EPIGENOME-WIDE ASSOCIATION STUDY OF RECOVERY OUTCOMES OF TRAUMATIC BRAIN INJURY PATIENTS

by

Yunqi Li

BS, Cornell University, 2016

BS, China Agricultural University, China, 2016

Submitted to the Graduate Faculty of

the Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Yunqi Li

It was defended on

April 16th, 2018

and approved by

Thesis Director:

Daniel E. Weeks, PhD, Professor, Departments of Human Genetics and Biostatistics
Graduate School of Public Health, University of Pittsburgh

Thesis Co-Director:

John R. Shaffer, PhD, Assistant Professor, Departments of Human Genetics and Oral Biology
Graduate School of Public Health, School of Dental Medicine
Graduate School of Public Health

Yvette P. Conley, PhD, FAAN, Professor
Department of Health Promotion & Development and Human Genetics
School of Nursing, Graduate School of Public Health
University of Pittsburgh

Daniel E. Weeks, PhD

John R. Shaffer, PhD

EPIGENOME-WIDE ASSOCIATION STUDY OF RECOVERY OUTCOMES OF TRAUMATIC BRAIN INJURY PATIENTS

Yunqi Li

University of Pittsburgh, 2018

ABSTRACT

Background: Traumatic Brain Injury (TBI) is a leading cause of morbidity and morbidity among individuals under 45 years old, worldwide. It is unknown why patients with similar extent of injury, similar care, and similar demographic factors have different recovery outcomes. Previous studies using animal models have identified robust DNA methylation changes post-TBI. This project aims to detect CpGs whose methylation levels associate with TBI patients' recovery outcomes in human subjects.

Methods: We obtained DNA methylation profiles of cerebrospinal fluid samples collected at three different time points, first or second day, third or fourth day, and fifth or sixth day post-TBI from 120 severe TBI patients. Measures of recovery were collected including Glasgow Outcome Scale (GOS), Disability Rating Scale (DRS), Neurological Rating Scale (NRS), Anxiety (ANX), Depression (DEP), and Deiner Satisfaction with Life Scale (SWLS) at third month, sixth month, twelfth month, and twenty-fourth month post-TBI, as well as covariates such as age, gender, BMI, and smoking. We dichotomized the third-month GOS to create a binary variable, which is the first phenotype used in the regression model. We also clustered the patients into poor recovery group and good recovery group based on the last available GOS, DRS, NRS, ANX, DEP, SWLS records, which is the second recovery phenotype we analyzed. After quality control and methylation data normalization, we used a linear regression model with empirical Bayes moderation to assess the association between DNA methylation at 307,187

cytosine-phosphate-guanine (CpG) sites and two recovery phenotypes with adjustment for age, gender, and surrogate variables.

Result: No significant associations between CpG methylation and recovery outcomes were observed at the genome-wide threshold for statistical significance (2.4×10^{-7}). 24 CpGs were suggestively associated with TBI recovery at p-value less than 1×10^{-5}. Most of these were located in/near genes which are associated with neurological phenotypes.

Public Health Significance: This pilot project provides a framework for a proposal to collect a larger dataset with higher power to detect potential genes and pathways related to methylation change post-TBI, and has the potential to develop novel interventions or improve the efficacy of existing interventions.

TABLE OF CONTENTS

PREFACE X

1.0 INTRODUCTION 1

1.1 SPECIFIC AIM 1

1.2 TRAUMATIC BRAIN INJURY 1

1.2.1 Traumatic Brain Injury Epidemiology 1

1.2.2 Traumatic Brain Injury Pathogenesis 1

1.2.3 Traumatic Brain Injury Severity and Recovery Outcome Measures 2

1.2.4 DNA Methylation and Traumatic Brain Injury 4

1.3 EPIGENOME-WIDE ASSOCIATION STUDY 4

1.3.1 Study Design 4

1.3.2 Goal of EWAS 6

1.3.3 Major Challenges 6

1.4 PUBLIC HEALTH SIGNIFICANCE 7

2.0 MATERIALS AND METHODS 8

2.1 STUDY POPULATION 8

2.2 DNA METHYLATION PROFILING 8

2.3 METHYLATION DATA QUALITY CONTROL 9

2.4 PATIENT CLUSTERING 9

2.5 ASSOCIATION ANALYSES 10

3.0 RESULTS 13

3.1 STUDY SAMPLE CHARACTERISTICS 13

3.2 PATIENT CLUSTERING 14

3.3 QUALITY CHECK OF THE DNA METHYLATION ARRAY DATA 15

3.4 SURROGATE VARIABLE ANALYSIS 16

3.5 EWAS OF RECOVERY OUTCOMES AND GROUPS 16

4.0 DISCUSSION 24

APPENDIX A : LIST OF CANDIDATE GENES IN OXPHOS PATHWAY 26

APPENDIX B : COMET, SCATTER AND BOX PLOTS OF SUGGESTIVE SIGNIFICANT CPGS 27

BIBLIOGRAPHY 52

LIST OF TABLES

Table 1. Measure, Range and Categories of TBI Severity and Outcome. 3

Table 2. Sample Characteristics. 13

Table 3. Number of Probes removed by the quality control steps 16

Table 4. Results for CpG sites association with outcomes. 18

Table 5. Ranks of 24 suggestively significant CpG across all 6 association analyses. 19

Table 6. Neurological Phenotypes related to CpGs or UCSC Reference Genes 22

LIST OF FIGURES

Figure 1. Flowchart of the study design. 12

Figure 2. Means of recovery scores of each group, and heatmap of each patient's scaled recovery scores. 15

Figure 3. Manhattan Plot 20

Figure 4. Quantile-Quantile Plot. 21

PREFACE

I would like to express the deepest appreciation to my family, fellow students, academic and research advisors, course instructors here at Graduate School of Public Health, University of Pittsburgh. Without the help and support, it would not have been possible to write this master thesis.

Above all, I would like to thank my committee members, Dr. Daniel E. Weeks, Dr. John R. Shaffer, and Dr. Yvette P. Conley, who gave me the opportunity to begin my first independent research project and demonstrated to me how to become a professional scientific researcher.

I would also like to acknowledge our group member, PhD student Annie I. Arockiaraj, who mentored me on data analytic and communication skills. Also, a thanks to all the class instructors and fellow students, who evoked my thinking about my future and career.

Lastly, thanks to my parents for their kind financial support to help me complete this two-year study and Hanxin Zhang, for his spiritual support, and their great patience and understanding at all times.

1.0 INTRODUCTION

1.1 SPECIFIC AIM

This project aimed to detect CpGs whose methylation level is different between good recovery TBI patients and poor recovery ones.

1.2 TRAUMATIC BRAIN INJURY

1.2.1 Traumatic Brain Injury Epidemiology

Traumatic brain injury (TBI) is the leading cause of morbidity and mortality among individuals under 45 years old, worldwide[1]. In the United States, approximate 5.3 million people suffer from the sequelae of TBI. Children, adolescents, and adults aged 75 years and older, are among those with the highest risk of having TBI-related emergency department visits and hospitalizations. The leading causes of TBI are falls, traffic accidents, and firearms.

1.2.2 Traumatic Brain Injury Pathogenesis

TBI has two major injury mechanisms, focal brain damage and diffuse brain damage[2]. Focal brain damage refers to contact injury, which causes contusion, laceration, and intracranial hemorrhage. Diffuse brain damage refers to acceleration/deceleration injury, which causes axonal injury and brain swelling. After TBI occurs, brain damage develops in two stages, the primary insult and the secondary insult. The primary insult, also known as primary damage or mechanical damage, arises at the same time as the impact occurs. The secondary insult, also known as secondary damage or delayed non-mechanical damage, involves all successive pathological changes that occur after injury such as cerebral ischemia and

intracranial hypertension. Specific pathophysiology processes occurring after TBI include decreasing/increasing cerebral blood flow, impairment of cerebrovascular autoregulation and carbon-dioxide reactivity, cerebral vasospasm (narrowing of the large and medium-sized intracranial arteries, which affects the anterior circulation supplied by the internal carotid arteries), cerebral metabolic dysfunction, brain tissue hypoxia, edema, inflammation, and cell necrosis and apoptosis.

1.2.3 Traumatic Brain Injury Severity and Recovery Outcome Measures

In our study, we used one injury severity measure and six recovery outcome measures to describe the patient's status. The Glasgow Coma Score (GCS) is one of the most commonly used assessments in quantifying the severity of TBI. It directly reflects the conscious state of a person by grading on eye opening, verbal response and motor response[3]. The Glasgow Outcome Scale (GOS) is the most referenced recovery measurement of brain injury due to its simplicity, time-saving, and reliability. GOS is scored on a 1 to 5 scale indicating different recovery status[4]. The Neurobehavioral Rating Scale (NRS) measures behavior and cognitive disturbance. It is measured based on a 10 to 20 minutes interview and patient observation including 27 aspects such as mental flexibility, anxiety, alertness attention, motivation, suspiciousness, and mental fatigability etc.[5] The Disability Rating Scale (DRS) assesses the patient's disability from eight aspects which are composed of eye opening, communication ability, motor response, cognitive ability to feed, toilet, and groom, functioning level, and employability[6]. Anxiety (ANX) and Depression (DEP) was measured using The Brief Symptom Inventory 18[7]. The Satisfaction with Life Scale (SWLS) assess the global cognitive judgment of life satisfaction. The score was calculated based on the agreement with five statements: "In most ways my life is close to my ideal. My life conditions are excellent. I am satisfied with my life. So far, I have gotten the important things I wanted in life. If I could live my life over, I would change almost nothing." [8] Ranges and categories of GCS, GOS, NRS, DRS, and SWLS are summarized in Table 1.

Table 1. Measure, Range and Categories of TBI Severity and Outcome.

Measure	Range	Categories
Glasgow Coma Scale (GCS)	3 to 15	3 - 8 Severe Brain Injury
		9 - 12 Moderate Brain Injury
		13-15 Mild Brain Injury
Glasgow Outcome Scale (GOS)	1 to 5	1 Death
		2 Persistent Vegetative State
		3 Severe Disability
		4 Moderate Disability
		5 Mild or no Disability
Neurobehavioral Rating Scale (NRS)	0 to 116	Higher score indicates more severe neurobehavioral impairment of behavioral disturbances
Disability Rating Scale (DRS)	0 to 29	0 None Disability
		1 Mild Disability
		2-3 Partial Disability
		4-6 Moderate Disability
		7-11 Moderately Severe Disability
		12-16 Severe Disability
		17-21 Extremely Severe Disability
		22-24 Vegetative State
		25-29 Extreme Vegetative State
Deiner Satisfaction with Life Scale (SWLS)	0 to 35	5-9 Extremely Dissatisfied
		10-14 Dissatisfied
		15-19 Slightly Dissatisfied
		20 Neutral
		21-25 Slightly Dissatisfied
		21-25 Slightly Satisfied
		26-30 Satisfied
		31-35 Extremely Satisfied

1.2.4 DNA Methylation and Traumatic Brain Injury

DNA methylation, as one of the most commonly studied epigenetic changes, has been assessed in a number of animal model studies of TBI. For example, in 2007, Zhang et al. detected global hypomethylation in microglia/macrophages in the early process after TBI[9]. In 2009, Lundberg et al. reported that they observed DNA-methyltransferase (Dnmts) enzyme re-localization as part of an epigenetic reprogramming of in situ reactive astrocytes post-TBI[10]. In 2015, Haghighi et al. identified cell-specific DNA methylation perturbations in neurons and glia associated with blast exposure in rats model. Therefore, it would be interesting to see how methylation relates to TBI in human patients[11].

1.3 EPIGENOME-WIDE ASSOCIATION STUDY

1.3.1 Study Design

In 2013, Michels et al. proposed a step-by-step framework for a successful Epigenome-wide Association Study (EWAS) study design[12].

A clear and concrete hypothesis or research question is the first step of a valid EWAS. Researchers need to hypothesize how the epigenetics variance is associated with a concrete phenotype. In this step, researchers are recommended to draw relationship graphs between epigenetics changes, environmental exposures, and outcomes.

Secondly, tissue for sampling needs to be specified based on existing hypothesis. For example, if the study aims to detect the epigenetics difference between tumor and non-tumor, the control tissue samples need to be retrieved from corresponding healthy control subjects instead of from the adjacent tissue of the tumor from case subjects. The reason is the epigenetic profile may change all over the

cancerous organ even in a histologically normal location. If the study aims to detect the association between epigenetic changes and Body Mass Index (BMI), blood might be selected as the sample tissue. Because BMI is a whole-body index and blood circulates around the whole body, blood might be an indicator of global change comparing to other solid tissues. In our study, we use cerebrospinal fluid (CSF) as the sampling tissue. CSF is a clear, colorless liquid surrounding and protecting the central nervous system (CNS). Advantages of choosing CSF are it is located near the brain, where the injury occurs, and it has been known to play a role in CNS repairing after injury[13].

Thirdly, researchers need to be aware of the population structure and biological variability when selecting samples. Population structure refers to genome/epigenome differences between the subgroups within the study samples. Confounding might be introduced to the association analysis if those characteristic differences are associated with the disease trait of interest. Therefore, researchers are recommended to select a homogeneous population towards the suspect characteristic or to enlarge the sample size.

After subject recruitment, the fourth step is choosing a suitable platform for methylation measure. The Illumina Infinium HumanMethylation Beadchip array, Reduced Representation Bisulfite Sequencing (RRBS), MeDIP with high-through sequencing (MeDIP-seq), Methyl-CpG binding domain protein sequencing (MBD-seq) are some commonly used protocols. Batch effects, a technical source of variation, may be introduced here. Possible solutions to reduce batch effects include: equal distribution of cases and controls within chips, balanced sample processing time, and careful application of quality control criteria. For data analysis, site-by-site association analysis, regional changes analysis, preclustering or grouping CpG sites analysis, and functional and gene set enrichment analysis could be employed. Lastly, to verify and validate the results, single-locus specific methylation techniques could be applied to verify the methylation measures. Comparable and different samples measured by a different methylation platform could also be used as the replication.

1.3.2 Goal of EWAS

Similar to genome-wide association study (GWAS), results from an EWAS could lead to the discovery of new biological mechanisms and target potential biomarkers for later drug development or disease intervention and prediction.

1.3.3 Major Challenges

In 2016, Birney et al. pointed out challenges in interpreting EWAS result, including cell type heterogeneity, transcriptomic variability, spurious associations, and reliable but non-causal association[14].

Cell type heterogeneity comes from the observations that different types of cells in the same sample tissue have distinct epigenetic profiles. For example, DNA methylation patterns may differ between natural killer cells and lymphocytes. Given that the current main methylation protocol measures the overall methylation level in the sample tissue, adjustment of the proportion of cells in the tissue will generate a different methylation level report. Researchers are encouraged to treat this problem seriously. Some possible solutions include: adding cell proportions as covariates in the association model, using single-cell techniques to measure epigenetics data, or applying statistical algorithms such as sparse PCA or surrogate variable analysis.

In an EWAS study, the identified target site can only be viewed to associate with an outcome instead of causing the outcome because the epigenetic profile could be altered by later environment and lifestyle factors. Therefore, when a significant association is detected between epigenetic change and disease phenotype, there are three possibilities: epigenetic change causes disease phenotype, disease phenotype causes epigenetic change (reverse causality), or there is no causal relationship between two, i.e. the observed association is induced by an undetected confounder.

1.4 PUBLIC HEALTH SIGNIFICANCE

TBI has been considered as a serious public health problem in the United States due to its incidence and chronic health effects[15]. It is noticed that there exists heterogeneity of prognosis of TBI patients, and traditional “one size fits all” intervention may not succeed clinically.

Previous animal studies have indicated robust methylation changes post-TBI. It would be interesting to explore the methylation changes in human subjects. The epigenome-wide association study approach is an efficient method for scanning the genome and identifying potential biological targets involved in TBI prognosis. Therefore, using an EWAS to detect the methylation differences between different recovery outcomes post-TBI could be a great opportunity to identify CpGs or genes related to TBI prognosis. This pilot study may also provide an analytic framework for the future collection of a larger data set that would have higher power to detect potential genes or pathways related to methylation change post-TBI and help to develop novel interventions or improve the efficacy of existing ones.

2.0 MATERIALS AND METHODS

2.1 STUDY POPULATION

A total of 120 severe traumatic brain injury patients from the Pittsburgh area were recruited. For each patient, GCS was obtained, and phenotypic measures of recovery, including GOS, DRS, NRS, ANX, DEP, and SWLS, were collected at three, six, twelve, and twenty-four months after TBI. Other demographic data including age, gender, ethnicity, weight, height, smoking status, education status, and marital status were also collected. Each subject's CSF samples were obtained at three different time points: first or second day (day 1 or 2), third or fourth day (day 3 or 4), and fifth or sixth day post-TBI (day 5 or 6).

2.2 DNA METHYLATION PROFILING

For all CSF samples (n=368, 3 samples for each of 120 patients and 8 technical replicates), the DNA methylation levels of CpGs across genome were measured using the Infinium Human Methylation 450k BeadChip, Illumina Inc. This DNA methylation array covers about 450,000 CpG sites. The DNA methylation level of each CpG was quantified by the beta value, which is a continuous variable ranging from 0, fully unmethylated, to 1, fully methylated. Methylation M values were calculated using Formula 1[16] based on β values and offset, α (we use α =100 in this study).

$$M = \log_2 \frac{\beta}{1 - \beta + \alpha} \quad (1)$$

2.3 METHYLATION DATA QUALITY CONTROL

Quality control procedures were performed for filtering and normalization of the methylation β values and M values using ENmix[17], Minfi[18], and CpGFilter[19] packages. First, we used ENmix and Minfi packages to remove low-quality samples, probes and outliers. Samples with bisulphite intensity less than 3 standard deviations below the mean across all samples or greater than 1% of CpGs with low quality values were considered low quality samples, and removed from analysis. Sample outliers identified based on total intensity or beta value distribution were also removed. Second, we use ENmix to perform background and dye bias correction. Third, we performed functional normalization[20] to control for potential batch effects. Lastly, we removed the CpGs which (1) overlapped with known single nucleotide polymorphisms (SNP), (2) were assayed by cross-reactive probes, (3) were located on the sex chromosomes, (4) exhibited multi-model distributions, (5) were low quality probes, defined as having >5% low quality methylation measurements across all samples, identified in Enmix, and (6) had a high ratio of technical variation to biological variations defined as intra-class correlation (ICC) score less than 0.55 across 8 technical replicate samples.

2.4 PATIENT CLUSTERING

Patients were classified into “good recovery” and “poor recovery” groups based on up to six outcomes measures using the k-POD method. GOS, NRS, DRS, SWLS, ANX, and DEP were measured at three, six, twelve, and twenty-four months after TBI. However, due to missing measurements and unavailability of SWLS, ANX, DEP scores in dead patients, we used the last available measure of GOS, NRS, DRS, SWLS, ANX, DEP to represent the final recovery status for each patient. Therefore, each patient then had up to 6 measurements used for clustering. Dead and persistent vegetative patients were first classified into the “poor recovery” group and eliminated from the following clustering procedure. Remaining patients were then clustered into two groups using the k-POD method from the kpodcluster package[21] . k-POD is

an extension of k-means clustering but allows for missing data and unknown missingness mechanisms. Hereafter, we referred the patients' clustering result from this section as Cluster-Based Recovery Group (CBRG).

