**Identification and Validation of Cell-free DNA Methylation Biomarker for Pan-cancer Early Screening**

**Research Goal**

The overarching goal of this study is to discover novel non-invasive cell-free DNA methylation-based diagnostic and prognostic biomarkers for human cancers. Low-grade glioma (LGG) and glioblastoma (GBM) are the most frequent types of brain cancer and account for 1.4% of the total new cancer cases and 2.8% of the total cancer deaths in US. Numerous studies have shown that perturbations in methylation profiles both within cancer tissues and in circulating cell-free DNA are recognized signatures of cancer biology (1-6). For glioblastoma patients between the ages of 55 and 64, the 5-year survival rate is 5% (American Cancer Society). Non-invasive early diagnosis for LGG and GBM are urgently needed to improve this survival rate through early interventions and therapies. In my previous study, we showed that tissue-specific DNA methylation fragments (10,000 copies/mL) could be detected in the plasma. Due to the blood-brain barrier and the difficulties to collecting brain samples, research on blood-based non-invasive diagnosis and prognosis tests have been impeded compared with other cancers. In this study, we collected the largest methylation dataset for brain cancer (N=1,196) and non-cancer brain samples (N=1,083) to identify a powerful DNA methylation-based diagnostic and prognostic biomarker panel. Building upon this work, the proposed study is to validate this diagnostic panel in Marshfield Clinic samples. This effort is highly significant and important as diagnostic utility from circulating cell free DNA methylation-based biomarkers will not only help clinicians to diagnose and provide information for disease surveillance, but will also provide important clues to better understand the molecular pathogenesis of LGG and GBM.

**Specific Aim 1:** **To discover DNA methylation-based Pan-cancer biomarkers with optimal diagnostic utility.** In order to identify the most powerful function DNA methylation biomarker for brain cancer research. In this section, we will collect the previous methylation data associated with brain cancer and non-tumor brain methylation with different data types such as microarray data, Genome-wide DNA methylation sequencing data, and reduced representation bisulfite sequencing (RRBS) data, methylated DNA immunoprecipitation sequencing (MeDIP-Seq) and DNA immunoprecipitation sequencing (DIP-Seq) data. Meanwhile, we will also collecte RNA-seq data, Hi-C data as well ChIP-seq data which are conducted to brain cancer or other brain tissues, such as bipolar, dementia, schizophrenia and autism. This dataset will provide a comprehensive biomarker or mechanism research for Marshfield Clinic clinician and scientist. We will identify the powerful functional DNA methylation biomarker applying multiple feature selection and machine learning algorithms.

**Specific Aim 2:** **To validate the high sensitivity, specificity and accuracy of the cell-free DNA methylation panel in Marshfield Clinic brain cancer plasma samples.** In order to validate the identified biomarkers in the first section. We validate the methylation status for our identified biomarkers in the brain cancer and control samples collected in Marshfield or Wisconsin. We will evaluate the cell-free DNA volume and size of the cell-free DNA fragment in the plasma from cancer and control samples. Meanwhile, DNA methylation status of the cell-free DNA methylation biomarkers will be evaluated and the performance of the prediction model will be assessed.

**Specific Aim 3:** **To validate the high sensitivity, specificity and accuracy of the cell-free DNA methylation panel in Marshfield Clinic brain cancer plasma samples.** In order to validate the identified biomarkers in the first section. We validate the methylation status for our identified biomarkers in the brain cancer and control samples collected in Marshfield or Wisconsin. We will evaluate the cell-free DNA volume and size of the cell-free DNA fragment in the plasma from cancer and control samples. Meanwhile, DNA methylation status of the cell-free DNA methylation biomarkers will be evaluated and the performance of the prediction model will be assessed.

**Public Health Relevance**

Approximately 350,000 individuals are diagnoses with a brain cancer in the US annually. Unfortunately, average survival durations have not increased appreciably over the past four decades. GBM is an unusually aggressive disease with extremely low 5-year survival rate. Pre-symptomatic diagnosis of GBM offers clinicians the opportunity to dramatically reduce GBM-related mortality and morbidity. Circulating cell-free DNA methylation signals have been demonstrated to be effective early diagnosis biomarkers for cancers. Currently, the FDA has approved one DNA methylation early screening biomarker for colon cancer (*SEPT9*) and another methylation-based biomarker (*SHOX2*) for lung cancer is currently under FDA evaluation. However, an efficacious DNA methylation biomarker for glioblastoma has not been developed.

**Background**

Brain cancer is the leading cause of cancer deaths in the US among children and young adults (SEER Cancer Statistics Review, Released April 15, 2019). Brain cancer and central nervous system (CNS) tumors are the most diverse cancers which includes 120 different subtypes (National Brain Tumor Society). In the past 30 years, only five treatments targeting to brain cancer have been approved by the FDA. Low-grade glioma (LGG) and glioblastoma (GBM) are the most common types of [primary brain tumor](https://www.mskcc.org/cancer-care/types/brain-tumors-primary). Importantly, LGG and GBM are not independent pathologies as low-grade gliomas may transition into high-grade gliomas over time. Early diagnosis of LGG and GBM brain cancers is an effective approach to increase the 5-year survival rate (7, 8). The ability to detect and treat a low-grade glioma (LGG), before it progresses to GBM, can be the difference between life and death. Currently, the most common diagnostic approaches for glioblastoma are Magnetic Resonance Imaging (MRI), Computerized Tomography (CT) and biopsy; however, the high cost of MRI and CT radiation risk has generated concern from the medical community. Additionally, MRI and CT cannot provide very early detection and extra clinical information such as glioma grading, subtype and prognostic measures and not reliably extractable from image data. Biopsy can provide more information, however, it also is not an early detection method and is an invasive treatment, bringing extra risk for cancer metastasis.

Molecular diagnosis has become one of the most important approaches to provide early and accurate diagnoses with fast speed and low-cost. Compared with other molecular variants such as SNPs, CNV, mRNA and miRNAs, DNA methylation have been demonstrated to be a highly powerful biomarker for cancer diagnosis. Genome-wide DNA hypo-methylation and local hyper-methylation in the promoter region of tumor suppressor genes have been observed for almost all cancer types (9, 10) and these abnormal changes have been observed earlier than most symptoms.(11, 12) DNA-methylation also has clinically beneficial features in comparison to MRI or CT(13-15). In my previous research and the work of others, DNA methylation was found to silence tumor suppressor genes(16, 17), miRNAs (18), mRNAs(19) and drug metabolic genes (20) to play important roles in cancer development(21), metastasis (22) and chemotherapy resistance(23). Recently, we demonstrated that circulating cell-free DNA methylation could provide a novel approach to help Clinicians diagnose or predict early stage cancers in a non-invasive way (1). Additionally, compared with DNA mutations, DNA methylation have more options/biomarkers to be selected for diagnosis and prognosis. According to recent research, abnormal DNA methylation sites in individual cancer patients could number up to 103-104, while DNA mutations (driver mutation and passenger mutation) number only 10-102.It is therefore reasonable to assume that the majority of gene expression changes are caused by DNA methylation rather than mutations. Further, mutations and DNA methylation could work together to silence certain genes, such as *TP53* (24).

In the past, DNA methylation research in glioblastoma was very limited and mainly focused on several identified genes, such as *MGMT* (25, 26), *CD133* (27), *ARF1* (28), and primarily applied to prognosis. Several non-coding RNAs were also reported to be abnormal in GBM, such as miR-153(29), miR-181(30). Although there have been several genome-wide DNA methylation studies to identify differential DNA methylation in GBM (31-34), limited sample sizes made it difficult to make solid conclusions (31-37) and normal controls were limited. Moreover, there are only few studies that investigated DNA methylation biomarkers for LGG or GBM based on circulating cell-free DNA. We only found six papers (38-43) in Pubmed with the [keywords](https://www.ncbi.nlm.nih.gov/pubmed/?term=((cell-free+or+circulating)+AND+glioblastoma)+AND+methylation) of cell-free or circulating, methylation and GBM/LGG in all the field. Recently, Giselle and colleagues found circulating cell-free DNA could be a prognostic and molecular marker for brain tumors under Perillyl Alcohol-based therapy (44), thereby indicating DNA fragments from the brain are released to blood (across the blood-brain barrier) and could be detected in the plasma.

