	-1.6001	0.3256	9.65E-05	0.030963					KIF3A	TSS1500	5	chr5:132072695-132073429	S_Shore
cg25453957	-0.9757	0.2014	0.000113	0.036121							17		
cg25584787	0.8155	0.1365	9.50E-06	0.003048					C5orf36	Body	5		
cg25679743	-2.8829	0.5776	8.11E-05	0.026047							19	chr19:42412375-42412584	Island
cg25871713	-2.3077	0.4592	7.51E-05	0.024117					ITGA9	Body	3		
cg26122004	-1.2898	0.249	5.32E-05	0.017091					C5orf45;C	Body;Body	5	chr5:179276846-179277068	Island
cg26606257	-1.3614	0.2829	0.000121	0.03886					MYT1L	Body	2		
cg27040463	-0.6793	0.1228	2.46E-05	0.007896					ABHD13;LI	5'UTR;TSS	13	chr13:108870502-108871328	Island
cg27055313	-0.5635	0.1188	0.000141	0.045259							14	chr14:106025533-106026386	Island
cg27324576	-1.1763	0.2473	0.000137	0.044019					VEPH1;VEI	Body;Body	3		
cg27417659	-1.5031	0.2939	6.16E-05	0.019789					TRRAP	Body	7		
cg27420264	-0.8188	0.1662	9.33E-05	0.029965					HSPA8;HSI	TSS200;TS	11	chr11:122932173-122933803	Island
ch_10_1024	-1.101	0.2143	5.86E-05	0.018807									
ch_4_13366	-0.8257	0.1534	3.40E-05	0.010912									
ch_5_19060	-1.2736	0.2484	6.00E-05	0.019255									

Table A2.2

CpG	ICC	P value
cg11029358	0.979056	4.45E-07
cg12349837	0.961259	1.60E-06
cg26015416	0.944679	1.39E-05
cg17207545	0.937823	2.46E-05
cg04385523	0.931517	0.000117
cg02299937	0.908664	1.29E-05
cg18469813	0.901578	9.13E-05
cg22590761	0.894415	2.65E-05
cg01750170	0.889368	0.000299
cg25288034	0.866309	0.000148
cg08482979	0.840104	0.00124
cg10083824	0.836534	0.001212
cg24699146	0.824922	0.000366
cg07708818	0.814555	0.000392
cg25637226	0.812584	0.002014
cg10002668	0.774514	0.00418
cg05132568	0.771446	0.002125
cg01769501	0.766981	0.004836
cg20458779	0.761597	0.004949
cg15549502	0.74647	0.001206
cg24600706	0.744432	0.001687
cg13984351	0.73477	0.00267
cg18773129	0.734442	0.001436
cg16383005	0.731928	0.006281
cg25499537	0.725816	0.008505
cg04990210	0.724765	0.003261
cg06829645	0.715809	0.002084
cg02448934	0.71478	0.001951
cg07805777	0.687972	0.008096
cg25471923	0.686291	0.002853
cg20075700	0.677139	0.007397
cg16019434	0.670026	0.016181
cg02346442	0.666076	0.008775
cg20431441	0.665892	0.009213
cg24677222	0.654635	0.019408
cg06998210	0.651392	0.005746
cg26624744	0.640641	0.008537
cg11413778	0.637632	0.00811
cg04614823	0.625649	0.009757
cg15367212	0.619501	0.007928
cg11823603	0.606962	0.027105
cg10548968	0.606855	0.007192
cg10343024	0.606723	0.02734
cg16998490	0.598492	0.031624
cg12143439	0.595929	0.01459
cg04723493	0.595728	0.011669

CpG	ICC	P value
cg06382167	0.593588	0.008852
cg25488567	0.593115	0.011583
cg12765716	0.588721	0.010329
cg11735008	0.579966	0.013543
cg00931181	0.574356	0.009914
cg08392484	0.569442	0.036974
cg05590053	0.559366	0.012538
cg00573504	0.539561	0.020963
cg02751327	0.535858	0.015777
cg02030454	0.524532	0.01608
cg13346967	0.519054	0.017335
cg11380128	0.513535	0.043542