2.5 ASSOCIATION ANALYSES

A total of six association analyses, including methylation in CSF at three-time points for two recovery outcomes, third-month GOS and CBRG, were conducted. Study-specific modeling details can be found in the Figure 1. A linear regression model with empirical Bayes moderation was used to estimate the association between DNA methylation M values for 307,187 CpGs and dichotomized 3-month GOS (GOS scores from 1 to 3 were recoded as 1, poor recovery, scores from 4 to 5 were recoded as 0, good recovery), denote as "GOS 3" hereafter, or CBRG, while adjusting for age, sex, and surrogate variables. The eBayes function from the limma package was used for all regression analyses[22]. Surrogate variables were computed using the sva package[23] in each model to remove potential batch effects and other unwanted variation, such as cell type heterogeneity, that was not adequately controlled by quality control steps. BMI and smoking were not included in the model as covariates due to the large amount of missing data. Race was not included because 95% percent of the patients are Caucasians. GCS was not included because patients in this study all had GCS scores between 3 to 8, which were classified in to severe brain injury. The threshold used to determine statistical significance was 2.4×10^{-7}, and the threshold to determine suggestive statistical significance was 1.0×10^{-5}.

Identified CpGs were annotated using the IlluminaHumanMethylation450k.db package[24]. For CpGs, which had p-value less than suggestive statistical significance threshold, coMET plots were drawn to visualize the regional association results and DNA co-methylation patterns using the coMET package[25] [26]. In addition, scatter plots and box plots were constructed to visualize the difference in methylation magnitude between poor recovery group and good recovery group. For all EWAS scans, Quantile-Quantile (QQ) plots were generated to check for confounder effects. EWAS results were summarized in

Manhattan plots. We also specifically looked at the results in candidate genes involving in the oxidative phosphorylation (OXPHOS) pathway, which has been shown to related to patient outcomes post-TBI[27] (Appendix A shows the OXPHOS candidate gene list).

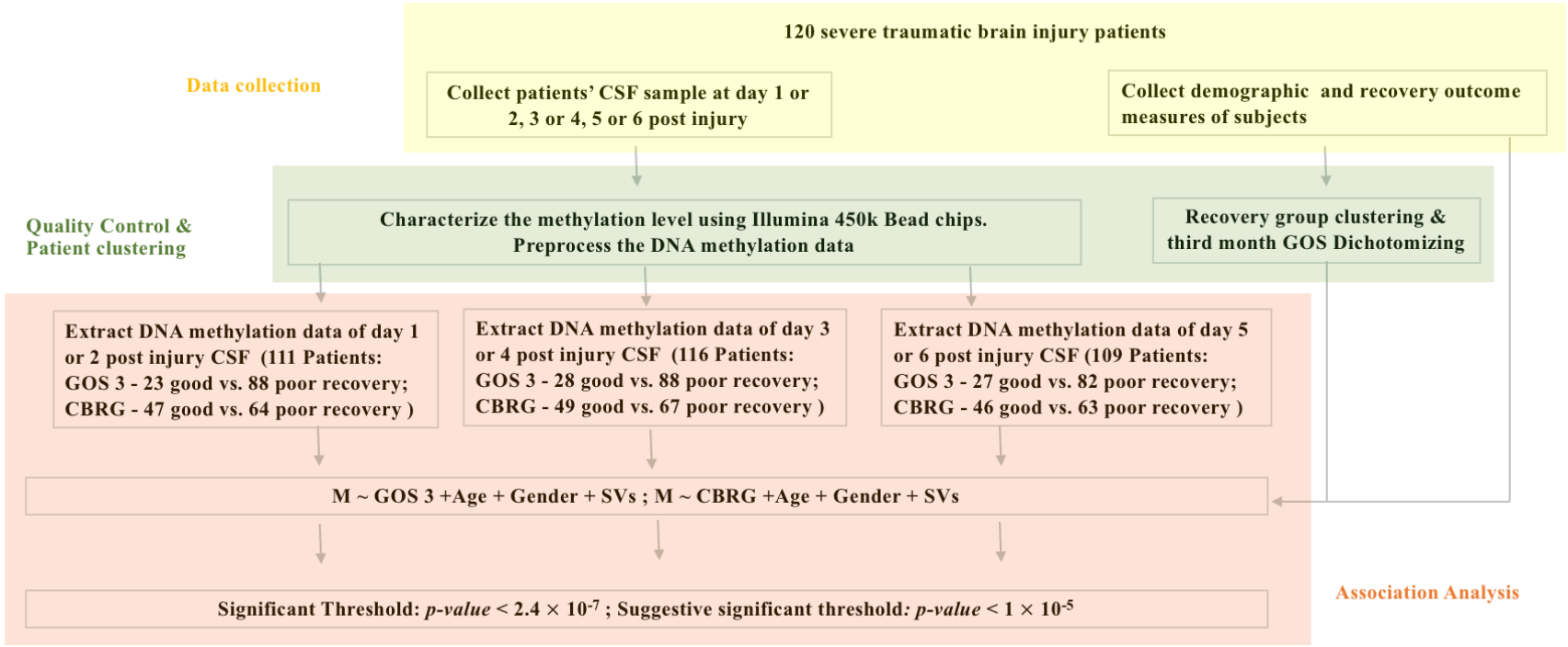


Figure 1. Flowchart of the study design.

Epigenome-wide association study was performed in order to identify methylation sites associated with TBI recovery. A total of 120 severe TBI patients were recruited. CSF sample was collected at three different time points. Demographic data including age, gender were collected. Recovery outcome measures including GOS, NRS, DRS, DEP, ANX, SWLS were also collected at three, six, twelve, and twenty-four months after TBI. Methylation levels were characterized using Illumina 450k Beadchip. Due to data availability, we used dichotomized third month GOS (GOS 3) as the first phenotype used in the regression model, and cluster-based recovery group (CBRG) as the second phenotype for analysis. Epigenome-association was performed with methylation M value as the response variable and the two phenotypes as independent variables, adjusting for age, gender and calculated surrogate variables. The genome-wide significance level is 2.4×10^{-7}, and suggestive significance level is 1×10^{-5}.

3.0 RESULTS

3.1 STUDY SAMPLE CHARACTERISTICS

The characteristics of 120 traumatic brain injury patients are summarized in Table 2. The age of patients ranged from 16 years old to 74 years old with mean at 37.24. Females account for 20.83% of the total of 120 patients. Most of the patients (95.83%) were Caucasian, 4 out of 120 patients (3.33%) were African Americans, and 1 patient (0.83%) was Asian. 51.67% of the patients recorded had severe disability at third month GOS.

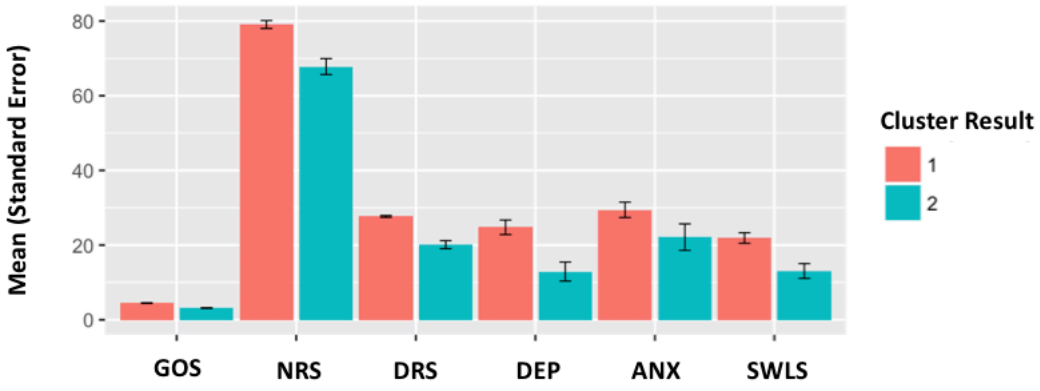
Table 2. Sample Characteristics.

Demographic Characteristics	Mean (SD) or Percent
Age, years	37.24 (16.87)
Sex, % female	20.83%
Race, %	
African Americans	3.33%
Caucasian	95.83%
Asian	0.83%
3rd month GOS, %	
1 Death	20.00%
2 Persistent vegetative state	4.17%
3 Severe disability	51.67%
4 Moderate disability	20.00%
5 Low disability	4.17%

3.2 PATIENT CLUSTERING

We used k-POD methods to dichotomize non-dead and non-persistent vegetative subjects into “good recovery” and “poor recovery” outcomes groups based on each patient’s last available GOS, NRS, DRS, DEP, ANX, and SWLS records. Means of GOS, reversely-scored NRS, reversely-scored DRS, reversely-scored DEP, reversely-scored ANX, and SWLS of each recovery group, and a heatmap of each patient’s scaled measure value are presented in Figure 2. We reversed scored DRS, NRS, DEP, and ANX (calculated as the maximum possible score minus the measured score) so that a lower value always indicates poor recovery and a higher value indicates good recovery. After combining dead and persistent vegetative subjects with the poor recovery cluster, there were a total of 67 subjects determined to have a poor recovery outcome and 53 subjects with a good recovery outcome.

(A)



(B)

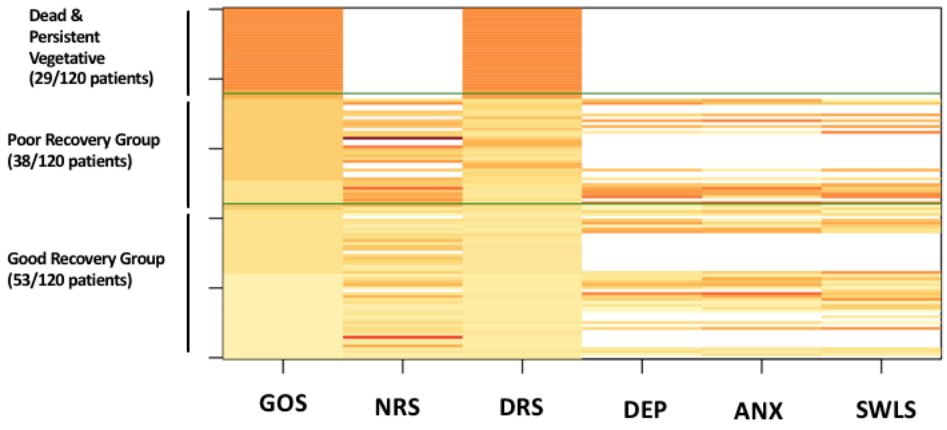


Figure 2. Means of recovery scores of each group, and heatmap of each patient's scaled recovery scores.

(A) Recovery outcome measures based on latest record in good and poor CBRG. Bar height stands for the mean value of each measure and the whisker stands for the standard error of mean. DRS, NRS, DEP, and ANX were reversely scored for visualizing these results. Cluster 2 has lower values in GOS, NRS, DRS, DEP, ANX, and SWLS, indicating a relatively poor recovery.

(B) Heatmap of subjects' recovery scores ordered by k-POD cluster based on latest record. The orange, red and darker colors indicate values representing poor recovery, yellow or lighter colors indicate values representing good recovery. The green lines indicated the boundary between dead patients and others, and between poor recovery group (cluster 2 in Figure 2A) and good recovery group (cluster 1 in Figure 2A).

3.3 QUALITY CHECK OF THE DNA METHYLATION ARRAY DATA

We performed quality control steps for data filtering and normalization. Out of the 368 samples, 8 samples were identified as low-quality samples, and 9 as outlier samples. These 17 were removed from later analysis. 485,512 CpGs were interrogated with the Illumina HumanMethylation450 BeadChip. 307,187 CpGs remained after step-by-step filtering of probes with SNPs, possible cross-reactive probes, sex-chromosome probes, multi-modal probes, bad probes identified by ENmix Package, and ICC lower than 0.55 (Table 3).

Table 3. Number of Probes removed by the quality control steps

Step	Number of Probes Retained	Removed
Raw	485512	-
Probes with SNPs	467971	17541
Cross-reactive Probes	431482	36489
Remove sex chromosome	421291	10191
Remove multi-model Probes	421291	0
Remove Enmix bad Probes	419895	1396
CpG Filter	307187	112708

3.4 SURROGATE VARIABLE ANALYSIS

We applied Surrogate Variable Analysis (SVA) to CSF methylation data collected at the day 1 or 2, day 3 or 4, day 5 or 6 post-TBI to calculate surrogate variables. The resulting surrogate variables were used in EWAS regression model to adjust for potential batch effects and other unknown sources of variation. Eight surrogate variables were identified using CSF methylation data collected at the day 1 or 2 post injury. Eight and seven surrogate variables were identified using CSF data collected at the day 3 or 4, and at the day 5 or 6 post injury.

3.5 EWAS OF RECOVERY OUTCOMES AND GROUPS

We investigated the association of GOS 3 and CBRG with DNA methylation at 307,187 CpGs assayed by the Illumina HumanMethylation 450 BeadChip in CSF. Due to the sample filtering in the quality control step and data availability, there were total 111 samples, 116 samples, and 109 samples for different sets of days post injury available for association analysis, as shown in Figure 1. The regression model used

methylation M-value as the dependent variable, used GOS 3 and CBRG as the independent variable, and was adjusted for age, gender, and surrogate variables.

Manhattan plots of association p-values are shown in Figures 3. From the Manhattan plot, there were 24 signals above the 1×10^{-5} suggestive significance threshold while there are no signals above the genome-wide significance threshold of 2.4×10^{-7}. Quantile-Quantile (QQ) plots of association p-values are shown in Figures 4. In the QQ plot of the day 1 or 2 dataset with CBRG as outcomes, deviation from the expected *p*-value distribution started from the middle part and become evident in the tail area, suggesting there might be remaining batch effects or other unknown confounding effects.

In our study, no significant associations between CpG methylation and recovery outcomes were observed at the genome-wide threshold for statistical significance, while 24 CpGs were identified to be suggestive association sites (Table 4). The Benjamini-Hochberg adjusted p-values were all greater than 0.05 indicating no genome-wide significant results. The ranking of 24 suggestively significant CpGs across all six association analyses was also extracted (Table 5). The rank of each CpG varied a lot across samples collected at different time-points. The rank of most CpGs also varied across different outcomes, while cg14507042, cg13505794, cg03899054, cg24845595 showed a consistency in rank in GOS 3 and CBRG.

Table 4. Results for CpG sites association with outcomes.

Data, Trait	CpG ID	CHR	Position	*p*-value	Adjust *p*-value	UCSC Reference Gene	Location with Respect to Gene	B	t
Day 1 2, GOS 3	cg08527161	chr11	126294215	3.89E-06	0.994	KIRREL3	3'UTR	2.245	-4.878
Day 3 4, GOS 3	cg15210596	chr1	246887325	6.54E-06	0.581	SCCPDH	TSS200	1.827	-4.739
	ch.2.3132178R	chr2	152476400	8.27E-06	0.581	NEB	Body	1.676	4.681
	cg21641458	chr7	27185136	5.05E-06	0.581	HOXA6	3'UTR	1.993	-4.802
	cg05756622	chr20	61147572	1.37E-06	0.420	C20orf166; C20orf200	TSS200;5'UTR	2.829	-5.113
Day 5 6, GOS 3	cg13505794	chr5	110075158	9.84E-06	0.427	SLC25A46	Body	1.365	-4.652
	cg14507042	chr9	139781209	8.98E-06	0.427	TRAF2	5'UTR	1.422	-4.675
	cg17167468	chr10	113943149	4.33E-06	0.427	GPAM	5'UTR	1.872	-4.854
	cg26276120	chr12	6977747	4.28E-06	0.427	TPI1	Body	1.879	4.857
	cg23401756	chr16	68013981	8.09E-06	0.427	DPEP3	Body	1.486	4.700
	cg12139369	chr17	71284979	4.49E-06	0.427	CDC42EP4	5'UTR	1.849	-4.845
Day 1 2, CBRG	cg24845595	chr1	156814488	3.59E-06	0.248	NTRK1; INSRR	Body; Body	2.826	4.897
	cg26812481	chr1	25976496	6.45E-06	0.248	MAN1C1	Body	2.411	-4.754
	cg25586848	chr3	194875931	5.66E-06	0.248	C3orf21	Body	2.504	-4.786
	cg15807035	chr4	140374443	4.55E-06	0.248	RAB33B	TSS1500	2.659	-4.839
	cg10336790	chr7	112725899	2.69E-06	0.248	GPR85	5'UTR	3.031	4.967
	cg03899054	chr8	143206674	1.20E-06	0.248			3.600	5.158
	cg23356310	chr12	130898775	4.84E-06	0.248	RIMBP2	Body	2.615	4.824
	cg26774156	chr15	74495384	9.50E-06	0.321	STRA6; STRA6; STRA6	5'UTR;1stExon; TSS200	2.136	4.658

Table 5. Ranks of 24 suggestively significant CpG across all 6 association analyses.

CpG Name	Day 1 2, GOS 3	Day 3 4, GOS 3	Day 5 6, GOS 3	Day 1 2, CBRG	Day 3 4, CBRG	Day 5 6, CBRG
cg08527161	1	63869	9472	2368	276647	108451
cg05756622	78565	1	115688	80735	9557	218993
cg21641458	397	2	7021	7248	655	16778
cg15210596	139325	3	1401	110500	6514	26267
ch.2.3132178R	141436	4	159393	241763	5609	70457
cg26276120	51983	10399	1	138285	150957	311
cg17167468	32773	38770	2	15300	177680	687
cg12139369	87661	177492	3	252320	283430	82
cg23401756	19715	107126	4	183824	202450	11633
cg14507042	276721	73070	5	255249	35931	6
cg13505794	110008	254226	6	71705	230191	38
cg03899054	5	189654	4409	1	71773	1727
cg27452922	372	67325	265880	2	228857	131890
cg10336790	11214	73599	55823	3	132	5609
cg24845595	3	788	32	4	1735	922
cg15807035	14355	162193	9619	5	111942	797
cg23356310	805	17426	74838	6	5740	9301
cg25586848	81675	78091	296749	7	280848	163835
cg26812481	1195	157772	49491	8	174336	287544
cg26774156	412	1301	272137	9	408	165344
cg19003904	135990	509	293544	245251	1	159612
cg15885734	36869	994	190827	22452	2	41485
cg15697902	122961	7757	47758	4648	3	1472
cg05794310	16778	273622	5958	3376	137175	1

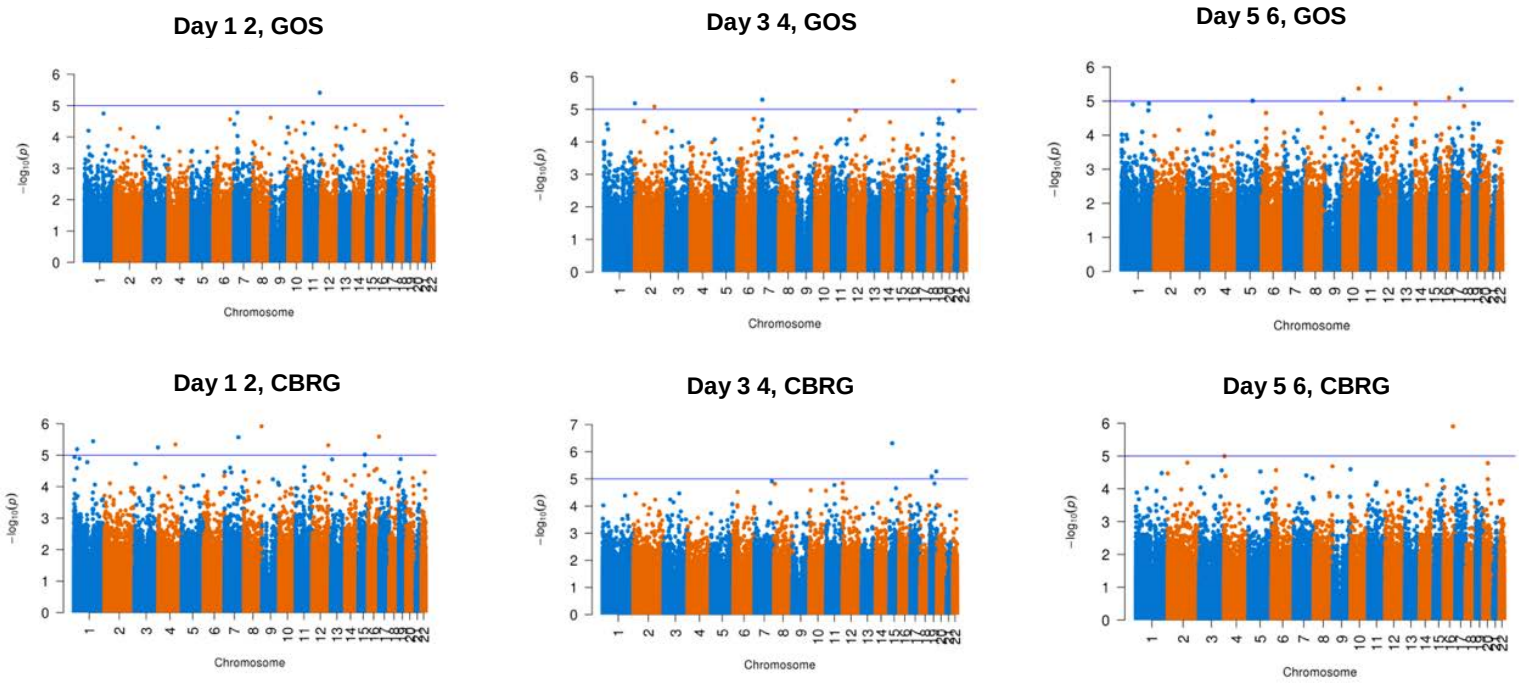


Figure 3. Manhattan Plot.