In this research, in order to identify the most powerful diagnostic and prognostic biomarkers for brain cancer, we amassed the largest genome-wide DNA methylation dataset using publicly-available data which includes 516 LGG, 673 GBM and 1,339 non-cancer brain samples. We also collected 516 lower grade glioma (LGG) and 1,339 normal PBMC genome-wide DNA methylation data to increase the power to identify GBM biomarkers. Among these brain cancer samples, 514 LGG and 155 GBM samples has detailed survival information (overall survival time), thereby enabling the evaluation of DNA methylation biomarkers in survival analyses. Furthermore, RNA-seq data were also available for 467 LGG and 382 GBM samples to help us understand the pathogenic mechanisms of LGG and GBM from the combined regulatory effects of DNA methylation and mRNA. With the help of the largest brain cancer genome-wide DNA methylation dataset, we can identify all the hyper-methylated genes which exhibit low-expression in glioblastoma or hypo-methylated genes with high-expression in glioblastoma. We also downloaded genome-wide DNA methylation data on 1,198 normal PBMC samples which can be used as background in a subtraction analysis to identify hyper-methylated DNA fragments which are non-methylated for white blood cells so that we can obtain signals from the hyper-methylation biomarkers for glioblastoma in the blood and without interference from GBM white blood cell DNA methylation signals. We will validate our diagnosis and prognosis methylation panels in Marshfield Clinic brain cancer samples (100 biopsy samples and 30 cell-free DNA from patient plasma).

Dr. Shicheng Guo, Dr. Schrodi and Dr. xx have recently completed several DNA methylation research and one of the manuscript was completed in which genome-wide DNA methylation sequencing to cell-free DNA collected from hepatitis, cirrhosis, early liver cancer and advanced liver cancer patients. The data showed cell-free DNA methylation signals in the plasma could provide a powerful biomarker for HCC diagnosis. Dr. Guo has published 20 SCI papers in DNA methylation research and have comprehensive DNA methylation research skills both in bench work and computational analysis. Dr. Schrodi have comprehensive statistical skills and will provided mythological supports in this project. We have confirmed with Dr. , Marshfield Clinic have stored more than 100 brain cancer biopsy samples which will provided the brain cancer tissues to validate the microarray data. Dr. xx will help us to enroll and collected 20-30 fresh blood from brain cancer and control individuals for the cell-free DNA methylation validation. In our study, we identified 5 interesting prognosis (overall-survival time) related DNA methylation biomarkers, validation of these biomarker will be not the priority works since it might require long time to wait. We will only validate the prognosis biomarkers in biopsy brain cancer samples since the survival time for these samples are available. Meanwhile, we will compare these biomarkers in cell-free DNA with brain cancer progress indicators, such as brain cancer stage (I, II, III, and IV) which will help us to evaluate the indirect performance of the prognosis prediction model.

**Preliminary Studies**

**Genome-wide DNA methylation of glioblastoma revealed numerous DNA methylation diagnostic biomarkers**

In order to identify DNA methylation biomarker for GBM diagnosis and prognosis, we collected genome-wide DNA methylation data from public database including GEO(45) and Arrayexpress(46). We found 15 public dataset are involved in genome-wide DNA methylation research (composed by methylation 27K, 450K and 850K microarray) in brain and brain cancer research (**Figure 1A**). These dataset include 1,379 brain cancer samples and 1,339 non-cancer brain. We showed the age and gender distribution among different sample types in **Figure 1B and 1C.** In order to design precision sequencing primers for target bisulfite sequencing, we collected normal brain genome-wide bisulfite sequencing data for human brain from Human Roadmap Project (47). In order to make our research more solid, we also collected other cancer samples so that we can evaluated the performance of the markers in other cancers. We collected NCI-60 cancer cell line methylome data based on HM450K array were downloaded from CellMiner database(48) and the cancer genome atlas (TCGA) DNA methylation data for 23 cancers (49).

A B C

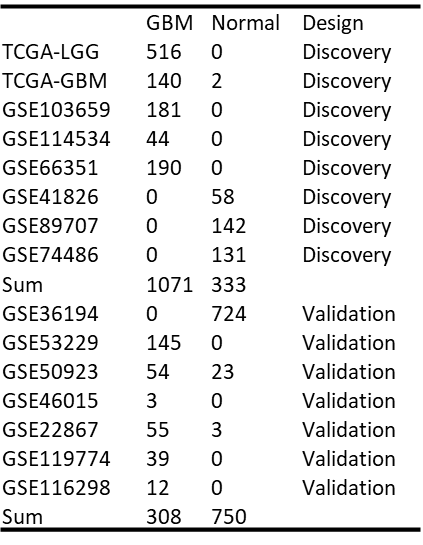
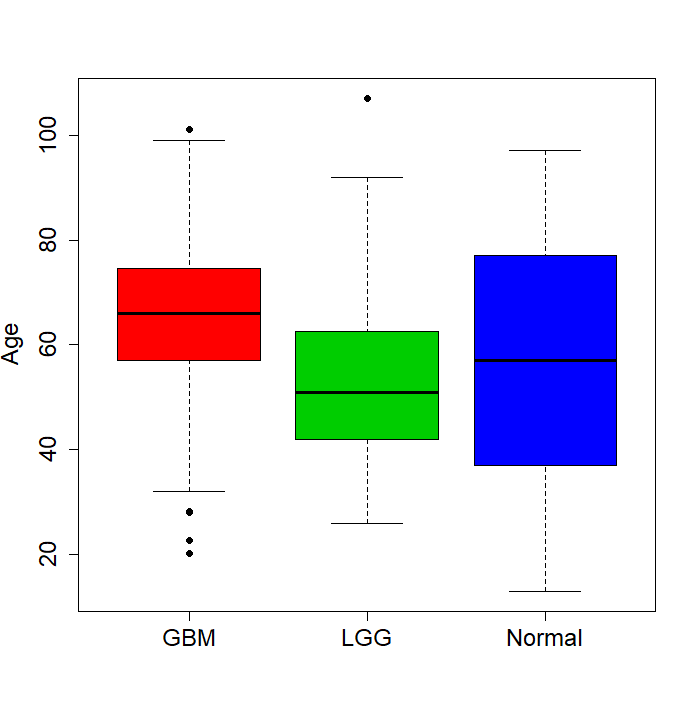
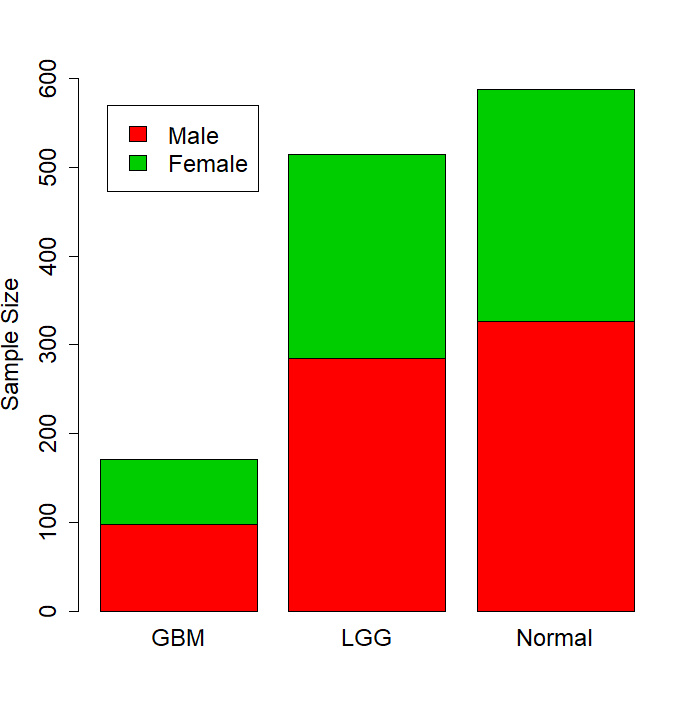
  

Figure 1. Dataset and clinical characteristic for MCRI LGG/GBM methylation biomarker identification project. A: sample size for each public dataset. B: age distribution for GBM/LGG and normal. C: gender distribution.