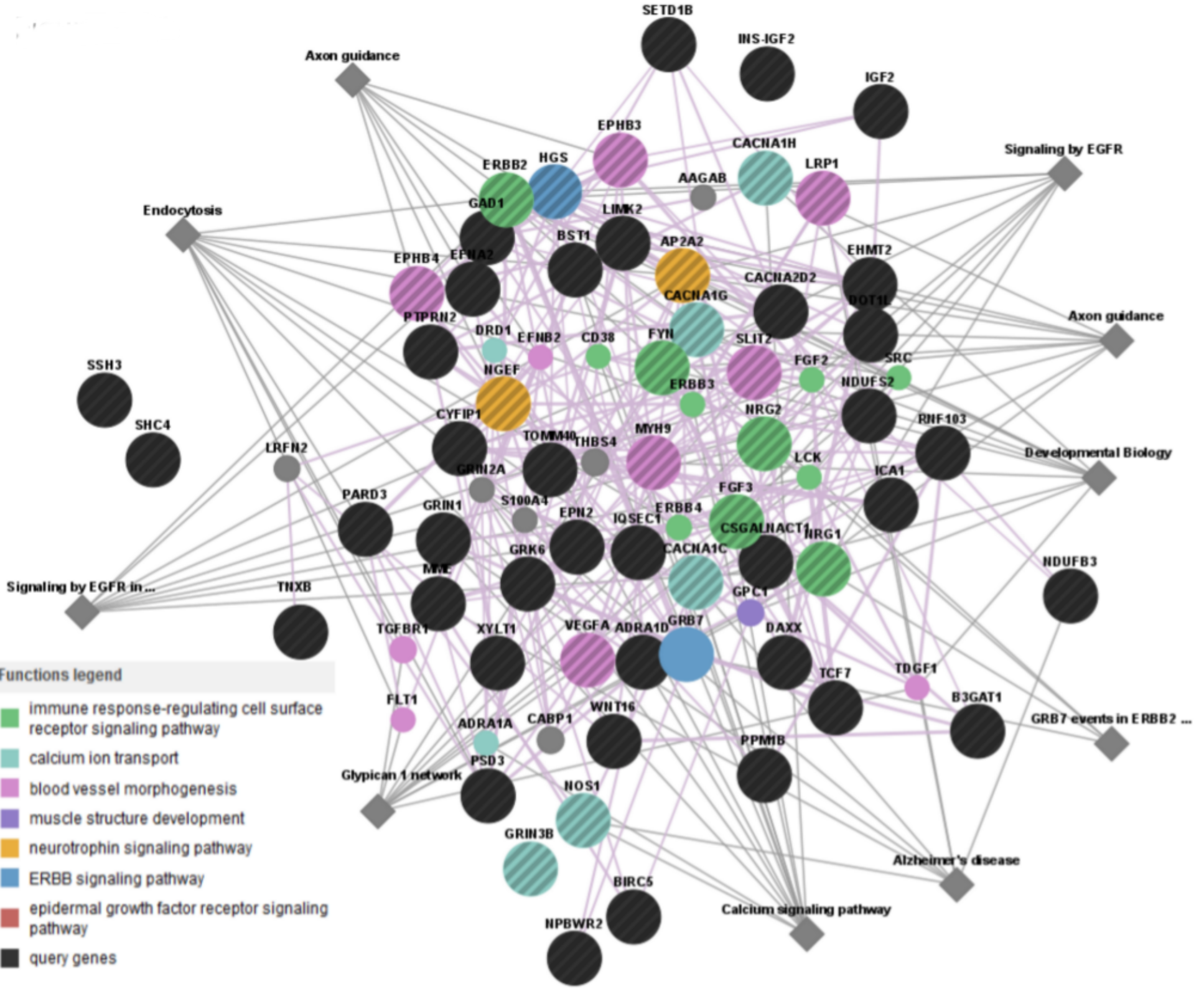

SETD1B
INS-IGF2
IGF2
Axon guidance
EPHB3
CACNA1H
Signaling by EGFR
ERBB2
HGS
LRP1
AAGAB
Endocytosis
GAD1
LIMK2
BST1
AP2A2
EHMT2
EPHB4
EFNA2
CACNA2D2
CACNA1G
DOT1L
PTPRN2
DRD1
EFNB2
CD38
FYN
SLIT2
Axon guidance
SSH3
NGEF
FGF2
SRC
ERBB3
NDUFS2
SHC4
CYFIP1
NRG2
RNF103
LRFN2
TOMM40
THBS4
MYH9
LCK
Developmental Biology
GRIN2A
ICA1
PARD3
S100A4
FGF3
GRIN1
EPN2
IQSEC1
ERBB4
CSGALNACT1
NRG1
GRK6
CACNA1C
NDUFB3
Signaling by EGFR in ...
MME
TNXB
GPC1
TGFBR1
XYLT1
VEGFA
ADRA1D
GRB7
DAXX
TCF7
TDGF1
FLT1
B3GAT1
ADRA1A
CABP1
WNT16
GRB7 events in ERBB2 ...
PPM1B
PSD3
Glypican 1 network
NOS1
GRIN3B
Alzheimer's disease
BIRC5
Calcium signaling pathway
NPBWR2
Functions legend
immune response-regulating cell surface receptor signaling pathway
calcium ion transport
blood vessel morphogenesis
muscle structure development
neurotrophin signaling pathway
ERBB signaling pathway
epidermal growth factor receptor signaling pathway
query genes