Each of the point represents 1 of the 307181 CpGs, colored according to chromosome. The x axis represents genomic location, and the y axis represents the negative logarithm of the p-value for CpG association calculated using a linear model. Threshold (blue line) is drawn at p-value < 1×10^{-5} for suggestive association.

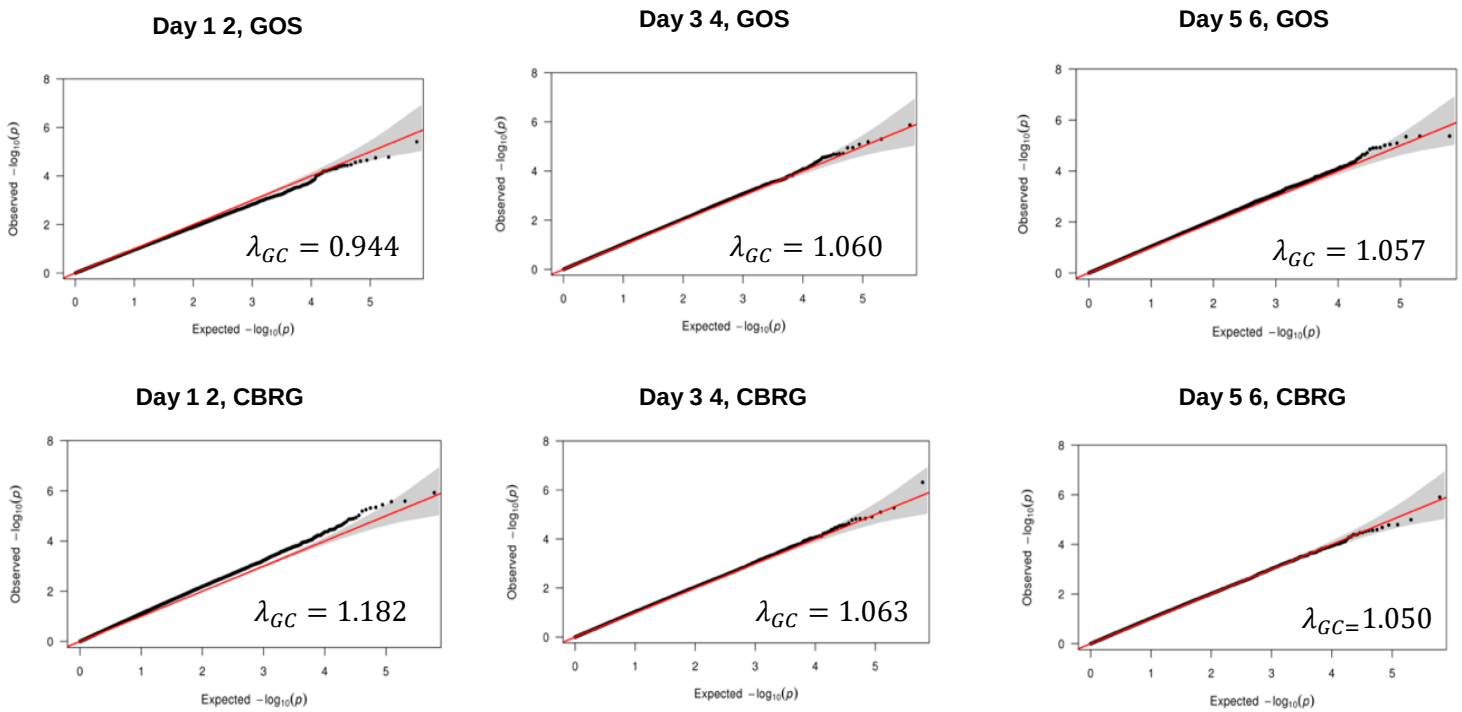


Figure 4. Quantile-Quantile Plot.

The negative logarithms of the observed (y axis) and the expected (x axis) p-value are plotted for each CpG. The red line indicates the null hypothesis of no true association. The grey shaded area indicates the 95% confidence interval. λ_{GC} stands for genomic inflation factor.

Among the 24 CpGs, many of them were located in or near genes associated with neurological phenotypes (Table 6). Regarding the magnitude of methylation level changes, no CpG had a β value change greater than 0.1 between the good and poor recovery groups. Some CpGs which had relatively large β value changes were: cg08527161, cg21641458, cg05756622, cg23401756, cg12139369, cg24845595, cg10336790, cg03899054 (Appendix 2).

Table 6. Neurological Phenotypes related to CpGs or UCSC Reference Genes.

CpG ID	UCSC Reference Gene	Gene Related Neurological Phenotype
cg08527161	KIRREL3	Intellectual disability and synapse development[28],
cg15210596	SCCPDH	
ch.2.3132178R	NEB	Actin filament stabilizer or length regulator in neurons of the human brain[29]
cg21641458	HOXA6	Biological pathway of neurogenesis[30], meningioma[31]
cg05756622	C20orf166; C20orf200	
cg13505794	SLC25A46	Neuropathy[32]
cg14507042	TRAF2	Cell survival and apoptosis[33]
cg17167468	GPAM	
cg26276120	TPI1	Hemolytic anemia which coupled with progressive, severe neurological disorder[34]
cg23401756	DPEP3	
cg12139369	CDC42EP4	Glutamatergic tripartite synapse configuration[35]
cg24845595	NTRK1; INSRR	Response to influenza vaccination (CpG related) *[36], Congenital insensitivity to pain, self-mutilating behavior, cognitive disability(NTRK1)[37]
cg26812481	MAN1C1	Parkinson Disease[38]
cg25586848	C3orf21	
cg15807035	RAB33B	
cg10336790	GPR85	Osteoarthritis (CpG related) *[39], Schizophrenia, Autism spectrum disorder[40]
cg03899054		
cg23356310	RIMBP2	Synaptic transmission, Spinocerebellar Ataxia[41]

Table 6 Continued

cg26774156	STRA6	
cg27452922		
cg19003904	lincRNA RP11−718O11.1	Waist circumference and weight*[42]
cg15885734	CTU1	
cg15697902	TECR	Mental Retardation[43]
cg05794310	VAC14	Childhood onset progressive neurological disorder [44]

*Not neurological phenotype or Phenotypes related to CpG

4.0 DISCUSSION

In our study, 24 CpGs were identified as suggestive sites which have p-value less than 1×10^{-5}. This 1×10^{-5} threshold is actually an arbitrary threshold commonly used in genome-wide association studies. We used a 1×10^{-5} threshold here is because our study has a relatively small sample size, only about 110, thus may have low power. Some true CpG signals may thus show a low p-value. Therefore, we chose 1×10^{-5} as the threshold and proposed these 24 CpG for further study in large samples.

The significant threshold we used in our study is 2.4×10^{-7}, which was suggested from the paper Estimation of a significance threshold for epigenome-wide association studies[45]. Because of the quality control and CpG filtering step, the actual number of CpG we test is 307,187 instead of about 450,000. The Bonferroni correction may suggest a more conservative threshold of 1.63×10^{-7} (0.05/307,187) which could also be used in this study.

In this study, we also explored the DNA methylation profile of genes involved in the OXPHOS pathway, which were previously reported to be involved in post-TBI recovery[46]. By extracting the association analysis results for those candidate genes, we did not find any significant differences in DNA methylation between good and poor recovery group or between good third-month GOS and poor third-month GOS outcomes.

Our study contains several limitations including the small sample size, lack of replication, use of CSF as the only tissue for measuring methylation, and exclusion of potential confounders from the regression model, and missing data. Firstly, the current study only contains 120 subjects which may result in a low power study. Secondly, there is no additional replication data set to confirm our findings. Thirdly, given that disease-associated methylation changes could be occurred across tissues[47], it could be possible that biological process, for example, immune response, in blood associated with TBI recovery.

Therefore, blood samples could be an addition to the CSF sample. Fourthly, variables such as type of injury, BMI, smoking, medication history, ethnicity, and nearby SNP genotypes were not considered in the regression model, yet such variables may influence DNA methylation. Including these potential sources of variation in the model could eliminate their effects on DNA methylation. One of the reasons for the exclusion of potential covariates is the small sample size limiting the number of predictors that can be simultaneously included in the regression model. Last but not least, several patients were lost to follow-up during the study, and therefore, their outcome measures in later time points, such as at twelfth months and twenty-fourth months, are missing. The traits used in the association study may misrepresent the true recovery status thus introduce noise to the result.

In conclusion, there are no significant associations between CpG methylation and recovery outcomes were observed at the genome-wide threshold for statistical significance. However, 24 CpGs were suggestively associated with TBI recovery. Some discovered CpGs are located in/near genes that are associated with neurological phenotypes.

For future research, large sample replication to confirm the findings, other methylation characterization methods to capture more CpGs, downstream pathway analysis, enrichment analysis to identify TBI related biological pathways or organs, and differentially methylated region analysis are needed to explore the underlying biological mechanisms within TBI recovery.

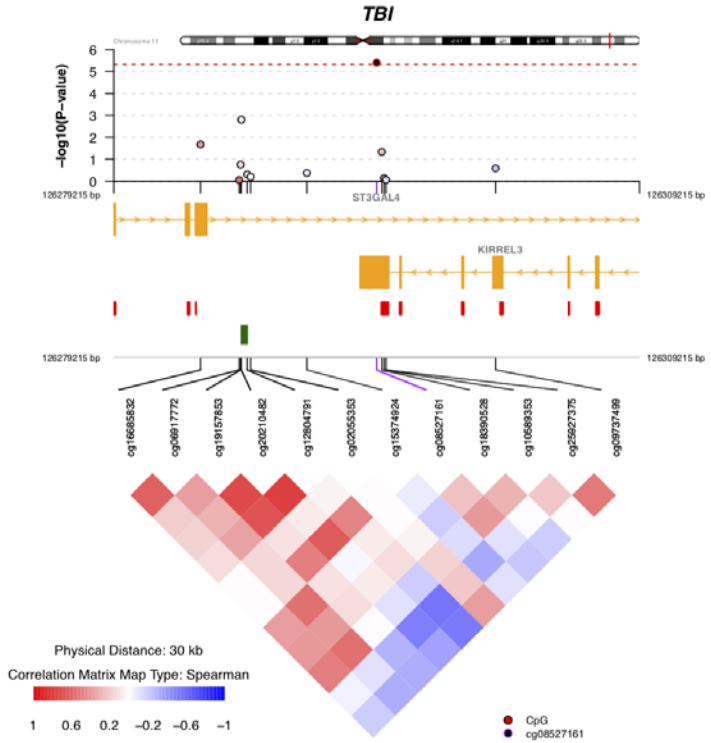
APPENDIX A: LIST OF CANDIDATE GENES IN OXPHOS PATHWAY

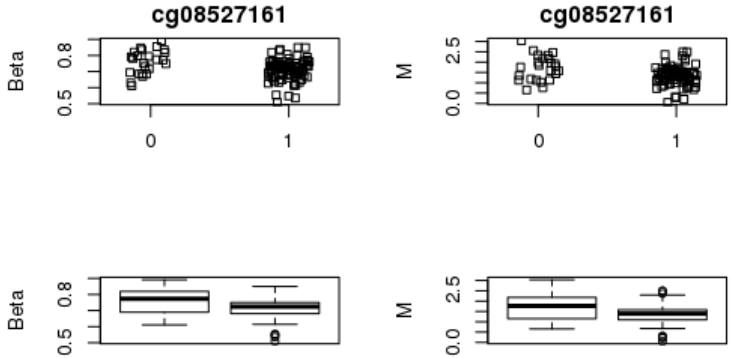
“NDUFS6”, “NDUFA5”, “NDUFS5”, “NDUFS4”, “NDUFA8”, “NDUFS7”, “NDUFS8”, “NDUFV2”, “NDUFS3”, “NDUFA9”, “NDUFA10”, “NDUFS2”, “NDUFV1”, “NDUFS1”, “NDUFA2”, “NDUFB3”, “NDUFA6”, “NDUFA7”, “NDUFC2”, “NDUFA11”, “NDUFB4”, “NDUFA13”, “GRIM19”, “NDUFB6”, “NDUFA12”, “DAP13”, “NDUFB7”, “NDUFB9”, “NDUFAB1”, “NDUFB8”, “NDUFA4”, “NDUFB1”, “NDUFB10”, “NDUFB5”, “NDUFB2”, “NDUFA1”, “SDHA”, “SDHB”, “SDHC”, “SDHD”, “UQCR”, “UCRC”, “UQCRH”, “UQCRB”, “UQCRFS1”, “CYC1”, “UQCRC1”, “UQCRC2”, “UQCRQ”, “COX4I1”, “COX4I2”, “COX5A”, “COX5B”, “COX6A1”, “COX6A2”, “COX6B1”, “COX6C”, “COX7A1”, “COX7A2”, “COX7C”, “ATP5A1”, “ATP5B”, “ATP5C1”, “ATP5D”, “ATP5E”, “ATP5F1”, “ATP5H”, “ATP5I”, “ATP5J2”, “ATP5L”, “ATP5J”, “ATP5G1”, “ATP”

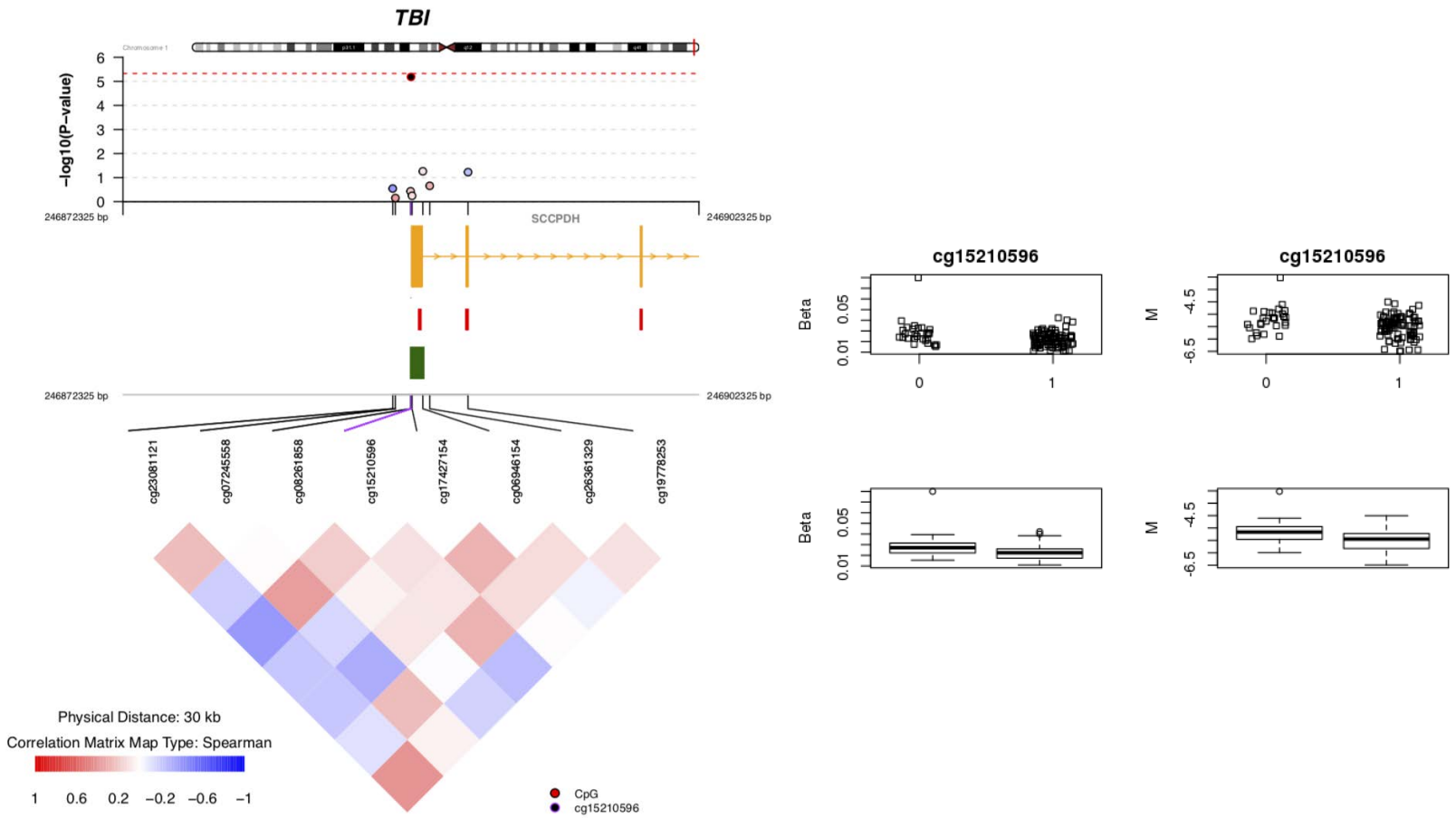
APPENDIX B: COMET, SCATTER AND BOX PLOTS OF SUGGESTIVE SIGNIFICANT CPGS

(Left) For each coMET plot, the upper panel shows the strength and extent of EWAS association signal; in the middle panel, the yellow track shows the gene ENSEMBL information, the red track indicates the SNP track, and the green track indicates the CpG island track. The lower panel shows the correlation between selected CpG in the genomic region. (Right) The scatter and box plots of methylation β value and M value between good and poor recovery group. Y axis: methylation β or M level; x axis: 1 stands for poor recovery group, 0 stands for good recovery group.