We plan to apply three stage research strategy to identify and evaluate the diagnostic biomarker for brain cancer including LGG and GBM (**Figure 2**). In the discovery stage, we integrate all the methylation 450K microarray data (including 516 LGG, 365 GBM and 589 non-tumor brain tissues) to identify potential diagnostic biomarkers. PCA analysis were applied to identify and to remove non-brain samples. We found 21 samples are actually blood samples which were confirmed by GEO clinical information and we also removed another 9 samples which are suspicious (**Figure 2**). For the discovery dataset, we have detailed clinical characteristics including age, gender. We collected these confounders and we found the samples are composed by 709 male, 565 female samples and 196 gender-missing samples. However, for the validation dataset (HM27K and 850K), clinical information are missing in GEO database. In order to identify the potential biomarkers which can be applied in non-invasive cell-free DNA (cfDNA) methylation based brain cancer diagnosis with plasma samples, we require the DNA methylation biomarker have distinct DNA methylation status with blood cells. We then adopted human PBMC, CD4+ and CD8+ T cell DNA methylation from our previous research (50-52) to be background noise control. In addition, cell-free DNA methylation data status of 75 normal samples were retracted from my previous cell-free DNA methylation research to increase the biomarker specificity (1). We removed all the probes which contain single nucleotide polymorphisms (SNPs) to avoid the influence from genetic variation in the population and differential hybridization between probe and DNA molecular. As we known, DNA methylation mainly played roles in gene expression regulation, we aim to identify function DNA methylation markers and therefore we collected transcriptional factor binding sites (TFBS), H3K4me1, H3K4me3, H3K27ac and DNase I hyper-sensitive site (DHS) data from ENCODE project (53). KEGG pathway and Reactome pathway database were downloaded and were applied for gene functional prediction. In our project, we set higher priority to tumor suppressor genes since we hope to identify hyper-methylation signals in human plasma. Suppose these functional are abnormal hyper-methylated in cancer cells and therefore the roles for these genes are belong to tumor suppressor genes. Epigenetics genes are another groups genes of our interest since current lots of epigenetics factors are targeted for cancer therapy. In the second stage of our study design is to validate the candidate biomarkers identified in the first stage (methylation 450K) in another independent dataset. This dataset are composed by 308 GBM and 750 non-cancer brain samples.

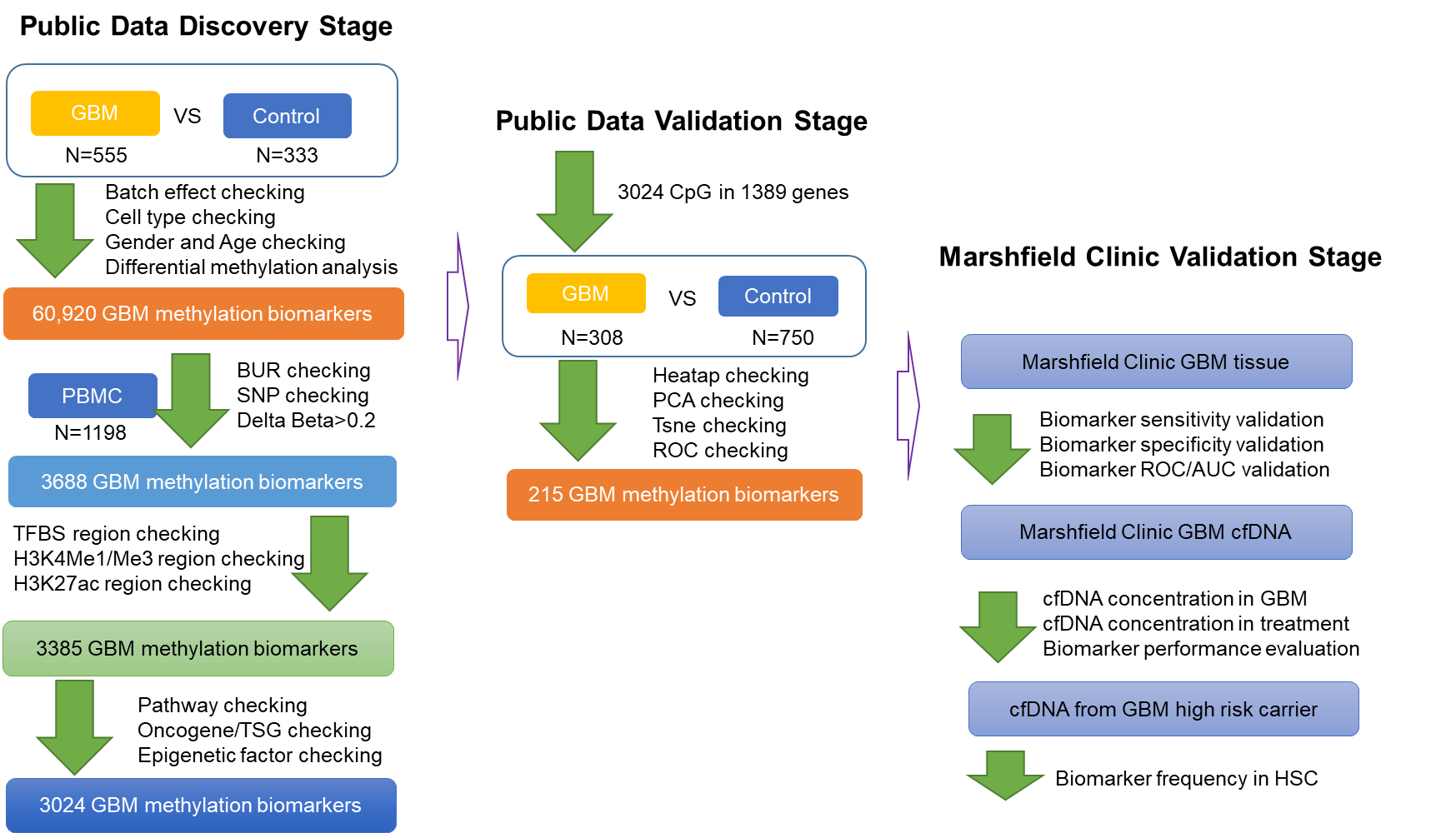


Figure 2. Flowchart of DNA methylation based biomarker for GBM. We designed a three-stage biomarker identification strategy to identify brain cancer (LGG and GBM) diagnostic and prognostic biomarker systems. We first scan the whole human methylome to obtain the most potential functional DNA methylation biomarkers and these makers were evaluated by another independent dataset between GBM and non-tumor brain samples and apply multiple machine learning method to form the best diagnosis and prognosis biomarker panel. Finally, we will evaluate the panel in MRCI samples including solid tumor, cell-free DNA from LGG/GBM cancer patient.

Methylation profiles for these samples are based on Illumina methylation 27K microarray.

We check the integrated DNA methylation dataset with principle component analysis and to evaluate clinical confounders in the whole dataset. We found gender and dataset don’t provided strong variations between LGG/GBM and normal samples (**Figure 3A and 3C**) while the major variation came from sample types including LGG, GBM and control brain samples(**Figure 3B**). The PC1 and PC2 contributed almost 80% variations. We apply logistic regression to identify LGG and GBM differential methylation loci adjusted by age and gender: (LGG/GBM ~ CpGs+ Age+ Gender) and the methylation status of 60, 920 CpGs are significantly different compared with non-tumor brain tissues.