Figure A1.1

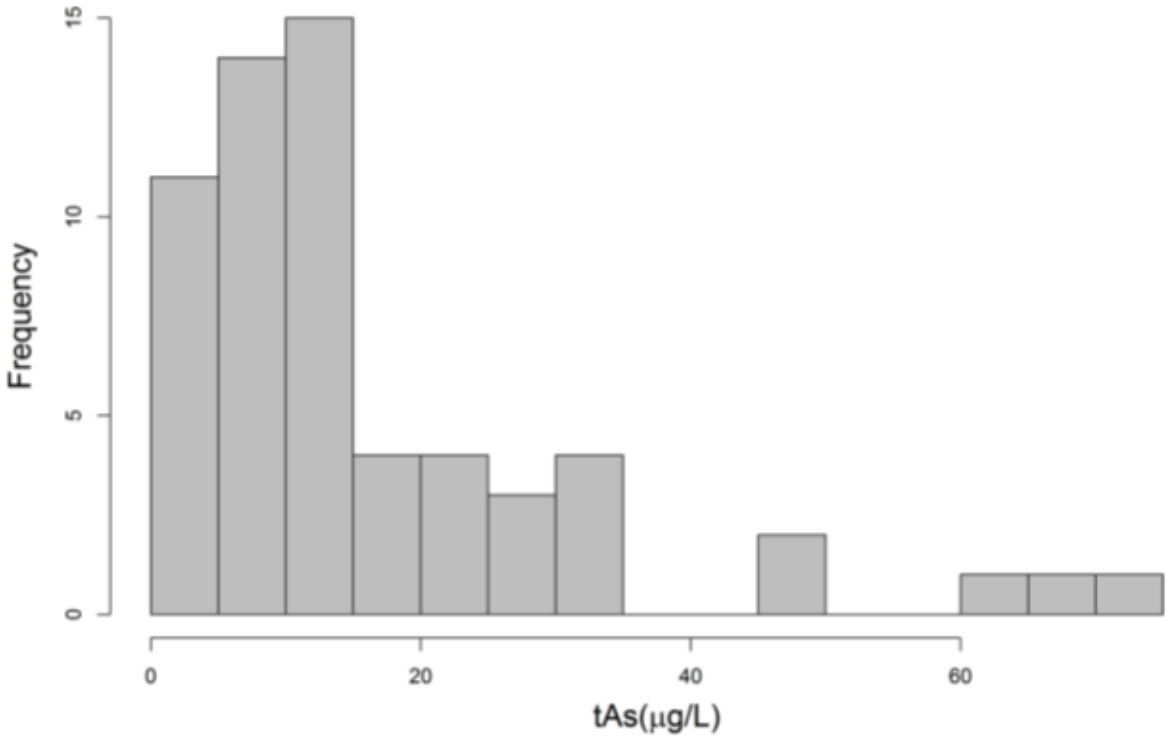
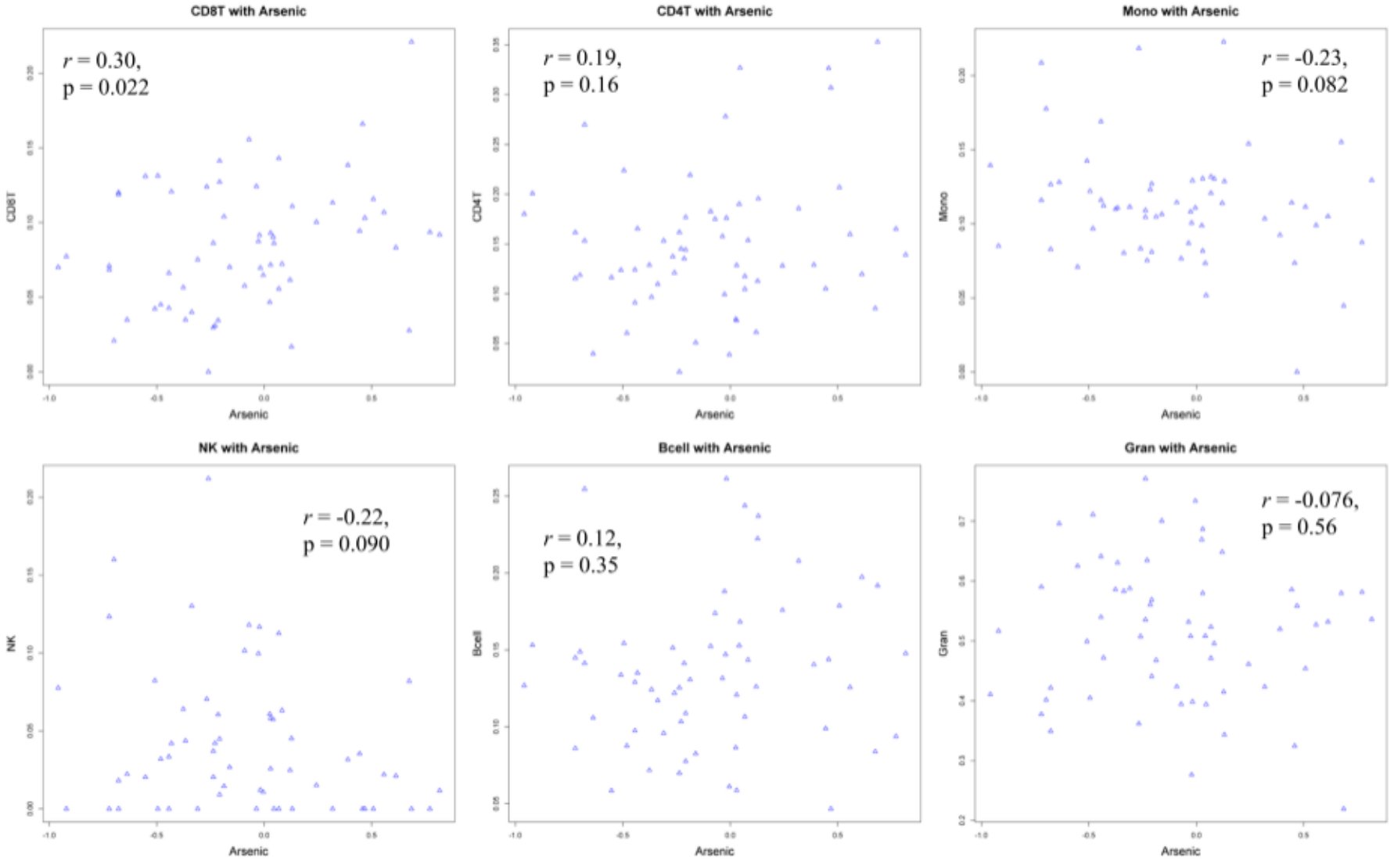


Figure A1.2



. Pearson correlations between **inorganic arsenic levels** (in log10 scale) and cell type proportions.