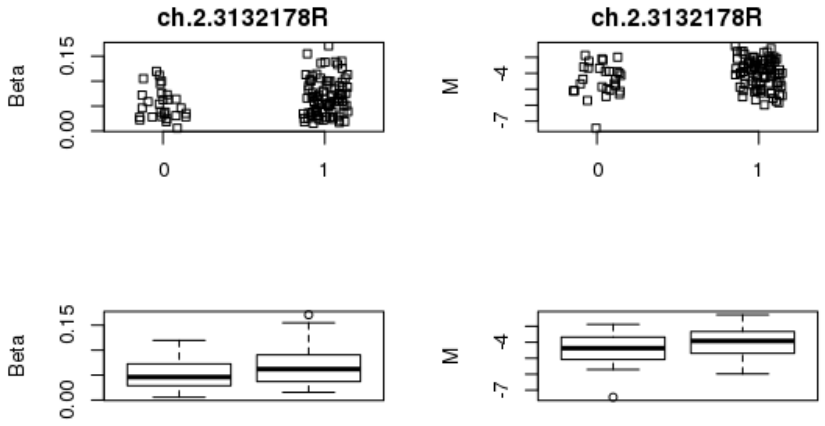
cg08527161

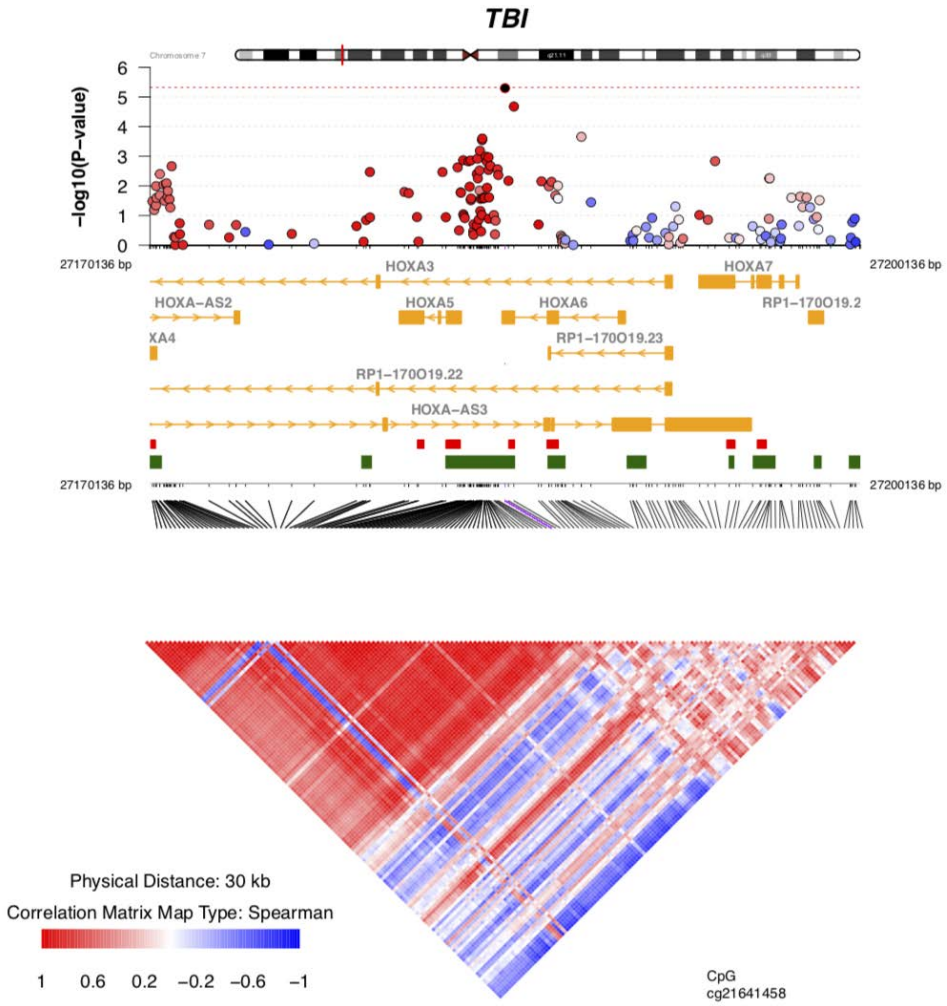

TBI
−log10(P−value)
126279215 bp
126309215 bp
ST3GAL4
KIRREL3
cg16685832
cg06917772
cg19157853
cg20210482
cg12804791
cg02055353
cg15374924
cg08527161
cg18390528
cg10589353
cg25927375
cg09737499
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg08527161

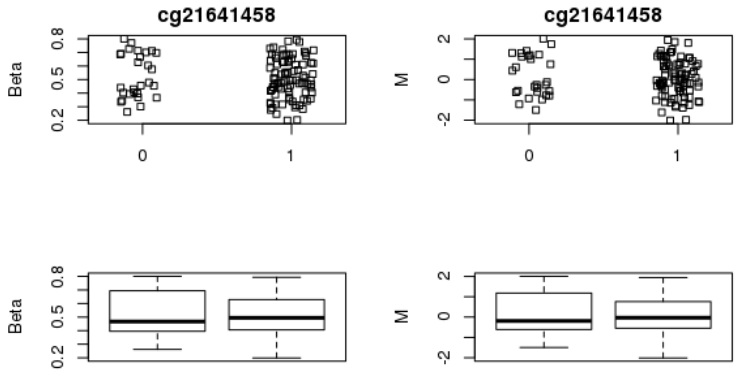

cg08527161
Beta
M
0
1


TBI
−log10(P−value)
246872325 bp
246902325 bp
SCCPDH
cg23081121
cg07245558
cg08261858
cg15210596
cg17427154
cg06946154
cg26361329
cg19778253
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg15210596
cg15210596
Beta
M
cg15210596
Beta
M

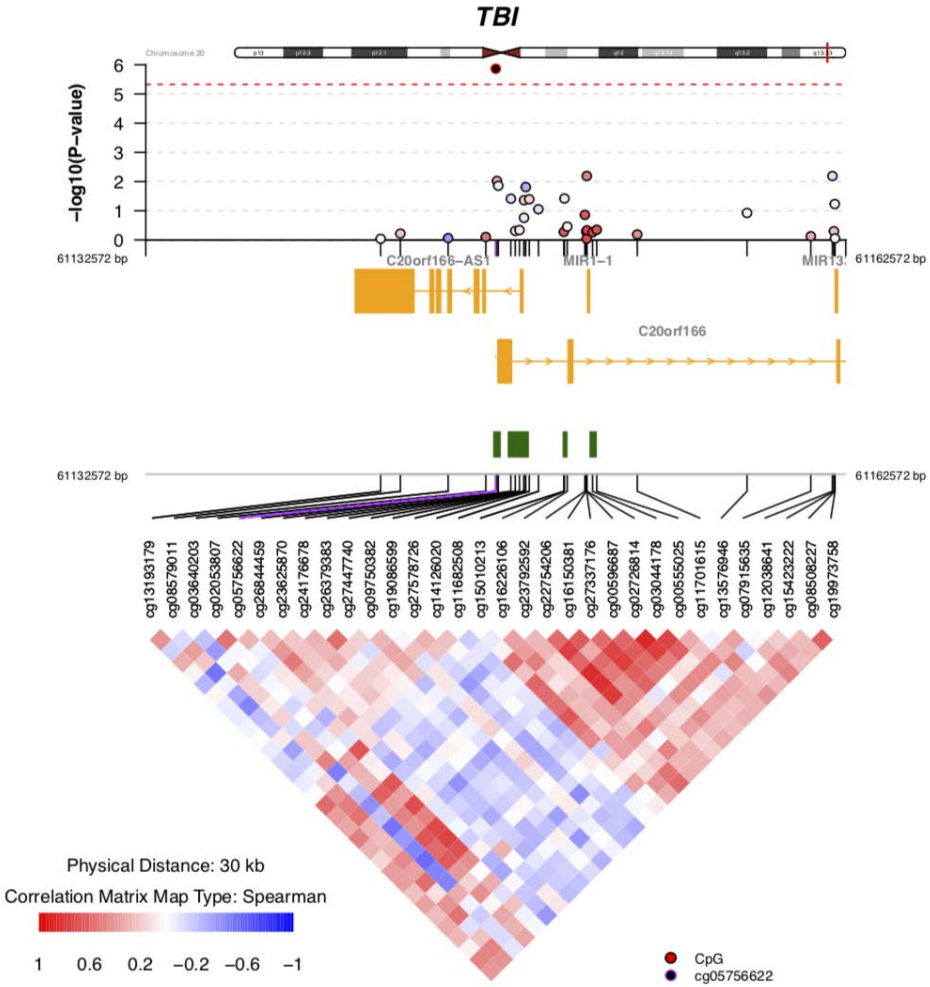
ch.2.3132178R (coMET is plot not available due to technical issue)

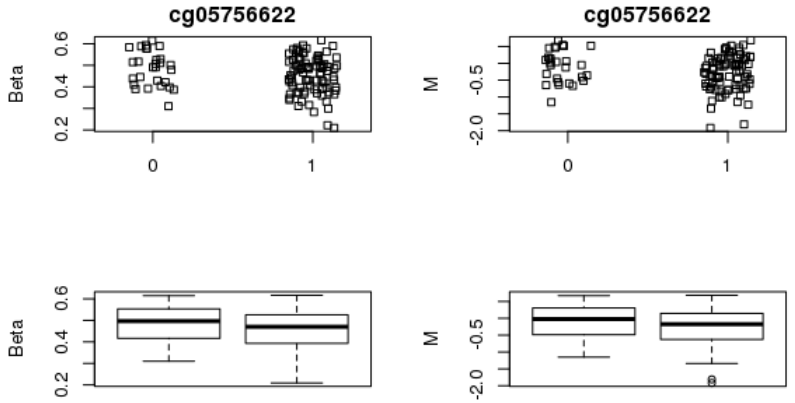



TBI
−log10(P−value)
27170136 bp
27200136 bp
HOXA3
HOXA7
HOXA−AS2
HOXA5
HOXA6
RP1−170O19.2
XA4
RP1−170O19.23
RP1−170O19.22
HOXA−AS3
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg21641458

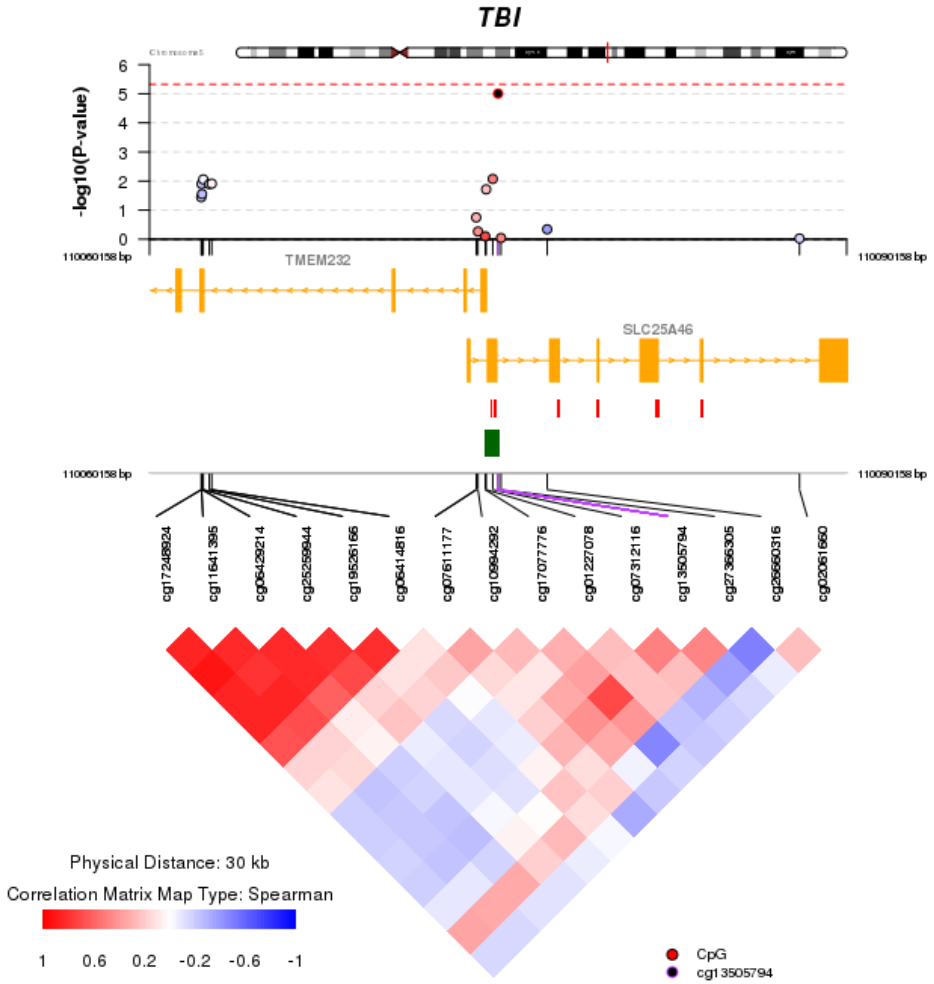

cg21641458
Beta
M
0
1

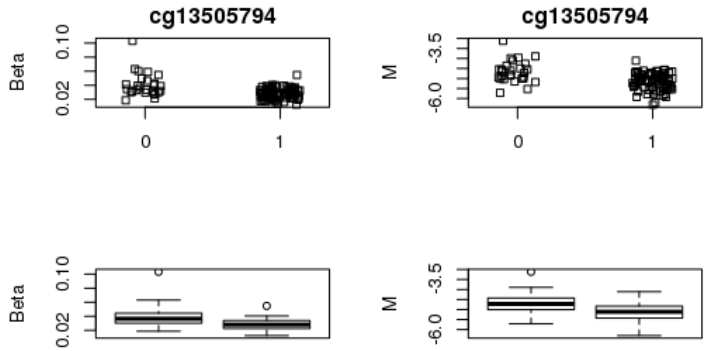
cg05756622


TBI
−log10(P−value)
61132572 bp
61162572 bp
C20orf166−AS1
MIR1−1
MIR13
C20orf166
cg13193179
cg08579011
cg03640203
cg02053807
cg05756622
cg26844459
cg23625870
cg24176678
cg26379383
cg27447740
cg09750382
cg19086599
cg27578726
cg14126020
cg11682508
cg15010213
cg16226106
cg23792592
cg22754206
cg16150381
cg27337176
cg00596687
cg02726814
cg03044178
cg00555025
cg11701615
cg13576946
cg07915635
cg12038641
cg15423222
cg08508227
cg19973758
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg05756622

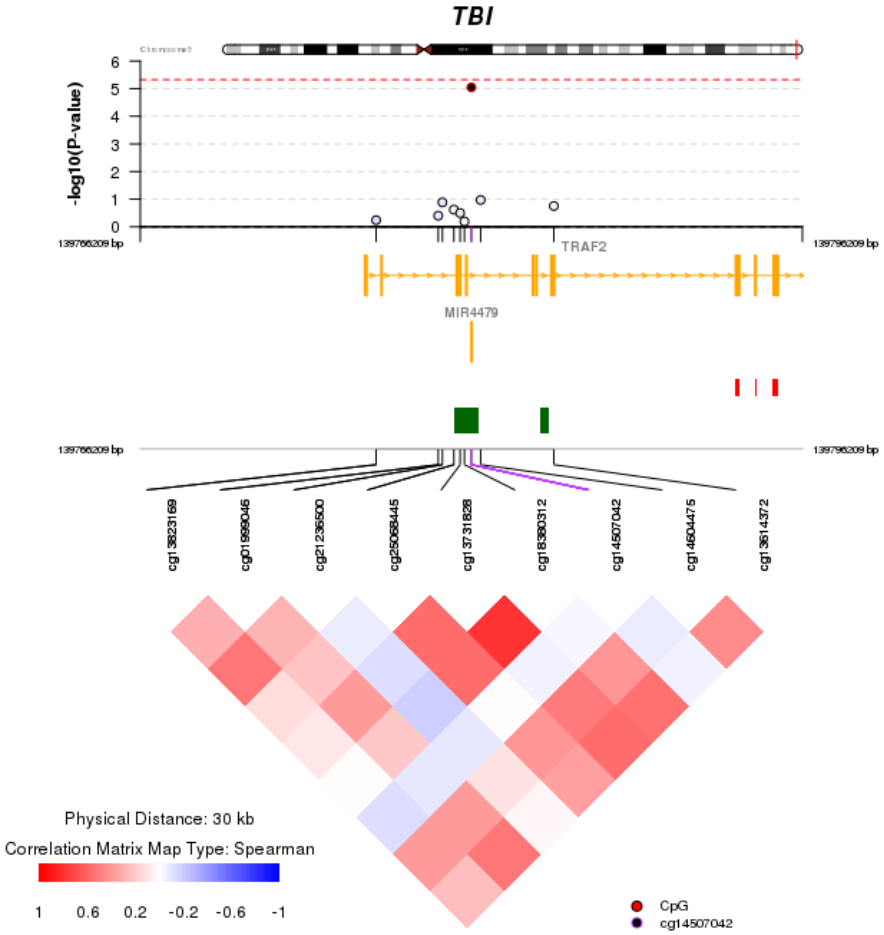

cg05756622
Beta
M
cg05756622
Beta
M

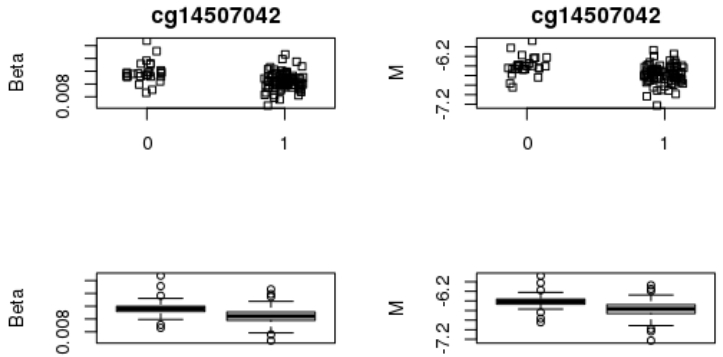
cg13505794


TBI
-log10(P-value)
TMEM232
SLC25A46
cg17248924
cg11641395
cg06429214
cg25259944
cg19526166
cg06414816
cg07611177
cg10994292
cg17077776
cg01227078
cg07312116
cg13505794
cg27366305
cg26660316
cg22061660
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
CpG
cg13505794