A B C

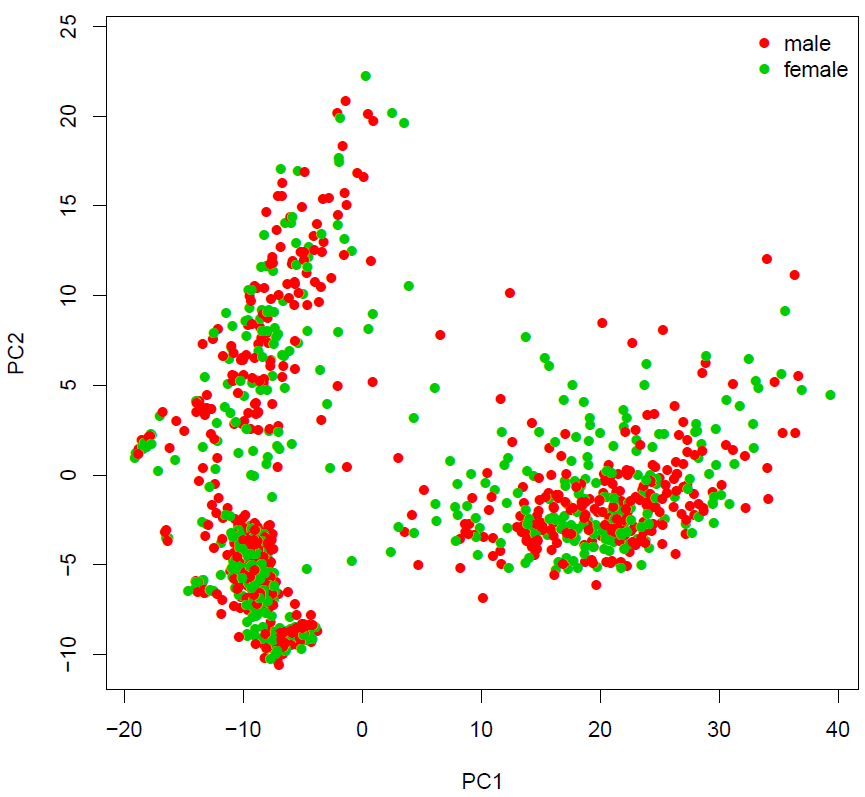
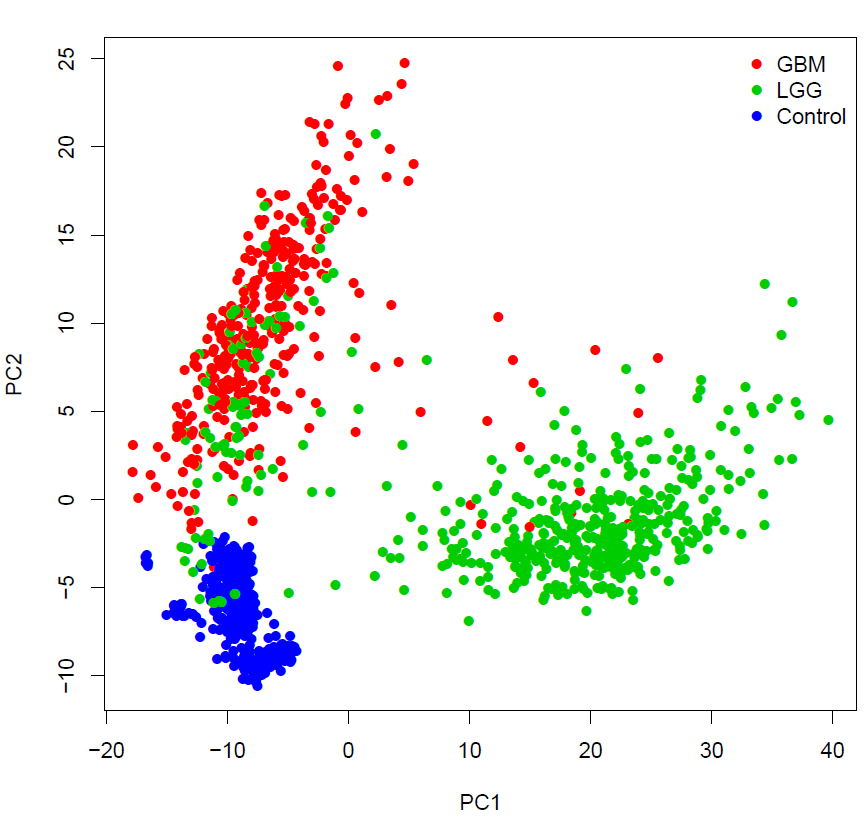
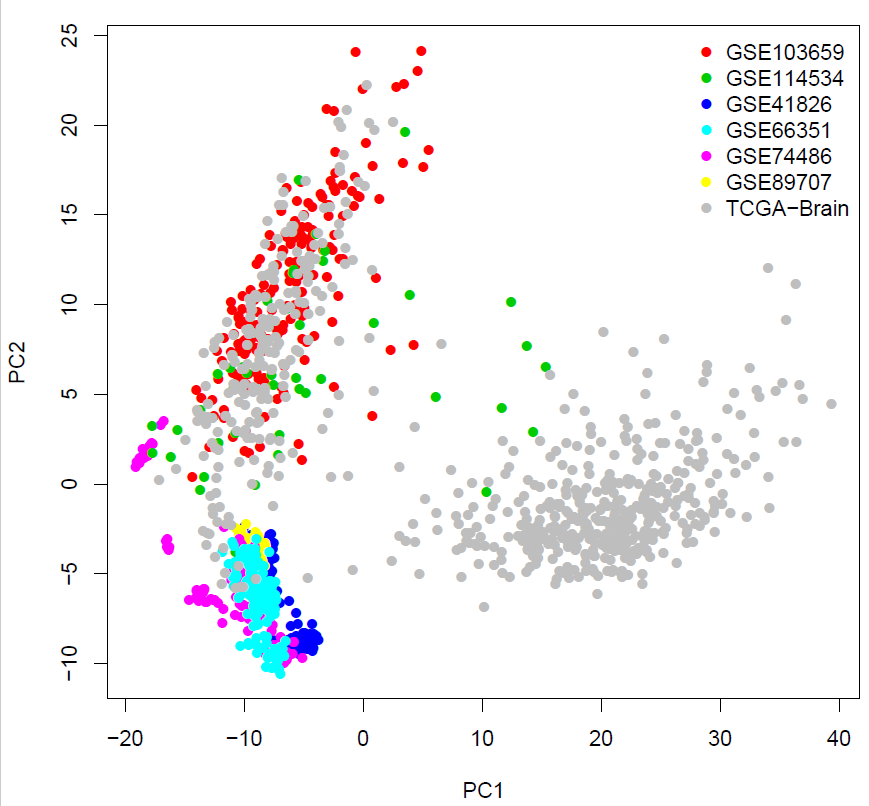


Figure 3. PCA analysis to discovery dataset to evaluate the data structure. In the discovery stage, we have 7 dataset which are merged by methylation 450K microarray (HM450K). HM450K includes 450,000 probes and provided genome-wide assay to human methylation profile.

When we filter with 1) blood un-methylated loci (BUL) from 1,198 PBMC samples 2) remove probes with SNPs and 3) methylation change (delta beta) > 0.2 and decrease the candidate biomarkers to 3,688. Actually, if we apply higher delta beta as the threshold, we can disease the candidate markers to 285 (delta beta=0.4, **Figure 4A**). With the functional genomic region as the filter, we decease the biomarker to 3,385 when we require the biomarker should be annotated by all the ENCODE regulatory elements (N=6, include TFBS, H3K4me1, H3K4me3, H3K27ac, DHS and BUR of PBMC).

We applied random forest for feature selection to check the distribution of importance for each biomarker. In this analysis, we conducted the feature selection evaluation in two ways: 1) binary prediction, in which, the samples were divided into cancer (LGG and GBM) and non-cancer samples. 2) multi-class prediction, in which, the samples were divided into LGG, GBM and non-cancer samples. We found both of the prediction performance are perfect. In the first scenario, the sensitivity and specificity estimated in out-of-bag (OOB) of the prediction model are 99.83% and 99.3% and the error rate of the prediction is only 0.48%. Within the model, 157 CpGs shown Mean-Decrease-Gini >1 which indicating have powerful prediction performance. In the second scenario, only 1 normal samples were mistakenly classified to LGG and the remained normal were all predicted to normal, indicating the specificity is as high as 99.83%. 94.5% GBM and 83% LGG were classified to its own group while 79 LGG were predicted to GBM while 20 GBM were predicted to LGG.