Figure A3.1

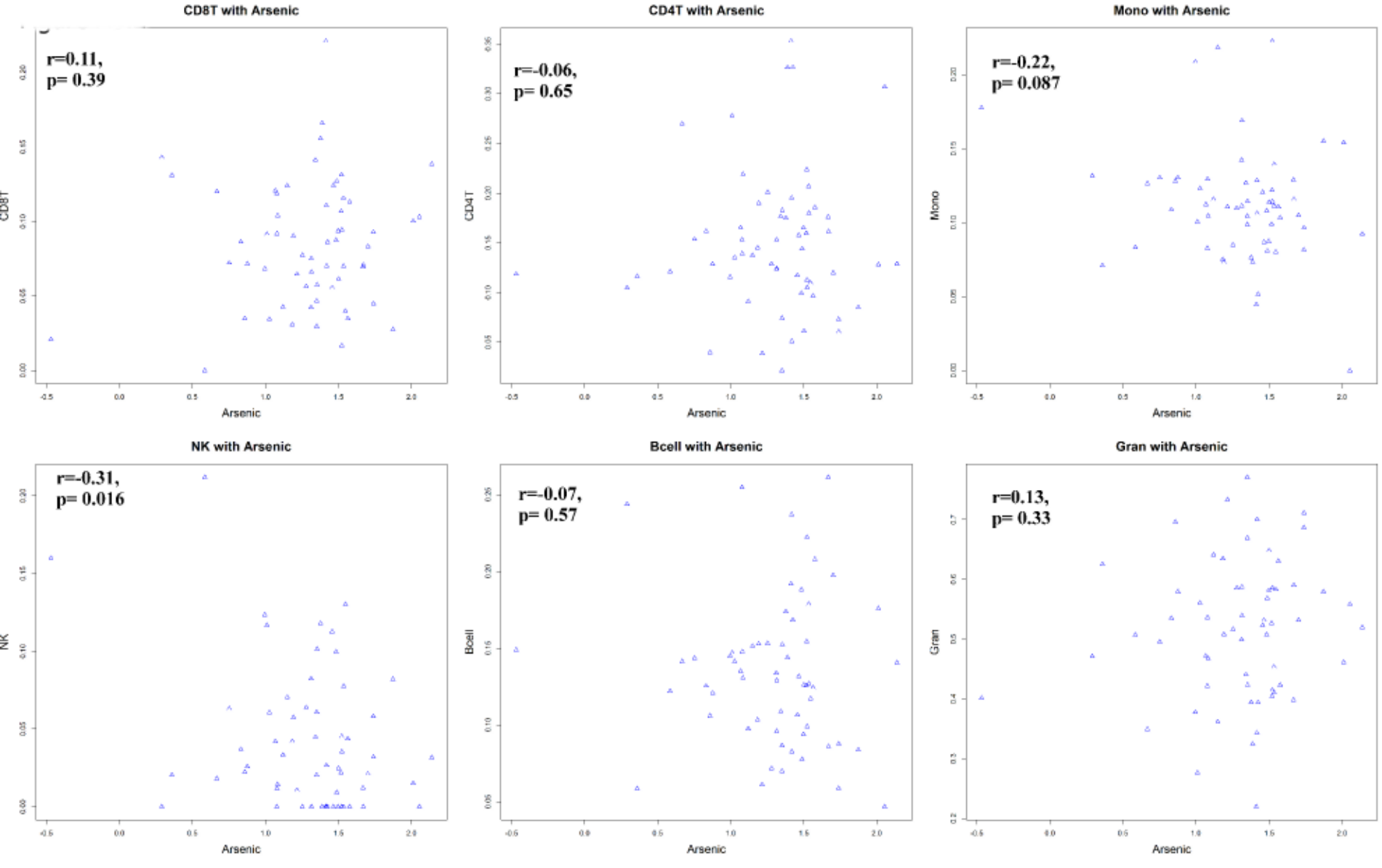

CD8T with Arsenic
r=0.11,
p= 0.39
CD8T
Arsenic
CD4T with Arsenic
r=-0.06,
p= 0.65
CD4T
Arsenic
Mono with Arsenic
r=-0.22,
p= 0.087
Mono
Arsenic
NK with Arsenic
r=-0.31,
p= 0.016
NK
Arsenic
Bcell with Arsenic
r=-0.07,
p= 0.57
Bcell
Arsenic
Gran with Arsenic
r=0.13,
p= 0.33
Gran
Arsenic