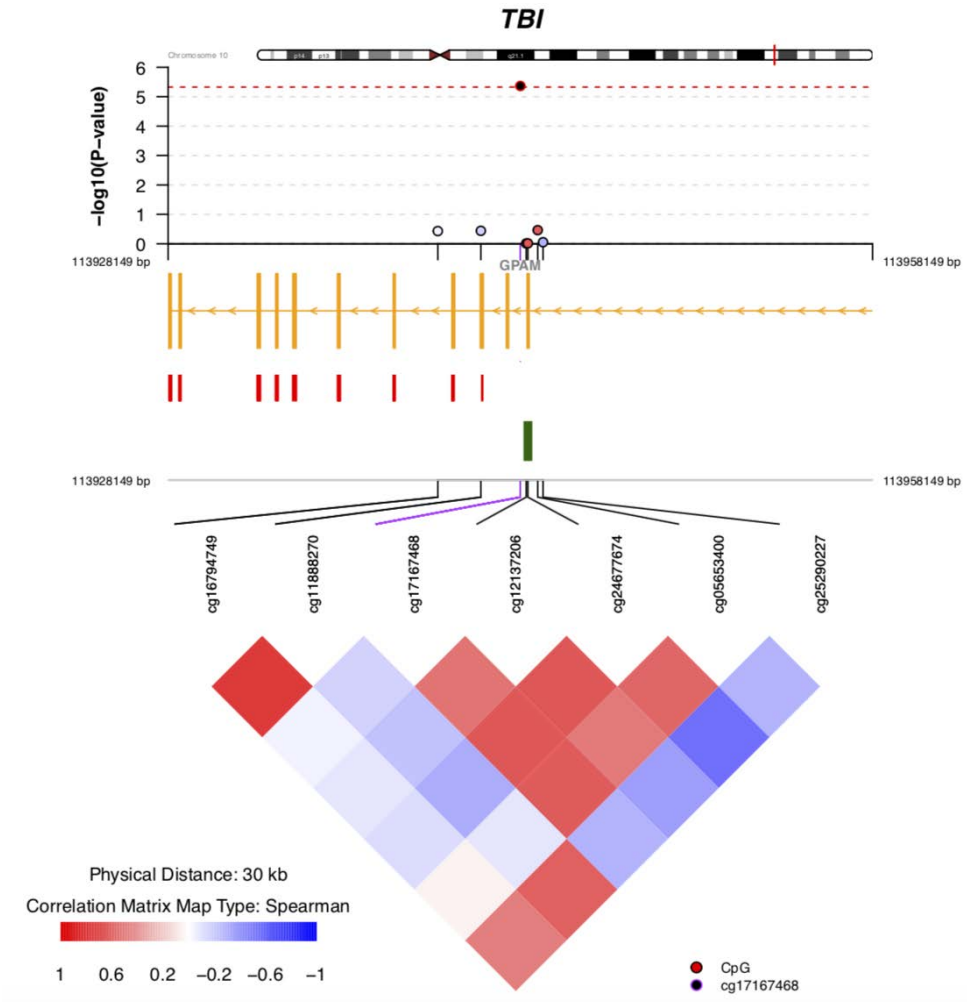

cg13505794
Beta
M
cg13505794
Beta
M

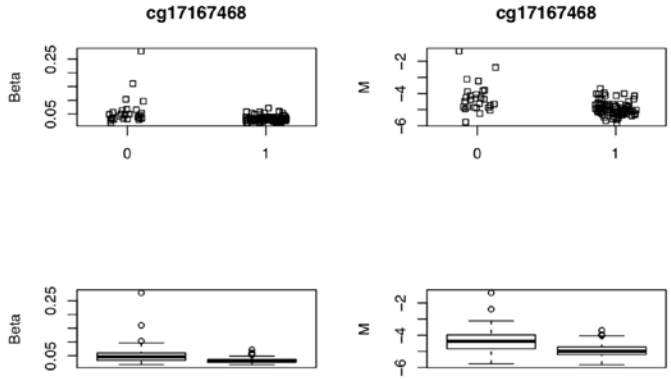
cg14507042


TBI
-log10(P-value)
6
5
4
3
2
1
0
139766209 bp
139796209 bp
TRAF2
MIR4479
cg13823169
cg01999046
cg21235500
cg25068445
cg17318281
cg18380312
cg14507042
cg14604475
cg13614372
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
-0.2
-0.6
-1
CpG
cg14507042

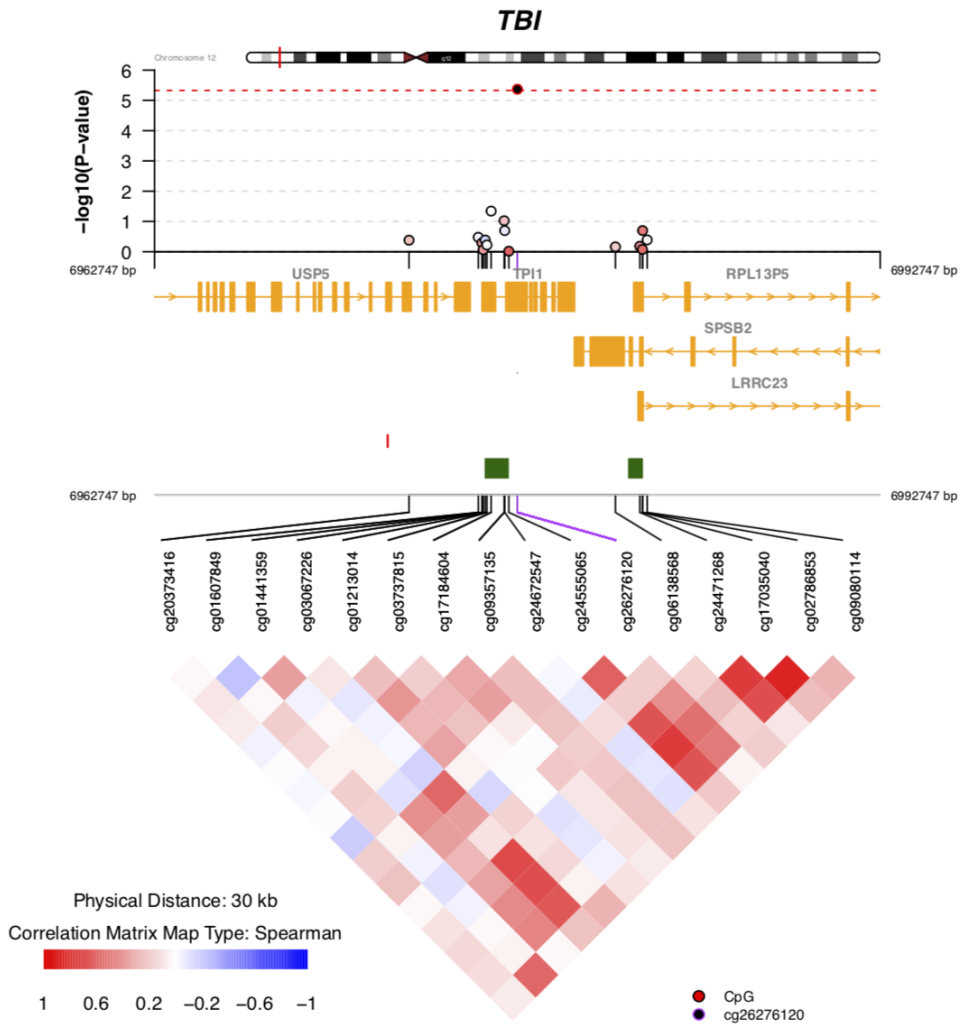

cg14507042
cg14507042
Beta
M
0.008
-6.2
-7.2
0
1
Beta
M
0.008
-6.2
-7.2

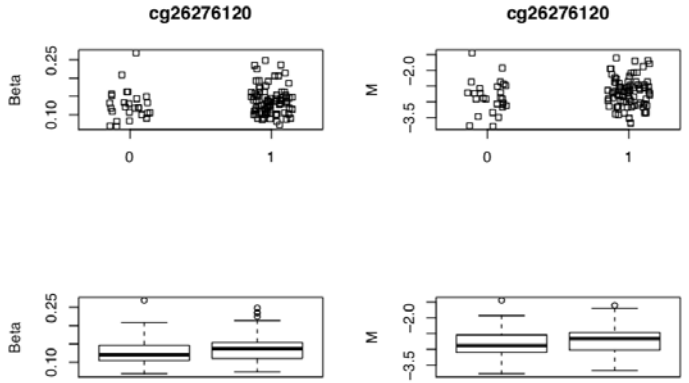
cg17167468


TBI
−log10(P−value)
113928149 bp
113958149 bp
GPAM
cg16794749
cg11888270
cg17167468
cg12137206
cg24677674
cg05653400
cg25290227
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg17167468

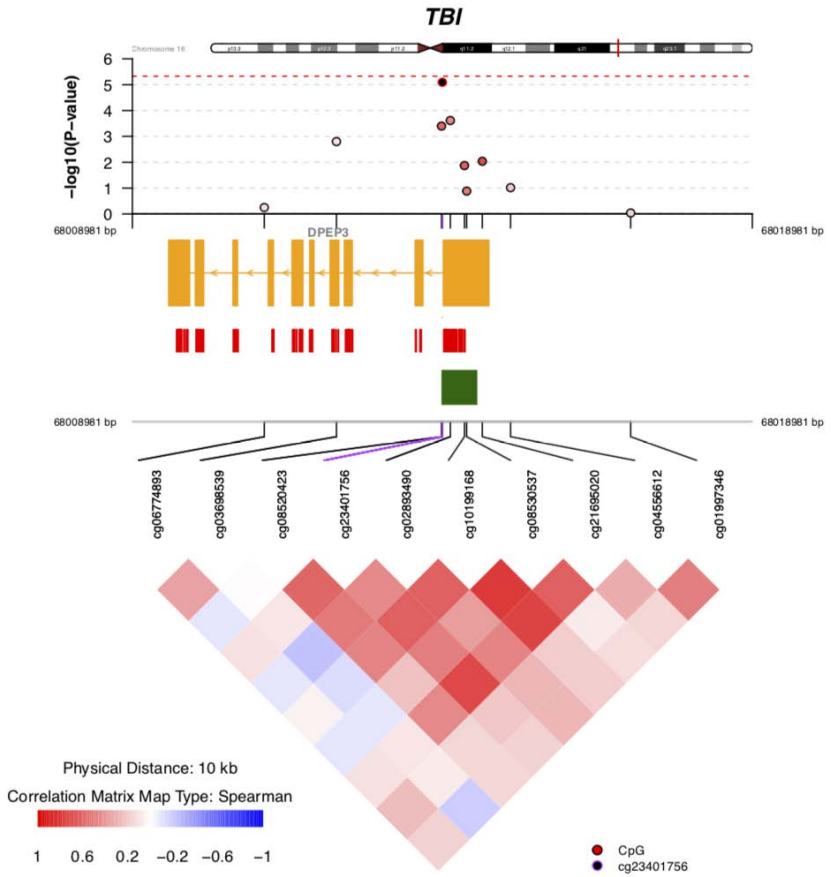

cg17167468
cg17167468
Beta
M

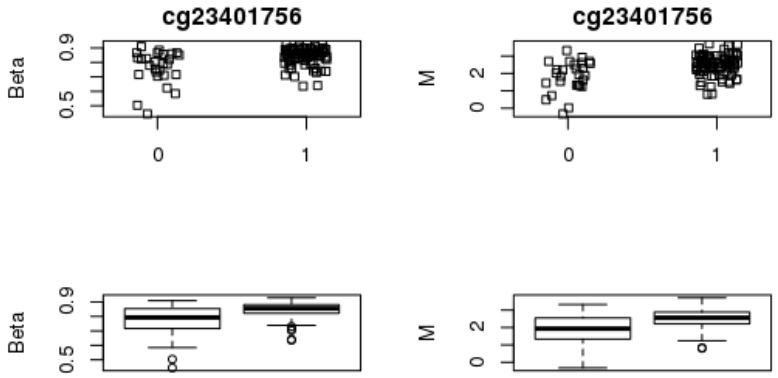
cg26276120

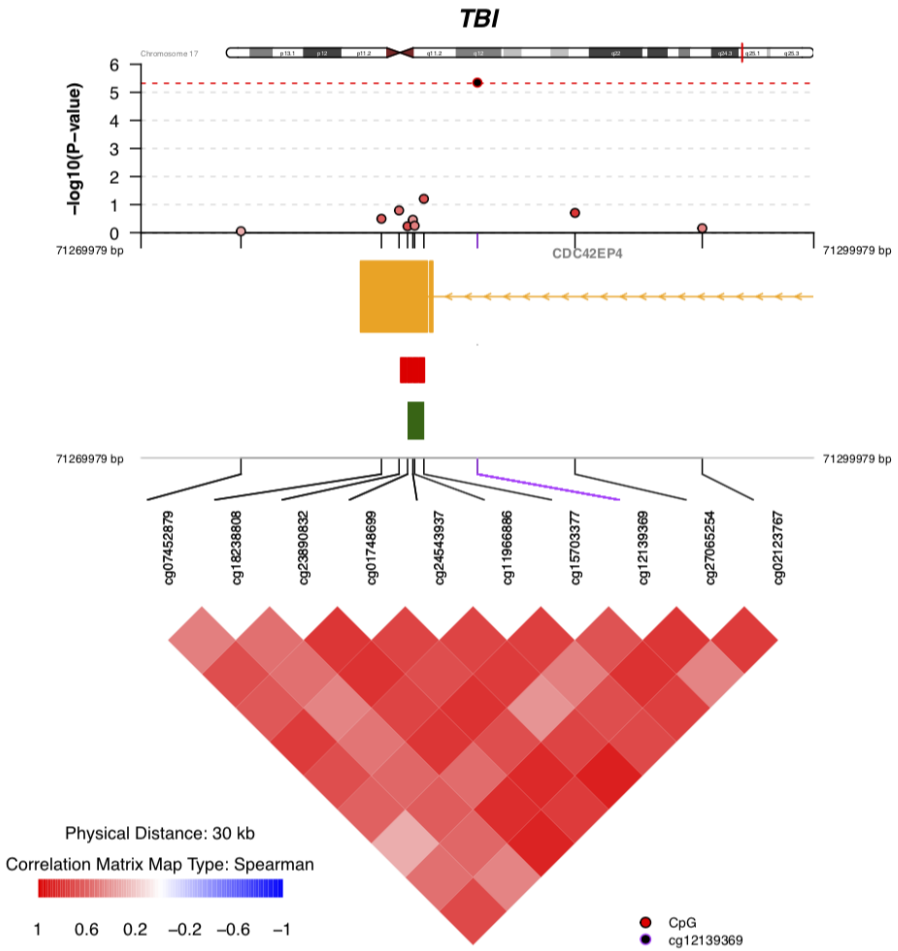

TBI
−log10(P−value)
6962747 bp
6992747 bp
USP5
TPI1
RPL13P5
SPSB2
LRRC23
cg20373416
cg01607849
cg01441359
cg03067226
cg01213014
cg03737815
cg17184604
cg09357135
cg24672547
cg24555065
cg26276120
cg06138568
cg24471268
cg17035040
cg02786853
cg09080114
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg26276120

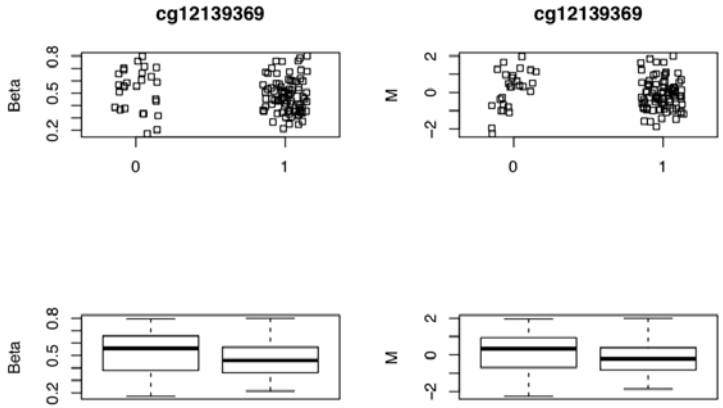

cg26276120
cg26276120
Beta
M

cg23401756

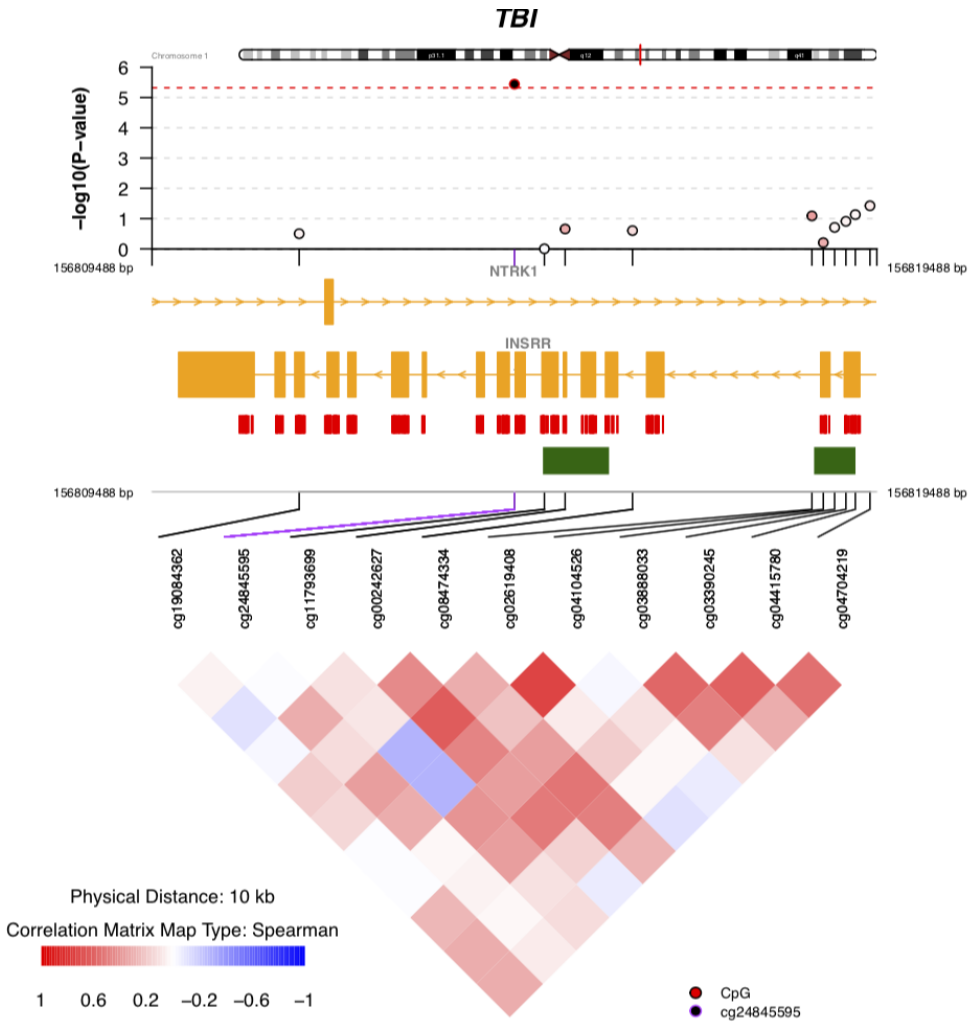

TBI
-log10(P-value)
68008981 bp
68018981 bp
DPEP3
cg06774893
cg03698539
cg08520423
cg23401756
cg02883490
cg10199168
cg08530537
cg21695020
cg04566612
cg01997346
Physical Distance: 10 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg23401756