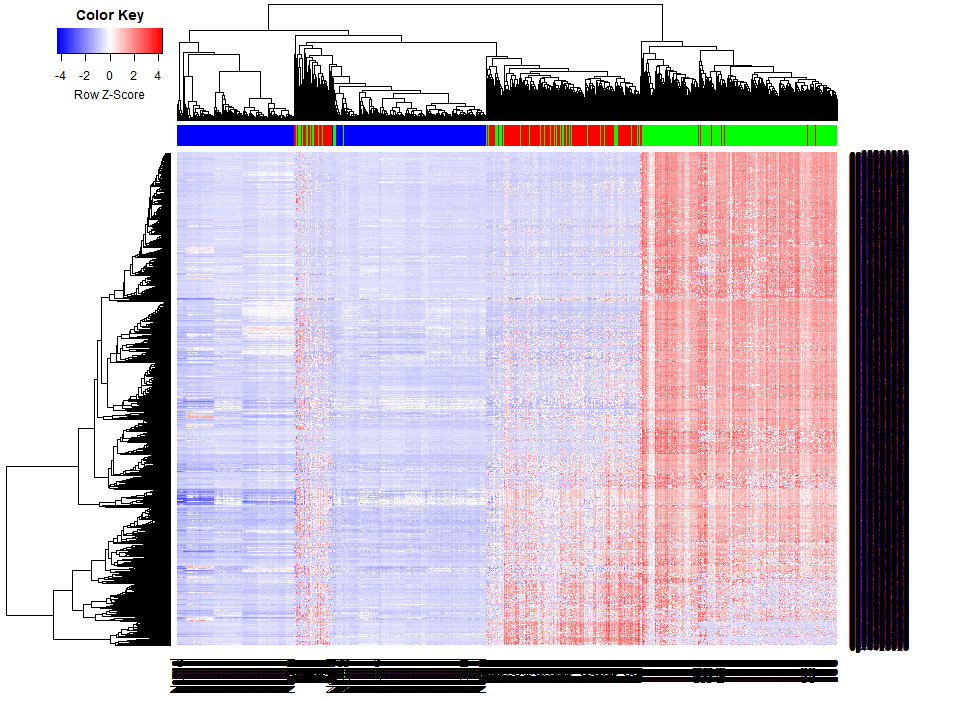
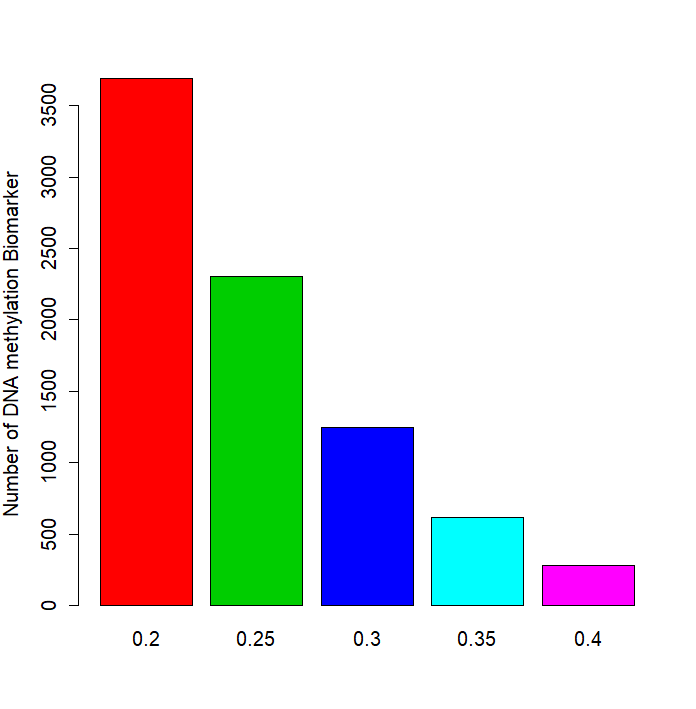


Figure 4. Final methylation biomarker and Cluster analysis to reveal diagnostic panel performance. With multiple filters, we eventually targeted 3,024 DNA methylation biomarkers in the discovery stage. This number can be decreased with higher threshold of effect size (ES) to methylation change between tumor and non-tumor samples. For example, when we set ES=0.4, we will have 285 markers remained and the cluster performance isn’t significantly weakened.

In order to select a best biomarker panel which can be applied in Marshfield Clinic samples, we need to evaluate the performance of DNA methylation biomarkers in GBM and LGG prediction with limited biomarkers. With random forest method, we found the prediction error rate is decreased as the increment of biomarker numbers. In the binary prediction scenario, the prediction accuracy could come up to 98% with ~3 biomarkers (**Figure 5A**) while in the 3-class classification scenario, the prediction accuracy could come up to 91% with ~ 5 biomarkers (**Figure 5B**). These result are consistent with our previous research in esophageal cancer (54) and lung cancer (55) in which we shown 3-5 methylation biomarker could provide well cancer prediction performance, usually AUC come up to 0.90 which is as good as traditional diagnosis method.