Figure A3.2

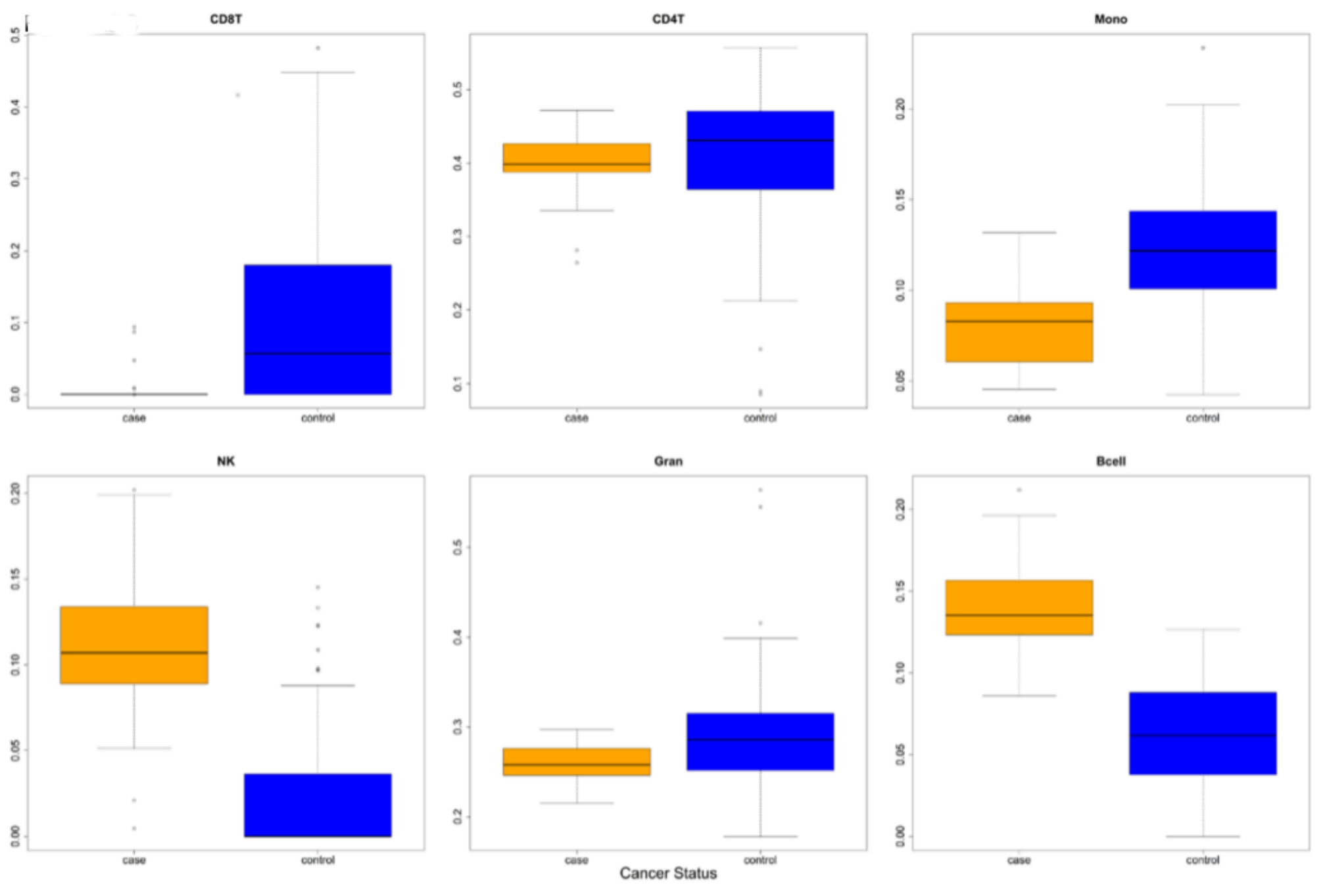


Figure A3.3

```
############################################################
### A) Installing and loading required packages
############################################################

if (!require("sva")) {
   install.packages("sva", dependencies = TRUE)
   library(sva)
   }
if (!require("limma")) {
   install.packages("limma", dependencies = TRUE)
   library(limma)
   }
if (!require("MASS")) {
   install.packages("MASS", dependencies = TRUE)
   library(MASS)
   }
if (!require("gplots")) {
   install.packages("gplots", dependencies = TRUE)
   library(gplots)
   }
if (!require("RColorBrewer")) {
   install.packages("RColorBrewer", dependencies = TRUE)
   library(RColorBrewer)
   }
if (!require("LPS")) {
   install.packages("LPS", dependencies = TRUE)
   library(LPS)
   }
if (!require("dendextend")) {
   install.packages("dendextend", dependencies = TRUE)
   library(dendextend)
   }
if (!require("dendextendRcpp")) {
   install.packages("dendextendRcpp", dependencies = TRUE)
   library(dendextendRcpp)
   }
if (!require("colorspace")) {
   install.packages("colorspace", dependencies = TRUE)
   library(colorspace)
   }
if (!require("VennDiagram")) {
   install.packages("VennDiagram", dependencies = TRUE)
   library(VennDiagram)
   }
#########################################################################
### B) Reading in data and transforming DNA-m beta values to M-values
#########################################################################
setwd("~User Directory")
beta<-read.csv("DNA-m filename.csv",header=T)  ##read in the DNA-m beta values
covar<-read.csv("covariate filename.csv", header=T)

beta1=beta[,-1]
rownames(beta1)=beta[,1]

miss = which(is.na(beta),arr.ind=TRUE)[,1];
beta1 = beta1[-miss,];
M_val = as.matrix(log2(beta1/(1-beta1)); ##Logit transforming Beta value to M-values

##Read in the covariates from covar dataframe into specific variables

covar1=covar[,2];covar2=covar[,3];covar1=covar[,4];etc.
mod=model.matrix(~covar1+covar2+covar3)
```

```
#############################################################################
### C) Performing Robust Regression
#############################################################################
fit<-lmFit(M-val,mod,method="robust")
fit1<-eBayes(fit1)
tab <- topTable(fit1, coef = "covar1",number=length(beta1[,1]), p.val=0.05,adjust =
"fdr")
write.table(tab,file='Robust_regression_results.csv',sep=",",row.names=TRUE)

#############################################################################
### D) Estimating Surrogate Variables (SVA) and using it in regression
#############################################################################
mod0 = model.matrix(~covar2+covar3)
svobj1 = sva(edata,mod,mod0,n.sv=NULL,method="two-step")
modSv = cbind(mod,svobj1$sv)
fit2<-lmFit(M_val,modSv,method="robust")
fit21<-eBayes(fit2)
tab21 <- topTable(fit21, coef = "covar1",number=length(beta1[,1]), p.val=0.05,adjust
= "fdr")

#############################################################################
### D) Estimating Cell proportions using minfi
#############################################################################
source("https://bioconductor.org/biocLite.R")
biocLite(c("minfi","quadprog","FlowSorted.Blood.450k",
"IlluminaHumanMethylation450kmanifest",
"IlluminaHumanMethylation450kanno.ilmn12.hg19"));

lib = c("minfi","quadprog","FlowSorted.Blood.450k",
"IlluminaHumanMethylation450kmanifest",
"IlluminaHumanMethylation450kanno.ilmn12.hg19");
lapply(lib, require, character.only = TRUE);

grSet1=read.table("input_data.txt",header=T); grSet=grSet1[,-1];
rownames(grSet)=grSet1[,1]; grSet=data.matrix(grSet);
referenceMset = get('FlowSorted.Blood.450k.compTable');

cell = c("CD8T","CD4T", "NK","Bcell","Mono","Gran","Eos");
compData = minfi:::pickCompProbes(referenceMset, cellTypes=cell);
coefs = compData$coefEsts;
coefs = coefs[ intersect(rownames(grSet), rownames(coefs)), ];
rm(referenceMset);

counts = minfi:::projectCellType(grSet[rownames(coefs), ], coefs);
rownames(counts) = colnames(grSet);
write.table(counts,file="Ewasher_data_minficell.csv",sep=",");

#############################################################################
### D) Estimating Cell proportions using Houseman
#############################################################################
##Visit this website
https://drive.google.com/folderview?id=0B6IN3-9RV-LBeUt5bENMSXlsY2s&usp= sharing
##to download all the relevant files

beta=t(as.matrix(read.table(".../input_data.txt",sep="\t",header=T)))

source("Rcodes_Cell_mixture.R")

#############################################################################