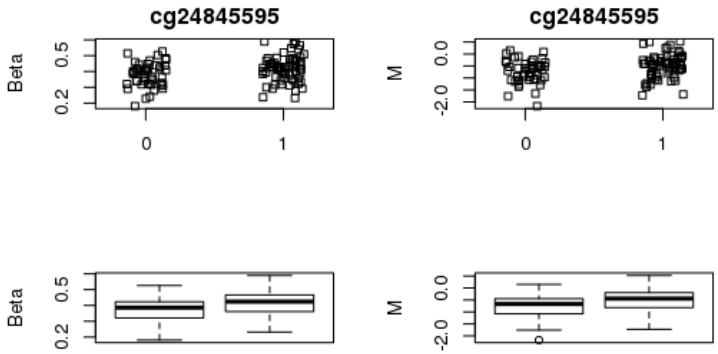

cg23401756
Beta
M
cg23401756
Beta
M


TBI
Chromosome 17
−log10(P−value)
71269979 bp
71299979 bp
CDC42EP4
cg07452879
cg18238808
cg23890832
cg01748699
cg24543937
cg11966886
cg15703377
cg12139369
cg27065254
cg02123767
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg12139369

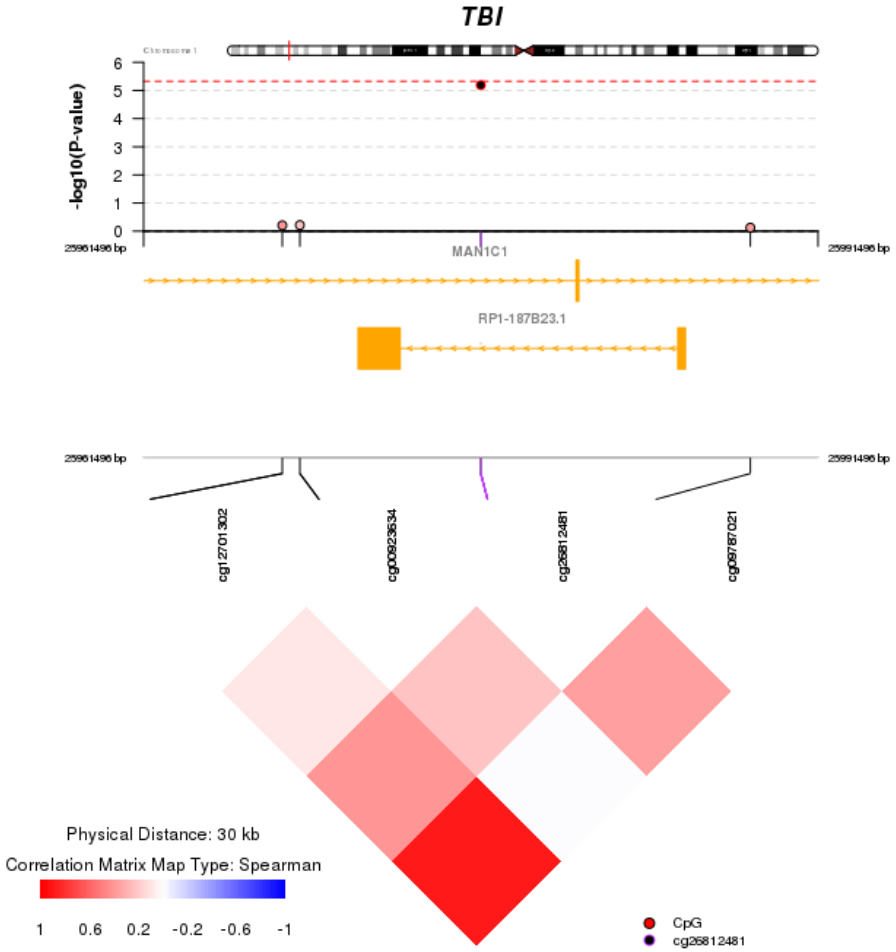

cg12139369
cg12139369
Beta
M

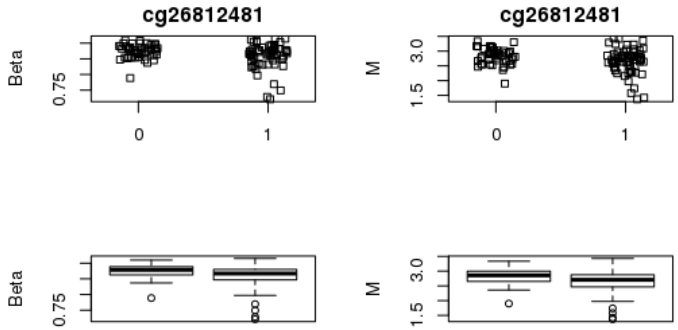
cg24845595


TBI
Chromosome 1
−log10(P−value)
156809488 bp
156819488 bp
NTRK1
INSRR
cg19084362
cg24845595
cg11793699
cg00242627
cg08474334
cg02619408
cg04104526
cg03888033
cg03390245
cg04415780
cg04704219
Physical Distance: 10 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg24845595

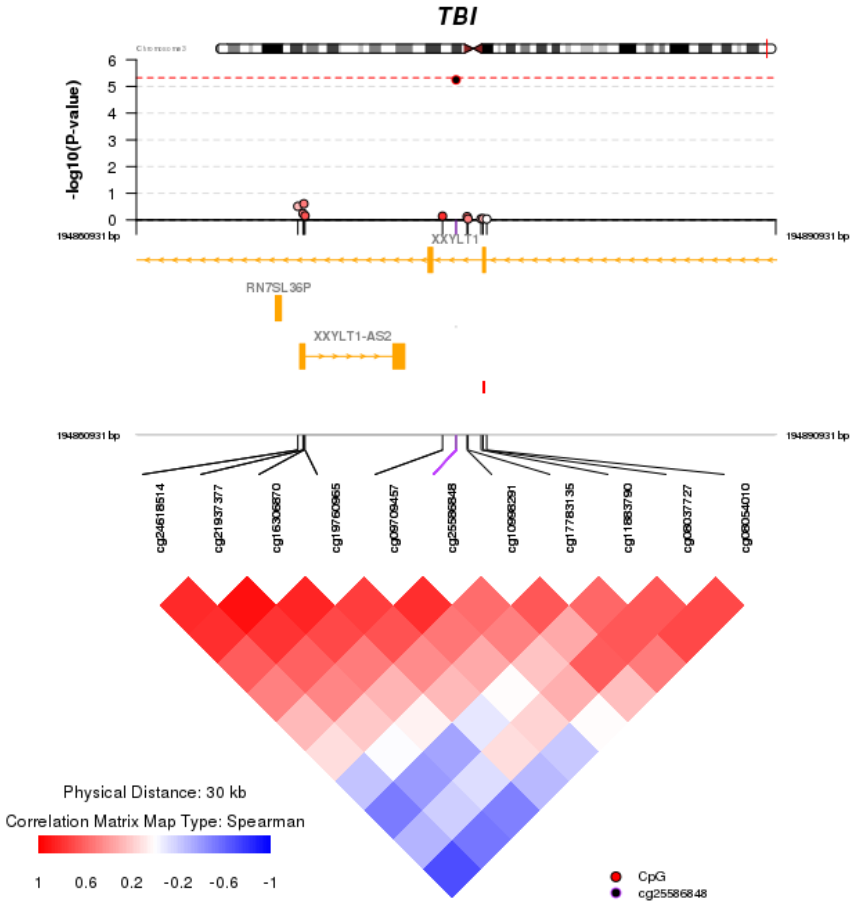

cg24845595
Beta
M
cg24845595
Beta
M

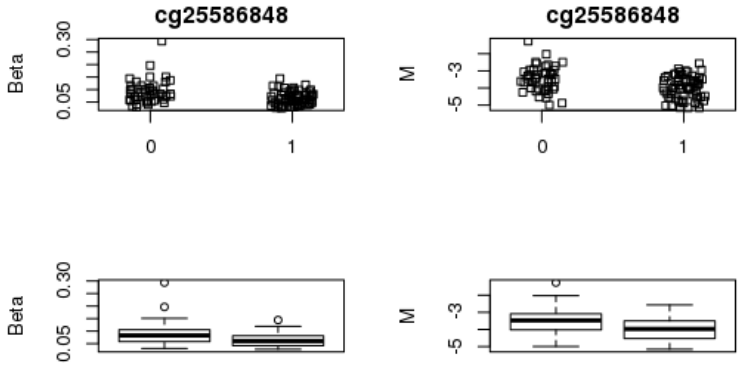
cg26812481


TBI
-log10(P-value)
6
5
4
3
2
1
0
25961496 bp
25991496 bp
MAN1C1
RP1-187B23.1
cg12701302
cg00923534
cg26812481
cg09787021
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
-0.2
-0.6
-1
CpG
cg26812481

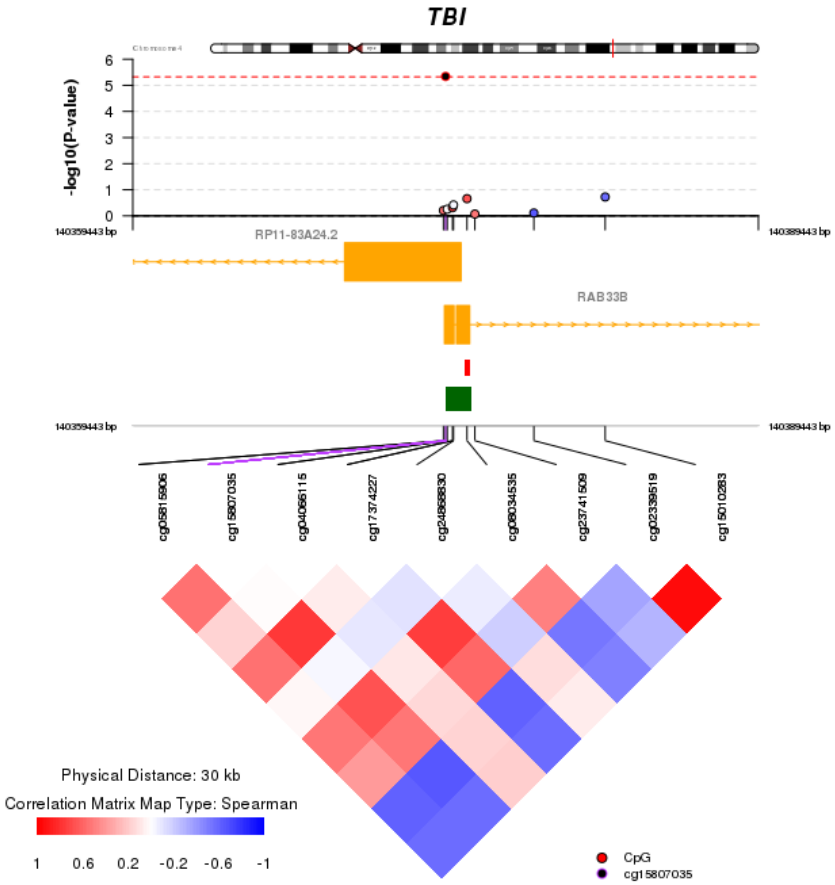

cg26812481
Beta
0.75
0
1
cg26812481
M
3.0
1.5
0
1
Beta
0.75
M
3.0
1.5

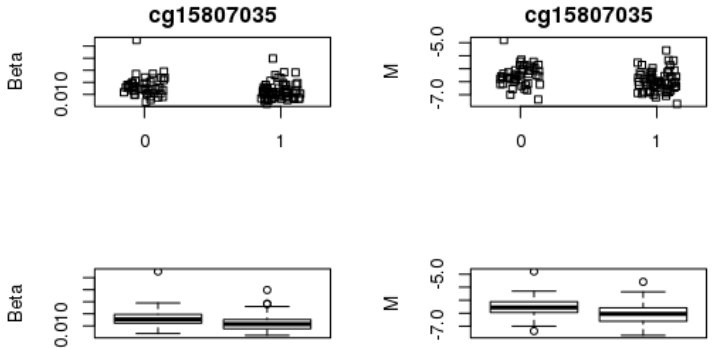
cg25586848


TBI
-log10(P-value)
194860931 bp
194890931 bp
XXYLT1
RN7SL36P
XXYLT1-AS2
cg24618514
cg21937377
cg16306870
cg19760965
cg09709457
cg25586848
cg10998291
cg17783135
cg11883790
cg08037727
cg08054010
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 -0.2 -0.6 -1
CpG
cg25586848

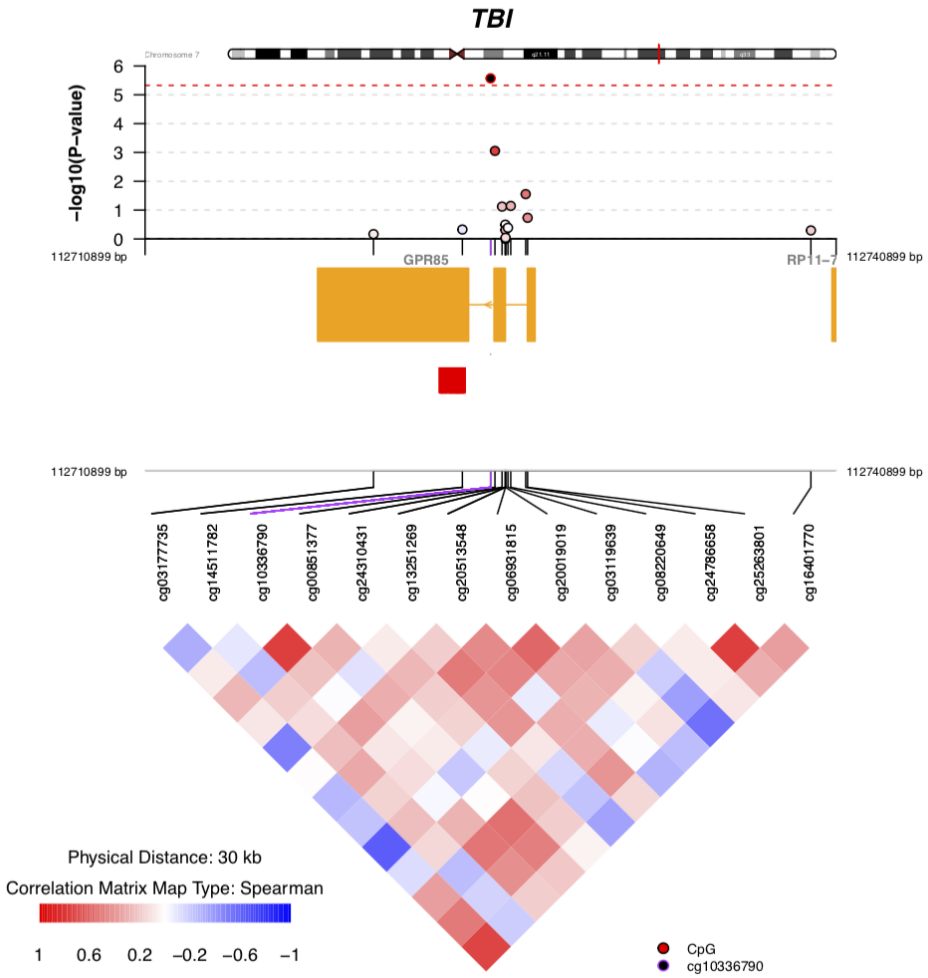

cg25586848
Beta
M
0
1

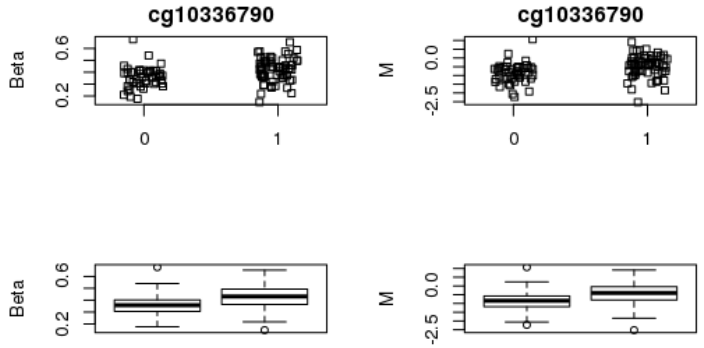
cg15807035

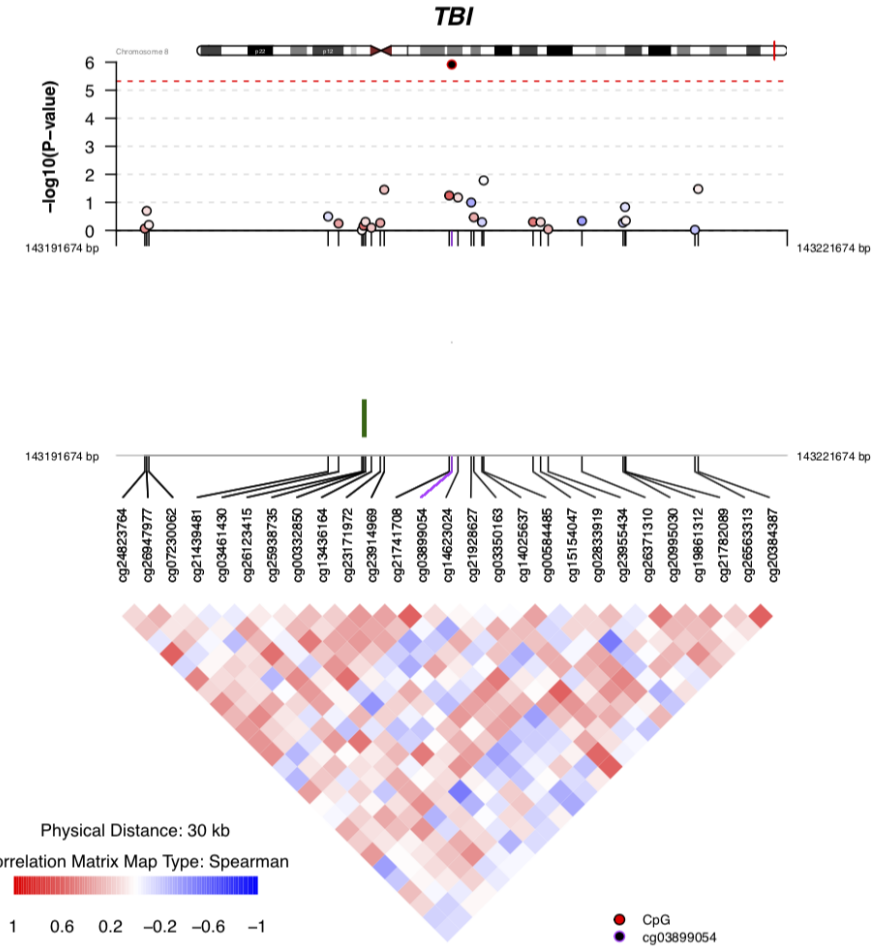

TBI
-log10(P-value)
RP11-83A24.2
RAB33B
cg05815906
cg15807035
cg04066115
cg17374227
cg24868830
cg06034535
cg23741509
cg02338519
cg15010283
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 -0.2 -0.6 -1
CpG
cg15807035