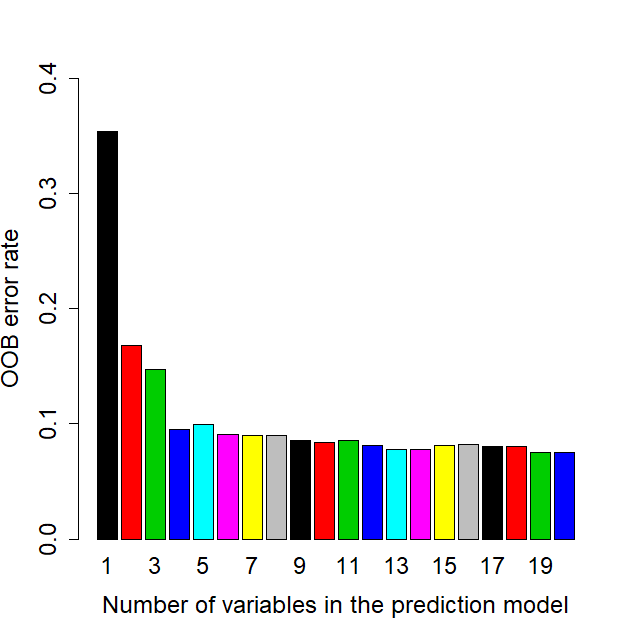
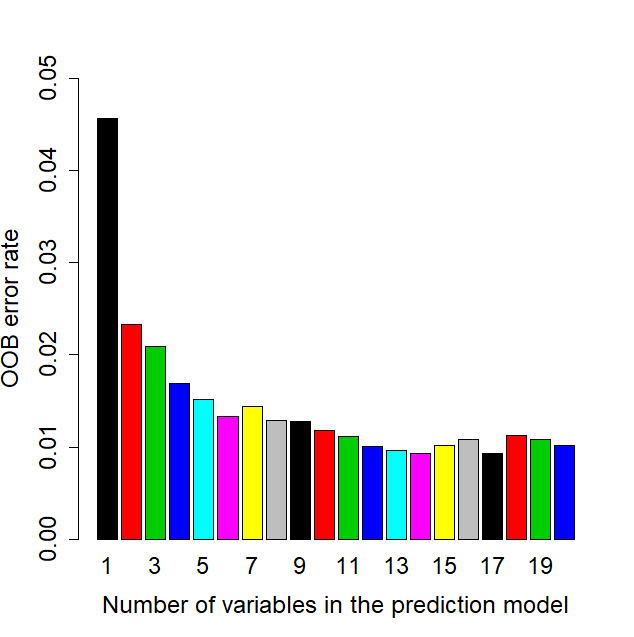
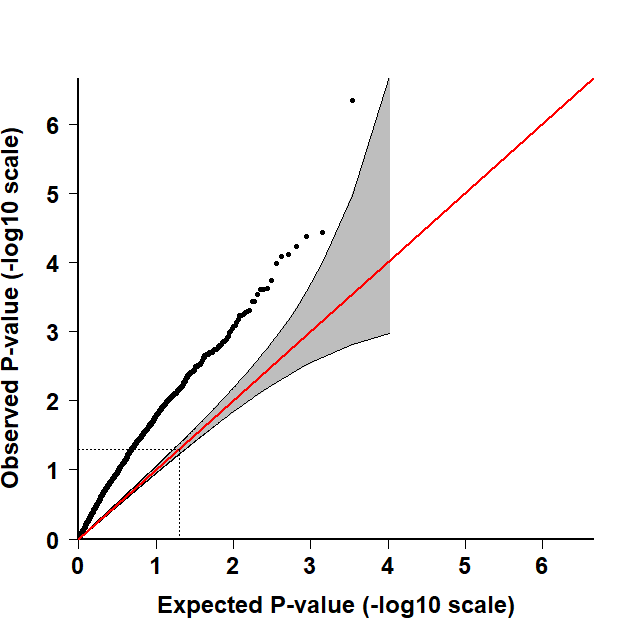
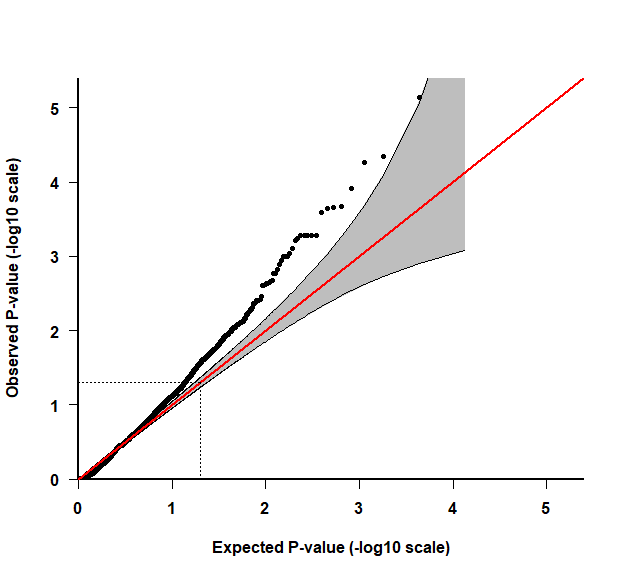


Figure 5. DNA methylation biomarker performance in random forest prediction models.

**Genome-wide DNA methylation of glioblastoma identified DNA methylation prognostic biomarkers**

DNA methylation have been demonstrated to be significantly associated with outcomes of human cancer. In order to identify prognosis biomarkers, I collected all the TCGA LGG and GBM methylation dataset which have comprehensive overall survival times (OS) so that we can identify the most interesting prognostic methylation biomarker for GBM and LGG. The overall survival time (OS) for 155 GBM and 514 LGG. With Cox-regression analysis, we identified 5 genes including NEUROD1 (cg01431993), SCHIP1 (cg01654862 and cg05580655), NEUROD1 (cg20709008), ARHGEF7 (cg20749916) and CDYL (cg25026237) shown significant prognostic biomarker for GBM or LGG outcome (Figure 6A and Figure 6B).

A (LGG) B (GBM)

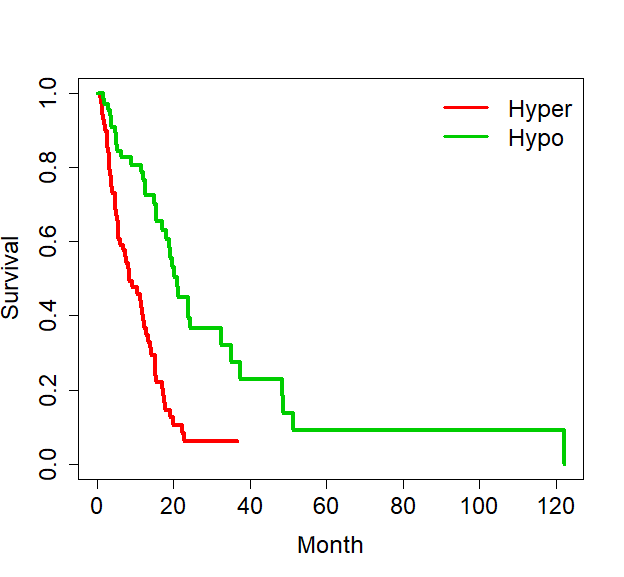
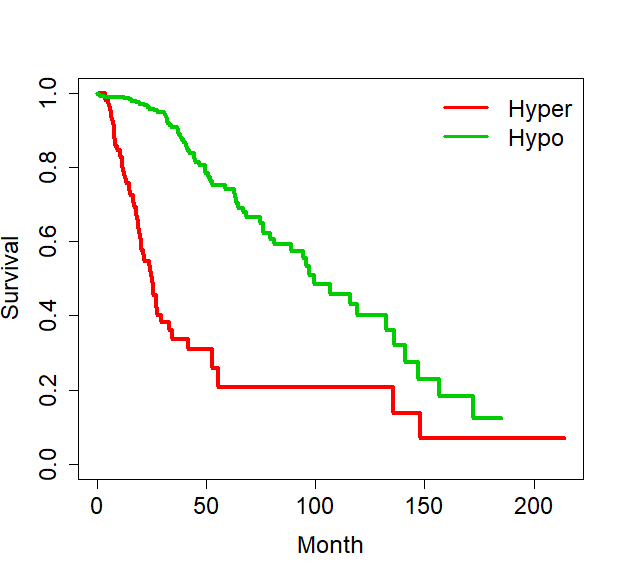
CDYL

ARHGEF7

NEUROD1

SCHIP1

Figure 6. QQ-plot for Cox-regression based survival analysis between DNA methylation and overall survival time.