### E) Using Refactor to adjust for Cell types
#############################################################################

##Download the Refactor source code from following link:
```

```
http://www.cs.tau.ac.il/~heran/cozygene/software/refactor.html

sd1=apply(beta1, 1, sd, na.rm = F)
include = which(sd1>=quantile(sd1,0.05))
O = beta1[include,]
M_val1 = M_val[include,]

cpgnames <- rownames(O)
for (site in 1:nrow(O))
 {
 model <- lm(O[site,] ~ covar1+covar2+covar3)
 O_adj[site,] = residuals(model)
 }
 O = O_adj
 print('Running a standard PCA...')
 pcs = prcomp(scale(t(O)));

 coeff = pcs$rotation
 score = pcs$x

 print('Compute a low rank approximation of input data and rank sites...')
 x = score[,1:7]%*%t(coeff[,1:7]);
 An = scale(t(O),center=T,scale=F)
 Bn = scale(x,center=T,scale=F)
 An = t(t(An)*(1/sqrt(apply(An^2,2,sum))))
 Bn = t(t(Bn)*(1/sqrt(apply(Bn^2,2,sum))))

# Find the distance of each site from its low rank approximation.
 distances = apply((An-Bn)^2,2,sum)^0.5 ;
 dsort = sort(distances,index.return=T);
 ranked_list = dsort$ix

 print('Compute ReFACTor components...')
 sites = ranked_list[1:500];
 pcs = prcomp(scale(t(O[sites,])));
 first_score <- score[,1:7];
 score = pcs$x


mod=model.matrix(~covar1+covar2+covar3+first_score)
fit1<-lmFit(M_val1,mod,method="ls")
fite1<-eBayes(fit1)
tab1 <- topTable(fite1, coef = "covar1",number=length(beta1[,1]), p.val=0.05,adjust
= "fdr")

write.table(tab1,file="Refactor_results.csv",sep=",",row.names=T)

########################################################################
### F) Using RefFreeCellMix to adjust for Cell types
########################################################################

library(RefFreeEWAS)
cell<-RefFreeCellMix(beta1,mu0=NULL,K=7,iters=5,Yfinal=NULL,verbose=TRUE)
mod1<-model.matrix(~covar1+covar2+covar3+cell$Omega)

fit1<-lmFit(M_val,mod1,method="robust")
fite1<-eBayes(fit1)
tab1 <- topTable(fite1, coef = "covar1",number=length(beta1[,1]), p.val=0.05,adjust
= "fdr")

write.table(tab1,file="RefFreeCellMix_results.csv", sep=",",row.names=T)

########################################################################
```

```
### F) Using RUV to adjust for Cell types
##########################################################################

library(data.table)
library(psych)
library(pracma)

##Specify beta values of top 600 CpG sites based upon the reference dataset

beta<-read.csv("Top_600_Minfi_27k.csv",header=T) ##Specify Top 600
beta1=beta[,-1]
rownames(beta1)=beta[,1]
beta1_t<-t(beta1)

pca_mval<-prcomp(beta1_t,center=T,scale.=T)
plot(pca_mval, type = "l")

###Based upon scree plot we can choose number of PC's##
rawLoadings <- pca_mval$rotation[,1:7] %*% diag(pca_mval$sdev, 7, 7)
rotatedLoadings <- varimax(rawLoadings,normalize = TRUE, eps = 1e-5)$loadings
invLoadings <- t(pracma::pinv(rotatedLoadings))
scores <- scale(beta1_t) %*% invLoadings

###x1,x2,x3,x4,x5 and x6 are primary and secondary covariates,
###PC1,PC2, PC3 and PC4 are principal components
##########################################################################################
######
mod<-model.matrix(~x1+x2+x3+x4+x5+x6+PC1+PC2+PC3+PC4)
fit<-lmFit(edata,mod,method="robust")
fite<-eBayes(fit)
tab <- topTable(fite, coef = "x1",number=length(edata[,1]), p.val=0.05,adjust =
"fdr")

##########################################################################################
#######
### F) Bar plots
##########################################################################################
#######

data<-read.csv("coef.csv",header=T)
data1=data[,-1]
rownames(data1)=data[,1]

category<-factor(colnames(data1))

cols<-c("orange1", "blue")[category]

barplot(t(data1),beside=T,col = cols,width = 0.82,space = NULL,names.arg = NULL,
legend.text =              NULL, horiz = FALSE, density = NULL,las=2,border =
par("fg"),main = NULL, sub = NULL             , xlab = NULL, ylab = NULL,xlim =
NULL, ylim = c(-0.5,0.5), xpd = TRUE, log = "",axes =        TRUE, axisnames = TRUE,
plot = TRUE,cex.names = 0.85, axis.lty = 1,offset = 0,add = FALSE        ,
args.legend = NULL)
abline(h=0)
title("Bar plot","","", "Methylation")
text(seq(1, 29, by=1), par("usr")[3] - 0.2, labels = rownames(data1), srt = 90, pos
= 2,offset =     0, vfont = NULL,col = NULL, font = NULL)

axis(1, at=seq(1, 29, by=1),tick = F,line = NA,pos = NA, outer = FALSE, font = NA,
     lty = "solid",labels = FALSE, lwd = 1.5,col = NULL, col.ticks = NULL,hadj = NA,

     padj = NA,lwd.ticks=1)
legend(1,0.5,c("CAT","HDM"),lty=c(1,1),lwd=c(3,3),col=c("orange1", "blue"))
```

```
box(bty="l")

#############################################################################################
###########