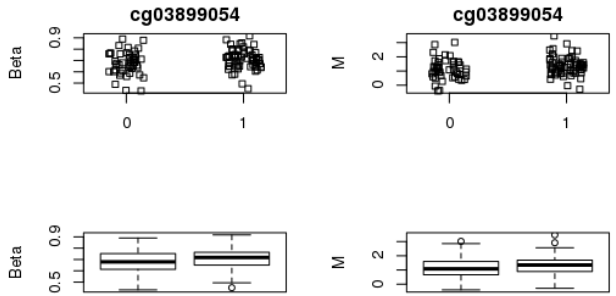

cg15807035
Beta
M
cg15807035
Beta
M

cg10336790

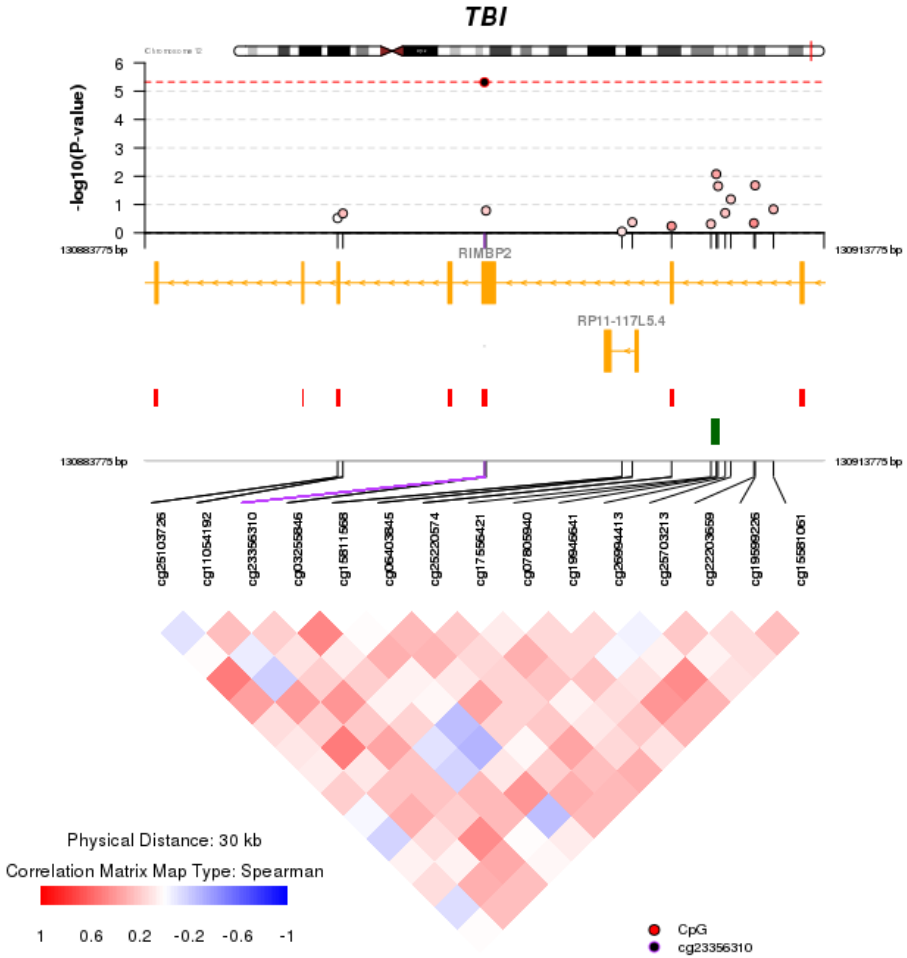

TBI
−log10(P−value)
112710899 bp
112740899 bp
GPR85
RP11−7
cg03177735
cg14511782
cg10336790
cg00851377
cg24310431
cg13251269
cg20513548
cg06931815
cg20019019
cg03119639
cg08220649
cg24786658
cg25263801
cg16401770
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg10336790

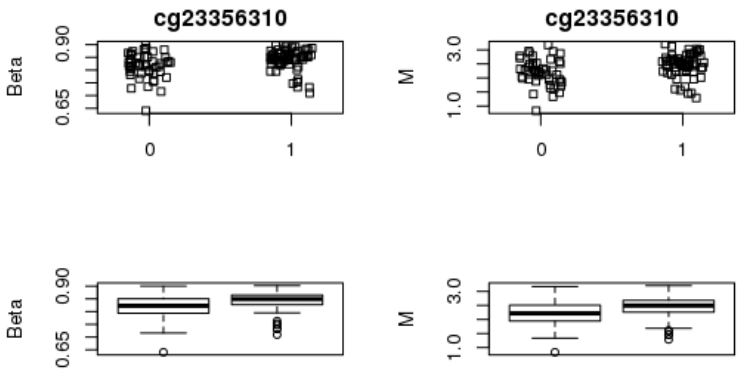

cg10336790
Beta
M
cg10336790
Beta
M

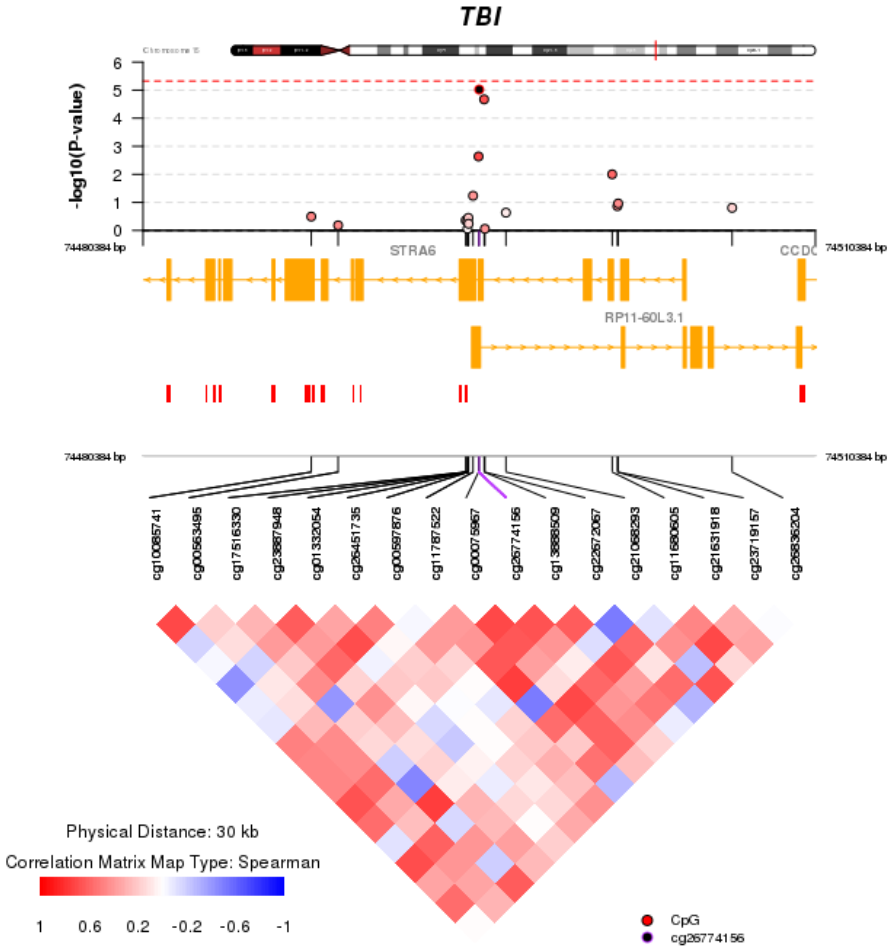

TBI
−log10(P−value)
143191674 bp
143221674 bp
cg24823764
cg26947977
cg07230062
cg21439481
cg03461430
cg26123415
cg25938735
cg00332850
cg13436164
cg23171972
cg23914969
cg21741708
cg03899054
cg14623024
cg21928627
cg03350163
cg14025637
cg00584485
cg15154047
cg02833919
cg23955434
cg26371310
cg20995030
cg19861312
cg21782089
cg26563313
cg20384387
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg03899054

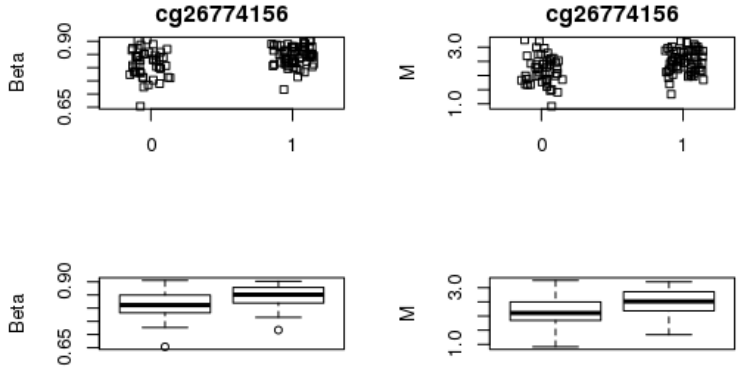

cg03899054
Beta
M
cg03899054
Beta
M

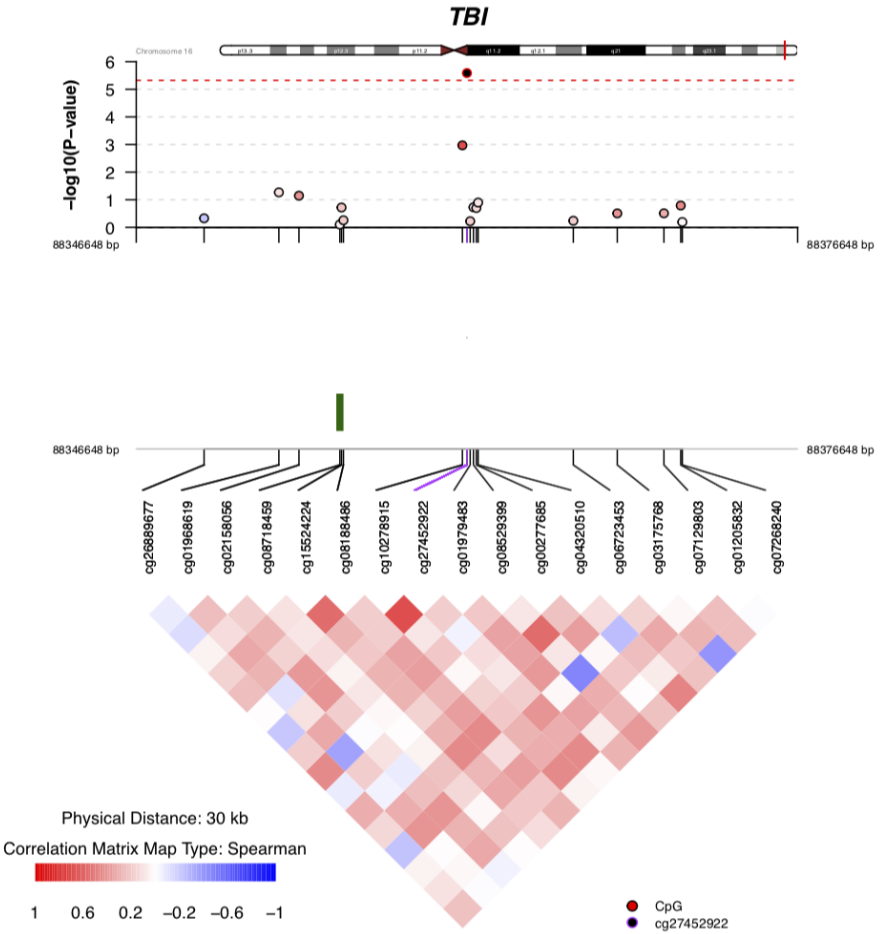
cg23356310

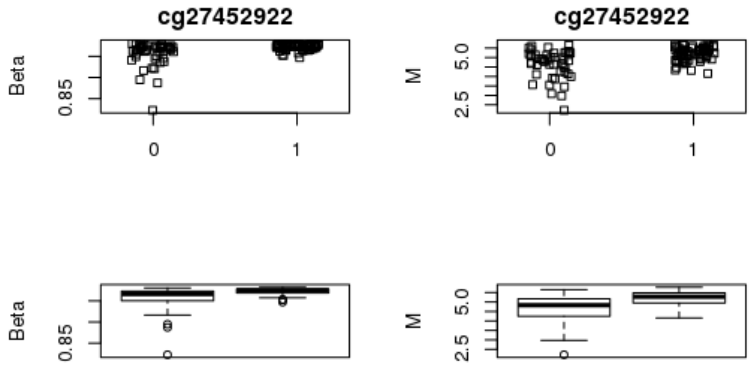

TBI
-log10(P-value)
6
5
4
3
2
1
0
130883775 bp
130913775 bp
RIMBP2
RP11-117L5.4
cg25103726
cg11054192
cg23356310
cg03256846
cg15811568
cg06403845
cg25220574
cg17556421
cg07805940
cg19946641
cg26994413
cg25703213
cg22203659
cg19599226
cg15681061
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
-0.2
-0.6
-1
CpG
cg23356310


cg23356310
Beta
0.90
0.65
0
1
cg23356310
M
3.0
1.0
0
1
Beta
0.90
0.65
M
3.0
1.0

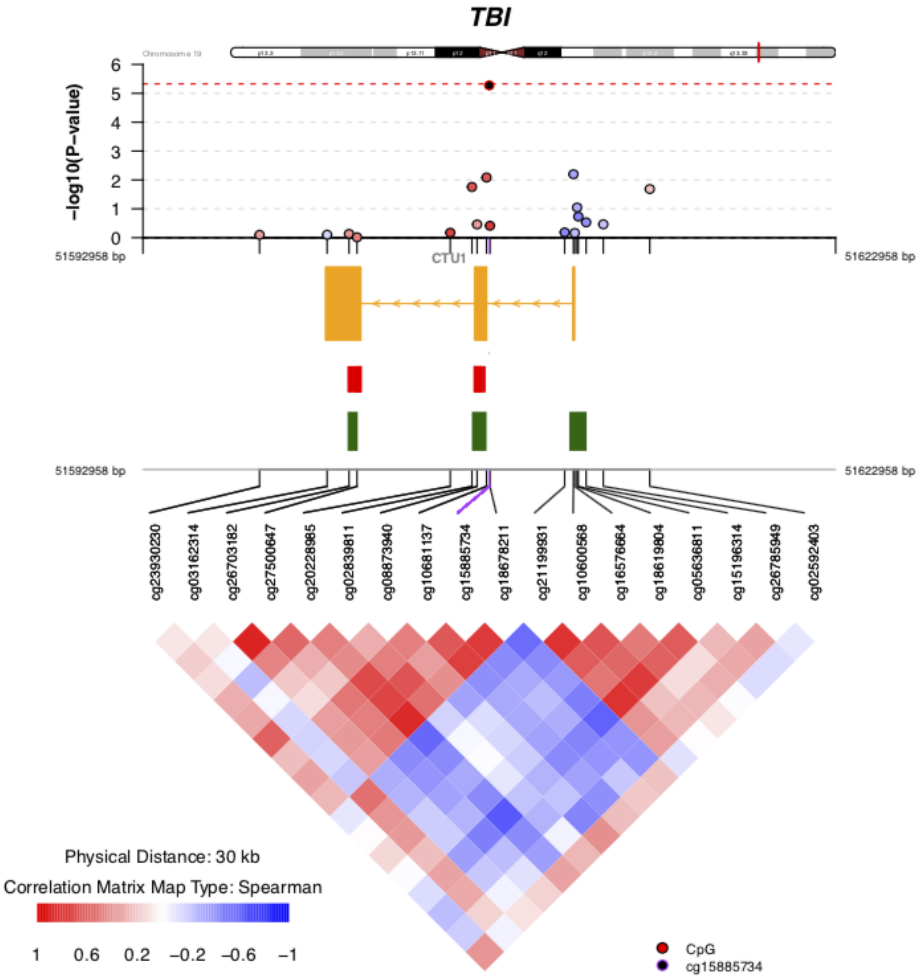

TBI
-log10(P-value)
6
5
4
3
2
1
0
74480384 bp
74510384 bp
STRA6
CCDC
RP11-60L3.1
cg10085741
cg00563495
cg17516330
cg23887948
cg01332054
cg25451735
cg00597876
cg11787522
cg00078967
cg26774156
cg13888509
cg22672067
cg21068293
cg11680605
cg21631918
cg23719157
cg26836204
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1
0.6
0.2
-0.2
-0.6
-1
CpG
cg26774156

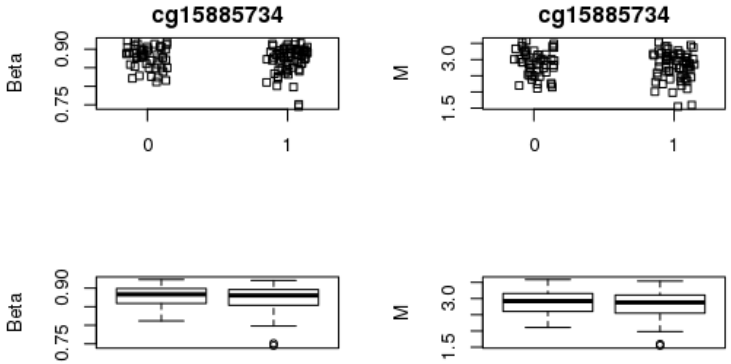

cg26774156
Beta
0.90
0.65
0
1
cg26774156
M
3.0
1.0
0
1
Beta
0.90
0.65
M
3.0
1.0

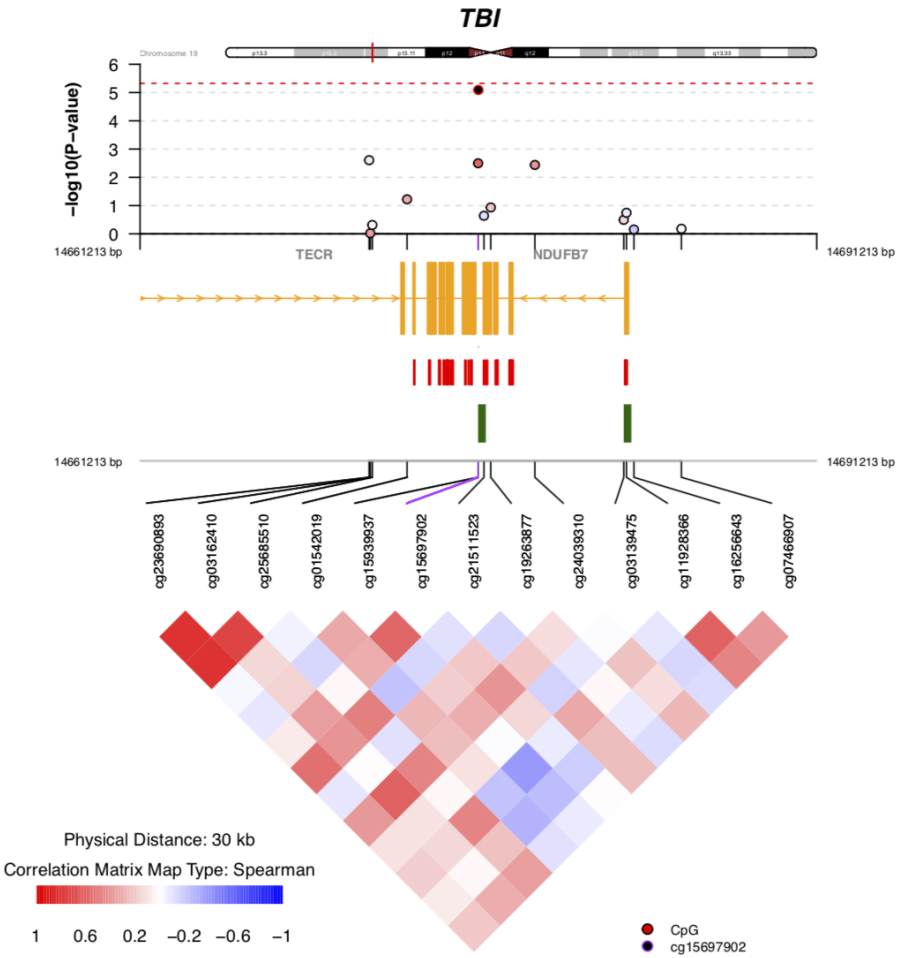

TBI
−log10(P−value)
88346648 bp
88376648 bp
cg26889677
cg01968619
cg02158056
cg08718459
cg15524224
cg08188486
cg10278915
cg27452922
cg01979483
cg08529399
cg00277685
cg04320510
cg06723453
cg03175768
cg07129803
cg01205832
cg07268240
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg27452922