****

cg20749916 (ARHGEF7) in LLG

**P<0.000001**

**P<0.000001**

cg25026237 (CDYL) in GBM

Figure 7. DNA methylation status revealed by Roadmap and GEO based WGBS data

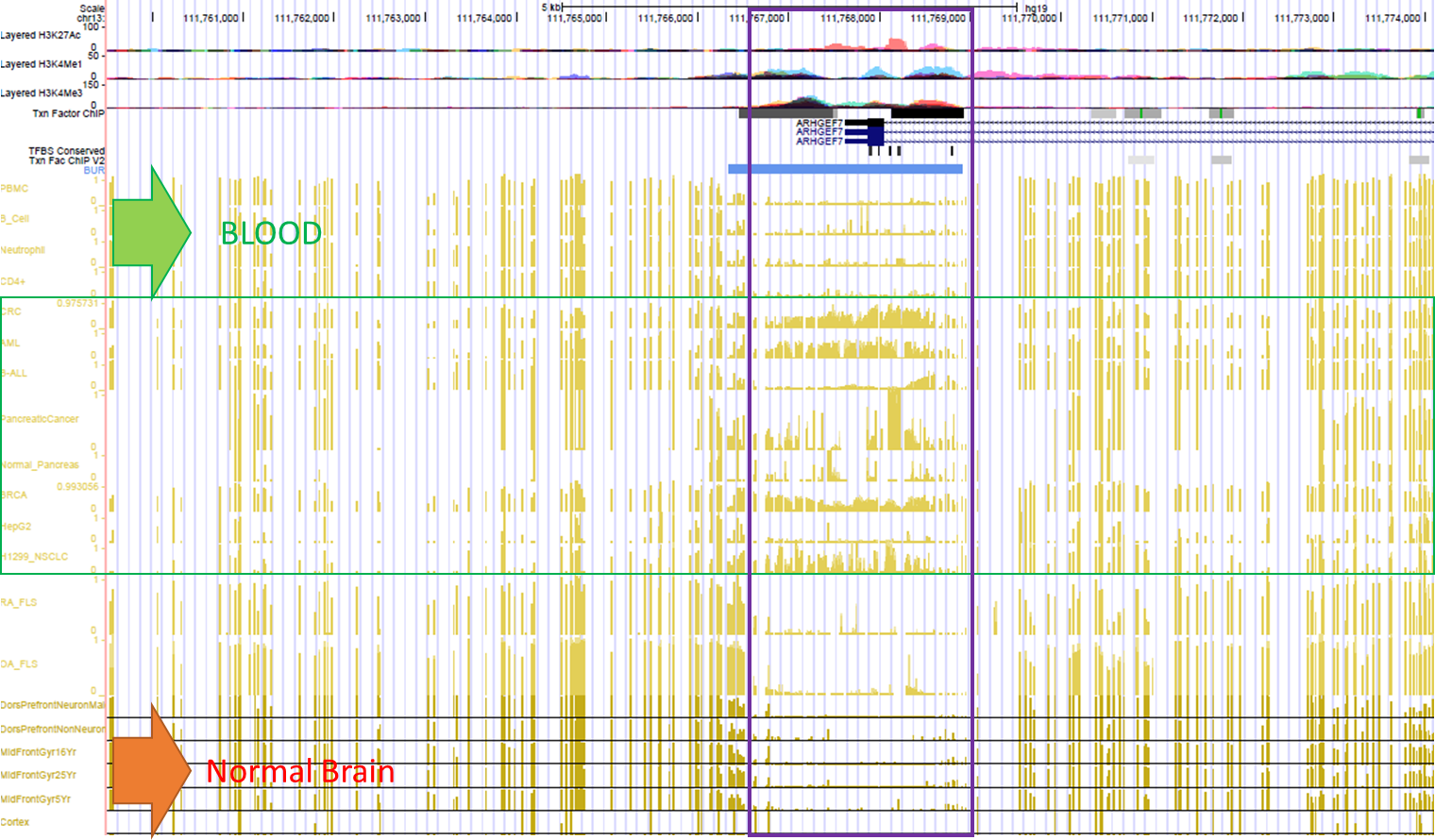


Figure 8. DNA methylation profiles of ARHGEF7 in human normal brain and other human normal or cancer tissues. We take ARHGEF7 as the example to show the hyper-methylated status of the promoter regions in human tissues (normal brain, normal blood and some other cancer tissues)

We checked some of the candidate DMRs and we found these biomarkers also shown hyper-methylated status in other human cancers such as liver cancer (HCC), breast cancer (BRCA), blood cancers (AML) as well pancreatic cancers while low-methylated in normal blood and normal human brain tissues. These evidences indicate the cell-free DNA methylation biomarker panel might have multiple roles in cancer diagnosis.

Overall, in this study, we identified 14 diagnostic biomarkers including SPI1, NFAM1, OSM, BIN2, ITGB2, IRF4, KLF4, MEF2C, TAL1, ZEB2, ZNF662, MTMG, EZH2, SOX11 and 5 prognostic methylation biomarkers including NEUROD1, SCHIP1, NEUROD1, ARHGEF7 and CDYL. We plan to validate these 16 biomarkers in about 100 brain cancer tissues and 30 cell-free DNA fragment collected from Marshfield Clinic to provide the final validation to these biomarkers.

**Power analysis and sample size estimation**

In order to estimate the minimum required sample size for the validation, we applied bootstrap resampling (N=1000) to evaluate the power under different sample size. For each single biomarker, we apply Bayesian Generalized Linear Models (BGLM) adjusted with age and gender to estimate the biomarker difference between cancer and non-cancer samples (A). We didn’t control the ratio between LGG and GBM so that we can receive generalized estimation. With the same strategy, we conducted the power analysis to all the biomarker and other potential biomarker without overlapped between 450K and 27K (which are excellent biomarker however didn’t occurred in the microarray in the phase 2 dataset). Take TEC (cg07849581) as the example, 22 samples could support the power for single biomarker (a=0.05) and 60 samples are enough to support for as many as 50 biomarkers adjusted with multiple test correction (q=0.05). We conducted power analysis to each biomarker one by one and we found 13 samples could provide 0.8 power for all the biomarkers and some of biomarker only require 10 samples to receive power of 0.8.

A B

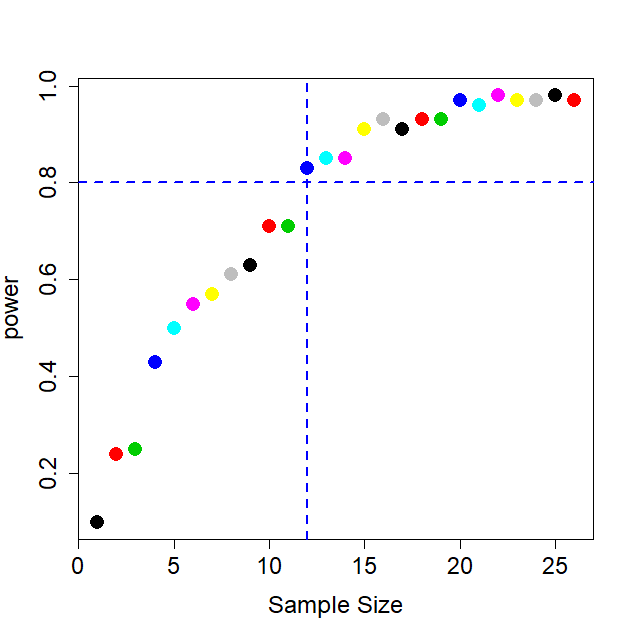
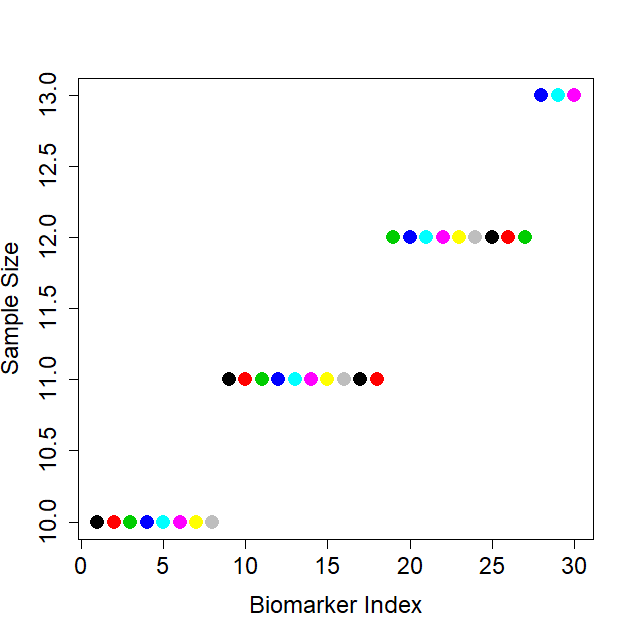
 

Figure 9. Power analysis to estimate the minimum sample size for the validation study. A) Power analysis for single biomarker (cg07849581). B) Power analysis to all the biomarkers to find the minimum sample size to receive the power of 0.8.

**Research Design and Methodology**

**Identification of DNA methylation based GBM/LGG diagnostic and prognostic biomarkers**

We designed a three-stage strategy to identify the most powerful DNA methylation based GBM/LGG diagnostic and prognostic biomarkers with comprehensive public dataset including GEO/Arrayexpress dataset, TCGA dataset, Roadmap Epigenomic dataset, NCI-60 DNA methylation dataset, ENCODE functional genome regulatory dataset. We also plan to collect some other related public dataset to provide more support to our final conclusion, including GSE121721, GSM941746, GSM669614, GSM669615, GSM669613, GSM669604, GSM669605 and GSM669603.