### F) Manhattan plots
#############################################################################################
###########

##Please refer to this website
http://genome.sph.umich.edu/wiki/Code_Sample:_Generating_Manhattan_Plots_in_R


#############################################################################################
###########
### G) Heatmap
#############################################################################################
###########

png("Plotname.png",    # create PNG for the heat map
  width = 5*300,        # 5 x 300 pixels
  height = 7*300,
  res = 300,            # 300 pixels per inch
  pointsize = 5)        # smaller font size

Rowv<-mat_data %>% dist %>% hclust %>% as.dendrogram %>%
   set("branches_k_color", k = 5) %>% set("branches_lwd", 1.5) %>%
   ladderize


heatmap.2(mat_data,    ###mat_data contains the correlation values or beta values
  cex.main=5.0,
  key=T,
  keysize=0.85,
  symkey=F,
  main = "Correlation", # heat map title
  notecol="black",      # change font color of cell labels to black
  density.info="density",  # turns off density plot inside color legend
  trace="none",         # turns off trace lines inside the heat map
  margins =c(13,11),     # widens margins around plot
  col=my_palette,       # use on color palette defined earlier
  breaks=col_breaks,
  dendrogram="row",     # only draw a row dendrogram
  scale="none",
  Rowv=Rowv,
  Colv="NA",            # turn off column clustering
  cexCol=2.0,
  cexRow=2.0
  )

dev.off()

#############################################################################################
########
### G) Venn Diagrams
#############################################################################################
########
ars<-read.csv("DNA-m data.csv",header=T)
venn.plot<-venn.diagram(
x = list(
SVA = ars$SVA,
Houseman = ars$Houseman,
minfi = ars$Minfi,
```

```
RefFreeEWAS = ars$RefFreeEWAS,
RefFreeCellMix = ars$RefFreeCellMix
),
filename = 'Venn_Ewasher1.png',
output = TRUE,
height = 35,
width = 65,
resolution = 300,
units = 'in', cat.pos = c(0,310, 215, 145, 50),
rotation.degree=15,
fill = gray.colors(5, start = 0.2, end = 0.9, gamma = 1.5, alpha = 0.7),
lty = "solid",
cex=5,
cat.cex=7,
Scaled=TRUE
);
```

```
                    SAS code for Epigenome wide mixed modeling
ods graphics off; ods html close; ods listing close;

/*DISPLAY P-VALUES TO A HIGHER NUMBER OF DECIMAL PLACES*/
ods path sasuser.templat(update) sashelp.tmplmst(read);
proc template;
   edit Common.PValue;
      notes "Default p-value column";
      just = r;
      format = pvalue15.9;
   end;
run;
/*
/  Unless you want to keep this edited template, delete it.
/  this only deletes the version of the template in your SASUSER
/  library, returning you to the regularly scheduled official one in
/  SASHELP.
/-----------------------------------------------------------------------*/

/* The first 5 rows of data below are ID and covariates and next rows are
DNA-methylation data*/
/* I have two rows for ID by the names ID_POS and ID*/

PROC IMPORT OUT= WORK.IgEt
            DATAFILE= "C:\Users\akaushl1\Documents\IgE_MIXED\iow11.csv"
            DBMS=CSV REPLACE;

RUN;


Data Work.IgEt1;
 SET Work.IgEt (obs=5);
run;
Data Work.IgEt2;
 SET Work.IgEt (firstobs=6);
run;

Data work.IgEt;
  Set Work.IgEt1 Work.IgEt2;
run;
proc transpose data= Work.IgEt let
     out= IgE_new (where=(upcase(_name_) ne 'ID_POs'));
     var _all_;
   id id_pos;
run;

Data IgE_new1;
Set IgE_new(firstobs=2);
run;

DATA IgE_new_long ;
  SET IgE_new1;
  IgE = serum_ige_10; time = 10; OUTPUT;
  IgE = serum_ige_18; time = 18; OUTPUT;

  DROP serum_ige_10 serum_ige_18;
RUN;

Data IgE_new_long;
 Set IgE_new_long;
 if IgE<0 then IgE=.;
 run;
data IgE_new_long1;
  set IgE_new_long;
```

```
  tcat=time;
  newIgE = input(IgE,best32.);
  cbige=input(cb_ige,best32.);
  bw=input(BIRTHWT,best32.);
  lIgE=log10(newIgE);
  drop IgE newIgE cb_ige;
  rename lIgE=IgE;

run;
proc print data= IgE_new_long1;run;

 data IgE1_new;
  set IgE_new1(drop= serum_ige_10 serum_ige_18  _NAME_ ID cb_ige BIRTHWT);
 run;

proc contents data = IgE1_new
out = vars(keep = varnum name)
noprint;
run;
proc print data= IgE1_new;run;


data _null_;

   mVars +1;

   len = 0;

   length str $10;

   call execute( cats( '%str( %%let name)', mVars , '=' ) );

   do WHILE( NOT eof );

      set work.vars(keep= name) end= eof;

      have+1;


          str = trim( name );

       len + ( 3 + length(str) );

      if EOF OR len gT 20000 then LEAVE;

   end;

   call execute( trim(str) !!' ; ' );

   putlog 'info: ' have= 'in ' mVars=;

   if eof then

      do;

         call symputx( 'n_name_vars', mVars );

         stop;

      end;

run;
```