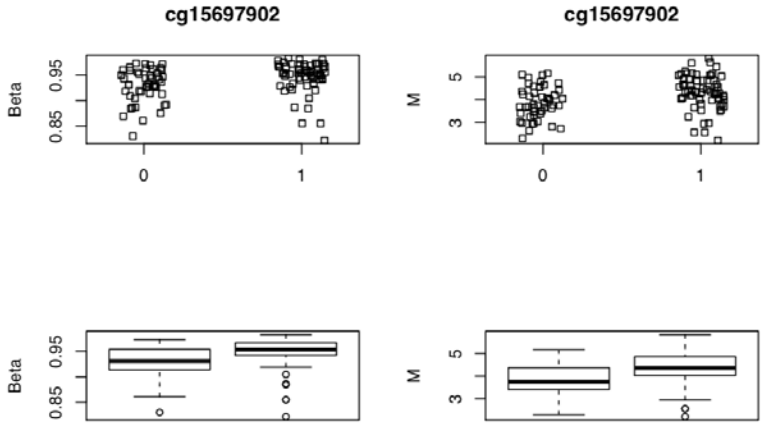

cg27452922
cg27452922
Beta
M
0.85
2.5
5.0
0
1

cg15885734

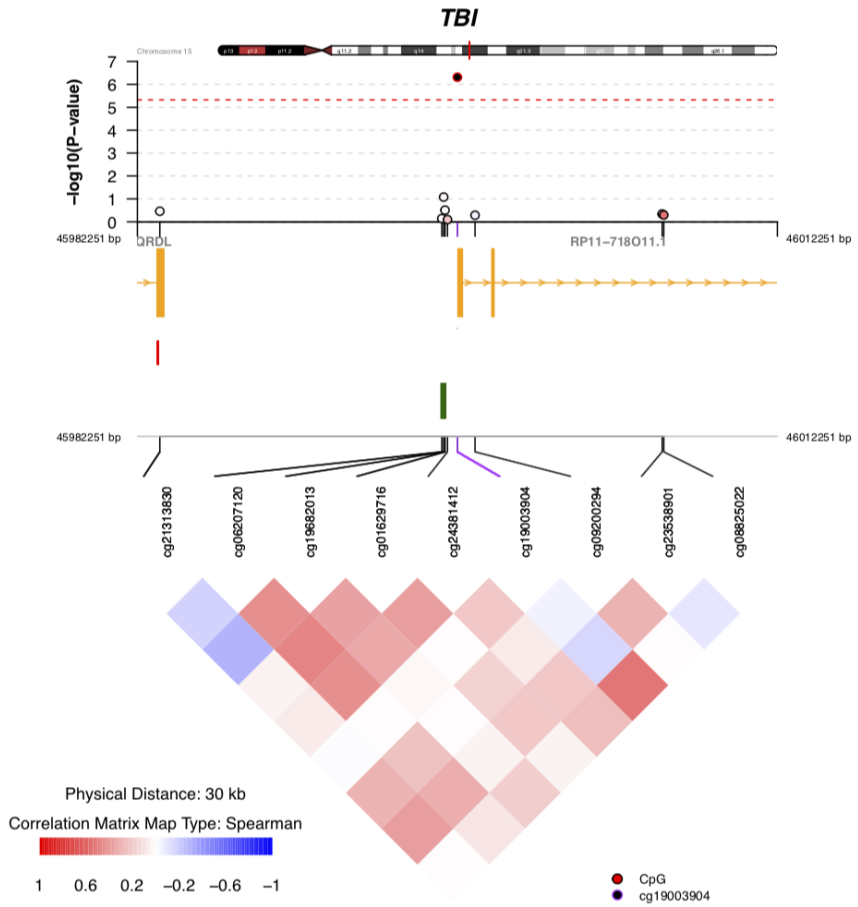

TBI
−log10(P−value)
51592958 bp
51622958 bp
CTU1
cg23930230
cg03162314
cg26703182
cg27500647
cg20228985
cg02839811
cg08873940
cg10681137
cg15885734
cg18678211
cg21199931
cg10600568
cg16576664
cg18619804
cg05636811
cg15196314
cg26785949
cg02592403
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg15885734

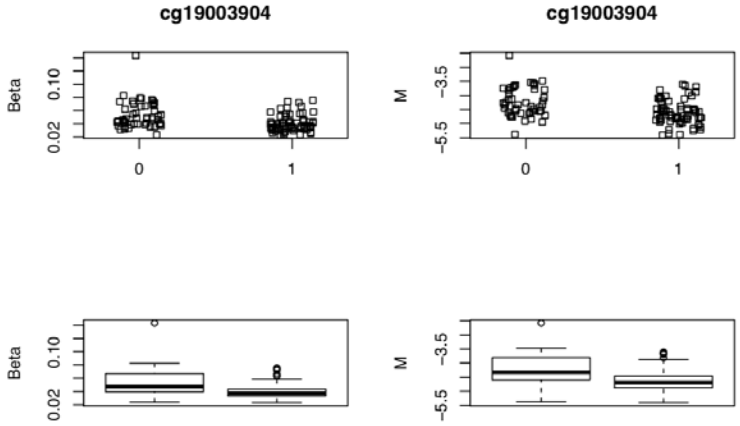

cg15885734
Beta
M
cg15885734
Beta
M


TBI
−log10(P−value)
14661213 bp
14691213 bp
TECR
NDUFB7
cg23690893
cg03162410
cg25685510
cg01542019
cg15939937
cg15697902
cg21511523
cg19263877
cg24039310
cg03139475
cg11928366
cg16256643
cg07466907
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg15697902

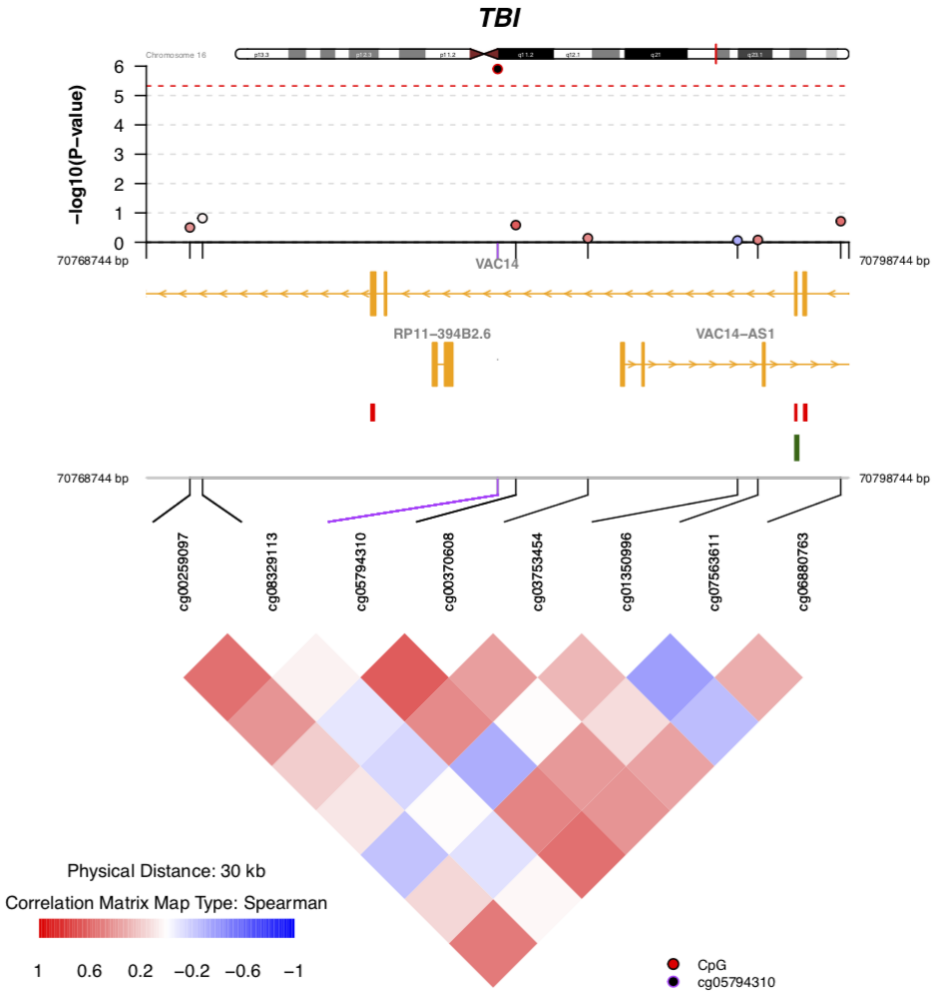

cg15697902
cg15697902
Beta
M

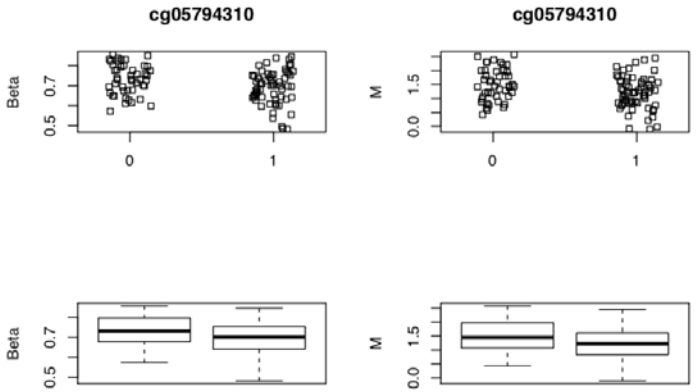
cg19003904


TBI
−log10(P−value)
45982251 bp
46012251 bp
QRDL
RP11−718O11.1
cg21313830
cg06207120
cg19682013
cg01629716
cg24381412
cg19003904
cg09200294
cg23538901
cg08825022
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg19003904


cg19003904
cg19003904
Beta
M

cg05794310


TBI
−log10(P−value)
70768744 bp
70798744 bp
VAC14
RP11−394B2.6
VAC14−AS1
cg00259097
cg08329113
cg05794310
cg00370608
cg03753454
cg01350996
cg07563611
cg06880763
Physical Distance: 30 kb
Correlation Matrix Map Type: Spearman
1 0.6 0.2 −0.2 −0.6 −1
CpG
cg05794310


cg05794310
cg05794310
Beta
M

BIBLIOGRAPHY

[1] U.S. Department of Health and Human Serivces, Centers for Disease Control and Prevention. "Traumatic brain injury in the United States: epidemiology and rehabilitation." *Congress Rep*. 2014.

[2] Wong, Victor S., and Brett Langley. "Epigenetic changes following traumatic brain injury and their implications for outcome, recovery and therapy." *Neuroscience letters* 625 (2016): 26-33.

[3] Risdall, Jane E., and David K. Menon. "Traumatic brain injury." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366.1562 (2011): 241-250.

[4] McMillan, Tom, et al. "The Glasgow Outcome Scale—40 years of application and refinement." *Nature Reviews Neurology* 12.8 (2016): 477.

[5] Levin, Harvey S., et al. "The neurobehavioural rating scale: assessment of the behavioural sequelae of head injury by the clinician." *Journal of Neurology, Neurosurgery & Psychiatry*50.2 (1987): 183-193.

[6] Bellon, Kimberly, et al. "Disability rating scale." *The Journal of head trauma rehabilitation* 27.6 (2012): 449-451.

[7] Derogatis, L. R. (2001). Brief symptom inventory 18. Johns Hopkins University.

[8] Diener, E. D., et al. "The satisfaction with life scale." *Journal of personality assessment* 49.1 (1985): 71-75.

[9] Zhang, Zhi-Yuan, et al. "Global hypomethylation defines a sub-population of reactive microglia/macrophages in experimental traumatic brain injury." *Neuroscience letters* 429.1 (2007): 1-6.

[10] Lundberg, Johan, et al. "Traumatic brain injury induces relocalization of DNA-methyltransferase 1." *Neuroscience letters* 457.1 (2009): 8-11.

[11] Haghighi, Fatemeh, et al. "Neuronal DNA methylation profiling of blast-related traumatic brain injury." *Journal of neurotrauma* 32.16 (2015): 1200-1209.

[12] Michels, Karin B., et al. "Recommendations for the design and analysis of epigenome-wide association studies." *Nature methods* 10.10 (2013): 949

[13] Johanson, Conrad, et al. "Traumatic brain injury and recovery mechanisms: peptide modulation of periventricular neurogenic regions by the choroid plexus–CSF nexus." *Journal of neural transmission* 118.1 (2011): 115-133.

[14] Birney, Ewan, George Davey Smith, and John M. Greally. "Epigenome-wide association studies and the interpretation of disease-omics." *PLoS genetics* 12.6 (2016): e1006105.

[15] Centers for Disease Control and Prevention. "Traumatic brain injury in the United States: epidemiology and rehabilitation." *Congress Rep*. 2014.

[16] Du, Pan, et al. "Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis." *BMC bioinformatics* 11.1 (2010): 587.

[17] Xu, Zongli, et al. "ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip." *Nucleic acids research* 44.3 (2015): e20-e20.

[18] Aryee, Martin J., et al. "Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays." *Bioinformatics* 30.10 (2014): 1363-1369.

[19] Chen, Jun, et al. "CpGFilter: model-based CpG probe filtering with replicates for epigenome-wide association studies." *Bioinformatics* 32.3 (2015): 469-471.

[20] Fortin, Jean-Philippe, et al. "Functional normalization of 450k methylation array data improves replication in large cancer studies." *Genome biology* 15.11 (2014): 503.

[21] Chi, Jocelyn T., Eric C. Chi, and Richard G. Baraniuk. "k-pod: A method for k-means clustering of missing data." *The American Statistician* 70.1 (2016): 91-99

[22] Ritchie, Matthew E., et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies." *Nucleic acids research* 43.7 (2015): e47-e47.

[23] Leek, Jeffrey T., et al. "sva: Surrogate variable analysis." *R package version* 3.25.4 (2017).

[24] Triche Jr, T. "Illuminahumanmethylation450k. db: Illumina human methylation 450k annotation data." R package version 0.6.0 (2016).

[25] Martin, Tiphaine C., et al. "coMET: an R plotting package to visualize regional plots of epigenome-wide association scan results." *QG14*. http://quantgen.soc.srcf.net/qg14/. (2014)

[26] Martin, Tiphaine C., et al. "coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns." *BMC bioinformatics* 16.1 (2015): 131.

[27] Conley, Yvette P., et al. "Mitochondrial polymorphisms impact outcomes after severe traumatic brain injury." *Journal of neurotrauma* 31.1 (2014): 34-41.

[28] Martin, E. Anne, et al. "The intellectual disability gene Kirrel3 regulates target-specific mossy fiber synapse development in the hippocampus." *Elife 4* (2015).

[29] Laitila, Jenni, et al. "Expression of multiple nebulin isoforms in human skeletal muscle and brain." *Muscle & nerve* 46.5 (2012): 730-737.

[30] Gao, Fan, et al. "DNA methylation in the malignant transformation of meningiomas." *PloS one* 8.1 (2013): e54114

[31] Kishida, Yugo, et al. "Epigenetic subclassification of meningiomas based on genome-wide DNA methylation analyses." Carcinogenesis 33.2 (2011): 436-441

[32] Summaries for SLC25A46 Gene. GeneCards, Human Gene Database, www.genecards.org/cgi-bin/carddisp.pl?gene=SLC25A46&keywords=SLC25A46. Accessed 22 Apr. 2018

[33] Summaries for TRAF2 Gene. GeneCards, Human Gene Database, www.genecards.org/cgi-bin/carddisp.pl?gene=TRAF2&keywords=TRAF2. Accessed 22 Apr. 2018

[34] Olah, Judit, et al. "Triosephosphate isomerase deficiency: a neurodegenerative misfolding disease." *Biochemical Society Transactions* 30.2 (2002): 30-38.

[35] Ageta-Ishihara, Natsumi, et al. "CDC42EP4, a perisynaptic scaffold protein in Bergmann glia, is required for glutamatergic tripartite synapse configuration." *Neurochemistry international* (2018).

[36] Gensous, Noémie, et al. "Responders and non-responders to influenza vaccination: A DNA methylation approach on blood cells." *Experimental gerontology* 105 (2018): 94-100.

[37] Summaries for NTRK1 Gene. GeneCards, Human Gene Database, http://www.genecards.org/cgi-bin/carddisp.pl?gene=NTRK1&keywords=NTRK1. Accessed 22 Apr. 2018

[38] Calligaris, Raffaella, et al. "Blood transcriptomics of drug-naïve sporadic Parkinson's disease patients." *BMC genomics* 16.1 (2015): 876.

[39] Rushton, Michael D., et al. "Characterization of the cartilage DNA methylome in knee and hip osteoarthritis." *Arthritis & rheumatology* 66.9 (2014): 2450-2460.

[40] Khan, Muhammad Zahid, and Ling He. "Neuro-psychopharmacological perspective of Orphan receptors of Rhodopsin (class A) family of G protein-coupled receptors." *Psychopharmacology* 234.8 (2017): 1181-1207.

[41] Summaries for RIMSBP2 Gene. GeneCards, Human Gene Database, www.genecards.org/cgi-bin/carddisp.pl?gene=RIMSBP2&keywords=RIMSBP2. Accessed 22 Apr. 2018

[42] Westra, Harm Jan. *Interpreting disease genetics using functional genomics*. University of Groningen, 2014.

[43] Summaries for TECR Gene. GeneCards, Human Gene Database, www.genecards.org/cgi-bin/carddisp.pl?gene=TECR&keywords=TECR. Accessed 22 Apr. 2018

[44] Summaries for VAC14 Gene. GeneCards, Human Gene Database,.www.genecards.org/cgi-bin/carddisp.pl?gene=VAC14&keywords=VAC14 Accessed 22 Apr. 2018

[45] Saffari, Ayden, et al. "Estimation of a significance threshold for epigenome-wide association studies." *Genetic epidemiology*42.1 (2018): 20-33.

[46] Singh, Indrapal N., et al. "Time course of post-traumatic mitochondrial oxidative damage and dysfunction in a mouse model of focal traumatic brain injury: implications for neuroprotective therapy." *Journal of Cerebral Blood Flow & Metabolism* 26.11 (2006): 1407-1418.

[47] Kaminsky, Z., et al. "A multi-tissue analysis identifies HLA complex group 9 gene methylation differences in bipolar disorder." *Molecular psychiatry* 17.7 (2012): 728.