SAS code for Epigenome wide mixed modeling

```
%macro split;
%do i = 1 %to &n_name_vars;
data IgE1_new&i.;
set IgE1_new (keep=name&i);
run;
%end; ;
%mend split;
%split;
/*
data IgE1_new1;
 set IgE1_new (keep=&x1);
 run;
*/

/* Macro to split the large dataset into bunch of small dataset */

%macro split1;
%do i = 1 %to &n_name_vars;

proc contents data = IgE1_new&i.
out = vars&i.(keep = name type)
noprint;
run;
%end;;
%mend split1;
%split1;

proc print data=vars1;run;

%macro split2;

%do k = 1 %to &n_name_vars;
data vars&k.;
set vars&k.;
if type=2 and name ne 'id';
newname=trim(left(name))||"_n";



proc sql noprint;
select trim(left(name)), trim(left(newname)),
       trim(left(newname))||'='||trim(left(name))
into :c_list separated by ' ', :n_list separated by ' ',
     :renam_list separated by ' '
from vars&k.;



data IgE_new_long2&k.;
set IgE_new_long1 (keep=IgE ID cbige TIME TCAT bw &&name&k.);
array ch(*) $ &c_list;
array nu(*) &n_list;
do i = 1 to dim(ch);
  nu(i)=input(ch(i),5.);
end;
drop i &c_list;
rename &renam_list;

run;

options nosymbolgen;
%end;
```

```
%mend split2;
%split2;

proc print data=IgE_new_long21;run;


%macro mylogita(indata, indvars, dep, myout =_out );
  %let k=1;
  %let ind = %scan(&indvars, &k);
  %do %while(&ind NE);
    title "The dependent variable is &dep";
    title2 "The independent variables are &ind";

ODS output SolutionF = est1&k Tests3=est2&i;

proc mixed data=&indata method=ML ;
      class ID;
      model &dep= &in cbige time bw/solution ddfm=bw;

      random int time/Subject=ID type=AR(1);*R=1,2 RCORR;
run;

ODS OUTPUT CLOSE;

%let k = %eval(&k + 1);
    %let ind = %scan(&indvars, &k);
  %end;

  data &myout;
    set
    %do i = 1 %to &k - 1;
      est1&i est2&i
    %end;
    ;
  run;

%mend;

*run the program;

%macro split3;

%do j = 1 %to &n_name_var;
%mylogita(work.IgE_new_long2&j., &&name&j., IgE, myout = myparms&j.)
%end;
%mend split3;
%split3;

*ods trace off;

%macro split4;

%do k = 1 %to &n_name_vars;

Data myparm1 (Keep = Effect Estimate StdErr df);
 Set myparms&k.(where=(probt NE . and Effect not in
('time','Intercept','tcat','cbige','bw')));
 run;
title;
Data myparm2 (Keep = Effect probf);
 Set myparms&k.(where=(probf NE . and Effect not in
('time','Intercept','tcat','cbige','bw')));
```

```
 run;

* 1. Sort myparms1 by "Effect" & save sorted file as mayparms11 ;
PROC SORT DATA=myparm1 OUT=myparms11;
  BY Effect;
RUN;

* 2. Sort myparms2 by "Effect" & save sorted file as mayparms21 ;

PROC SORT DATA=myparm2 OUT=myparms21;
  BY Effect;
RUN;

* 3. Merge myparms11 and myparms21 by Effect in a data step ;

DATA myparms12&k. ;
  MERGE myparms11 myparms21;
  BY Effect;
RUN;
%end;
%mend split4;
%split4;


%macro combine;

data big;
  set
  %do i = 1 %to &n_name_vars;

    myparms12&i.
  %end;
  ;
run;

%mend;

%combine;

Data big1;
set big(rename=(probf=RAW_P));
run;
proc multtest pdata=big1 ADAPTIVEHOCHBERG out=big3p;
run;



PROC EXPORT DATA= WORK.big3p
            OUTFILE=
"C:\Users\akaushl1\Documents\IgE_MIXED\Results_mixed_iow1_common_with_tai_june1.csv"

            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;
```

IRB Approval letter

Hello,

The University of Memphis Institutional Review Board, FWA00006815, has reviewed and approved your submission in accordance with all applicable statuses and regulations as well as ethical principles.

PI NAME: Akhilesh Kaushal
CO-PI:

PROJECT TITLE: Comparison of different cell type correction methods for genome-wide epigenetics studies
FACULTY ADVISOR NAME (if applicable): Hongmei Zhang

IRB ID: #4075
APPROVAL DATE: 6/62016
EXPIRATION DATE:
LEVEL OF REVIEW: Exempt

Please Note: Modifications do not extend the expiration of the original approval

Approval of this project is given with the following obligations:

1. If this IRB approval has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.

2. When the project is finished or terminated, a completion form must be completed and sent to the board.

3. No change may be made in the approved protocol without prior board approval, whether the approved protocol was reviewed at the Exempt, Exedited or Full Board level.

4. Exempt approval are considered to have no expiration date and no further review is necessary unless the protocol needs modification.

Approval of this project is given with the following special obligations:

Thank you,

James P. Whelan, Ph.D.

Institutional Review Board Chair

The University of Memphis.

Note: Review outcomes will be communicated to the email address on file. This email should be considered an official communication from the UM IRB.