**Patient enrollment, blood collection, cell-free DNA extraction and bisulfite conversion**

In this project, we will enroll patients diagnosed as LGG and GBM. In order to decrease the noise and false negative, other brain cancer subtype will be excluded. According to our power analysis, we will collect ~30 LGG/GBM solid cancer tissues which as we known are existed in Marshfield Clinic biobank. For the cell-free DNA from the LGG/GBM patients, we will enrolled within Marshfield Clinic working together with Dr. xx. Meanwhile, equal fresh normal plasma will be collected with gender and age matching.

5ml of whole blood was collected in EDTA-collection tubes (EDTA Monovettes, Sarstedt, Germany) and was immediately centrifuged at 1600×g and 4 °C for 10 min. Plasma (2.5ml, ~55% of the whole blood) was transferred to a fresh tube followed by a second 10 min of centrifugation at 16,000×g and 4 °C. Finally, plasma was passed through a 0.8 μm filter and aliquots were stored at −80 °C until further analysis. DNA will be extracted using a modified viral DNA/RNA extraction kit (QIAamp Circulating Nucleic Acid Kit, Qiagen, Hilden, Germany). Plasma samples were thawed at room temperature and extracted following the directions of the kit with modifications. Samples were lysed and treated with protease at 56°C for 10 min in a 50 mL Falcon tube. 100 μL of magnetic particles and 15 mL of binding buffer were then added, and binding was performed for 60 min at room temperature on a rotator. Magnetic particles were captured for 4 min, the supernatant discarded and the pellet was resuspended in 3 mL of wash buffer. 1.5 mL of partaicle solution were transferred to a 2 mL SafeLock, the beads captured and the supernatant discarded. This was repeated to complete the 3 mL transfer. Tubes were briefly centrifuged and the residual wash buffer was removed by pipetting after bead separation. The tubes were then placed in a 56°C dry block for 5 min, 100 μl of elution buffer was added, the tubes incubated at 65°C with shaking on a thermomixer for 15 min, the particles separated on a magnetic stand and the eluted DNA transferred to a 0.5 mL SafeLock tube (Eppendorf). A 5μl aliquot of the DNA sample was transferred to 45 μl of elution buffer for the measurement of genomic DNA.

95-100ul of extracted DNA in elution buffer will be applied for bisulfite treatment. The bisulfite reagents (for 25 reactions) were prepared as follows: Sodium bisulfite (4.71gm) and sodium sulfite (1.13gm) were dissolved in 10mL of ddH2O in a falcon tube, by vigorous shaking and heating to 50°C, and the adjust pH to 5.4-5.5 with 0.2M NaOH. 188 mg of 6-hydroxy-2, 5, 7, 8-tetramethyl-chroman-2-carboxylic acid was dissolved in 1.5 mL diethyleneglycoldimethylether (DME), vortexing to ensure an uniform solution. 190 μL of bisulfite solution and 30μl DME-radical scavenger solution were added to the 95-100 ul DNA sample in 0.5 mL SafeLock tubes. The tubes were incubated in a Eppendorf Mastercycler (Eppendorf) according to the following protocol: 5 min 99°C, 25min 50°C, 5 min 99°C, 1h 25min 50°C, 5 min 99°C, 4h 55min 50°C, hold 20°C. This approach allowed overnight bisulfite conversion.

Following bisulfite conversion, DNA was purified. The bisulfite reaction (320ul) was transferred to a 2 mL tube, and 1 μL of polyA (500 ng/ul) and 1.5 mL of binding buffer were added.10ul chemagen magnetic particles were added and the sample was mixed by vortexing. The samples were incubated at room temperature on a thermal mixer at a rotation of 1000 rpm for 60 min. Magnetic particles were separated on a magnetic stand, the liquid discarded, the tubes briefly centrifuged and the residual liquid removed following magnetic separation. The particles were washed twice with wash buffer II from the kit, and once with 70% ethanol. Following the ethanol wash, the tubes were centrifuged again, and the residual liquid removed following magnetic separation. The particles were dried by placing open tubes in a 55°C heat block, and 55 μL elution buffer (10 mM Tris pH 7.2) added. Samples were incubated at 55°C for 15 min on a thermal mixer with rotation set to 1000 rpm, placed on a magnetic separator and the eluate containing DNA transferred to a new tube. A 5 µL aliquot of bis-DNA was added to 45 µL of elution buffer for the measurement of a 10 fold diluted sample. This purification method leaves out the desulfonation step following bisulfite conversion, to make the DNA amenable to carry over prevention by UNGase treatment. PCR amplification of sulfonated DNA requires an extended activation time to allow desulfonation to occur prior to amplification.

Positive controls for DNA extraction were 25ng/mL methylated DNA diluted in 5mg/mL bovine serum albumin (BSA), while negative extraction controls were BSA without spiked DNA. Positive controls for bisulfite conversion were composed of 10ng fully methylated DNA spiked into 90 ng of human genomic DNA prepared from buffy coat cells (Roche Applied Sciences) in a 100 µL elution buffer. Negative bisulfite conversion controls were composed of elution buffer alone

**Liquid biopsy tests and DNA methylation measurement to solid tissue and cell-free DNA**

DNA methylation measurement have several different steps including DNA extraction (see above), DNA bisulfite conversion (see above) and targeted-PCR based methylation sequencing (NGS). We already have parts of GBM/LGG solid tissues and cell-free DNA from tumor patients can be enrolled with different disease stages. Parts of cell-free DNA can be extracted from MCRI personalized medicine research project (PMRP). DNA extraction is basic molecular biology experiments. DNA methylation bisulfite conversion is routine molecular assay which can be completed with EpiTect Bisulfite Kit (QIAGEN). We will apply target bisulfite sequencing (TBS) to targeted sequencing the biomarker regions identified in our Aim 1. We have applied target bisulfite sequencing in our previous research for multiple times in esophageal cancers (16, 54). First, we will design BSP primers for our biomarkers identified in the Aim 1 and then provided the barcode for each samples and then pooled them into one single sequencing library. Theoretically, one MiSeq lane could provide the target methylation sequencing for 15 biomarkers in 384 samples with 5,000x sequencing depth with 150bp reads.

Study Quality Control

**Machine learning and multi-class classification**

We will apply multiple classification/machine learning algorithm to build the diagnostic and prognostic models based on solid tissue, cfDNA methylation from cancer patients and high-risk allele carrier. We will prepare a package for the multiple machine learning algorithms including random forest (RF), logistic regression (LR), support vector machine (SVM), Bayes tree (BT), K-nearest neighbors (KNN). According to my previous research, these method do have certain different performance in same dataset causing by the preferred data structure and characteristics. We will evaluate the performance of the different methods in GBM/LGG methylation dataset and provided a reference for the further methylation biomarker research. Five-fold cross-validation will be applied in the solid tissue validation since the sample size is enough for the model training and testing. However, in the cell-free DNA methylation validation stage, we will apply logistic regression based ROC analysis to show the prediction performance since the sample size is not large enough for the cross-validation (N=20-30). Again, current proposal will be used to collect preliminary data for external grant application. We will continue the research with larger sample size.

**Literature Cited**

1. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 2017;49:635-642.

2. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 2018;563:579-583.

3. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9:5068.

4. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet 2011;43:768-775.

5. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 2009;41:1350-1353.

6. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, Yi S, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater 2017;16:1155-1161.

7. Chu TPC, Shah A, Walker D, Coleman MP. Where are the opportunities for an earlier diagnosis of primary intracranial tumours in children and young adults? Eur J Paediatr Neurol 2017;21:388-395.

8. Fry CW, Perrow R, Paul SP. Brain tumours in children: importance of early identification. Br J Nurs 2014;23:1202-1207.

9. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301:89-92.

10. Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. Biochem Biophys Res Commun 1983;111:47-54.

11. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glockner S, et al. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 2008;358:1118-1128.

12. Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, Kozuki T, et al. Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res 2005;11:1219-1225.

13. Han L, Kamdar MR. MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. Pac Symp Biocomput 2018;23:331-342.

14. Wei J, Yang G, Hao X, Gu D, Tan Y, Wang X, Dong D, et al. A multi-sequence and habitat-based MRI radiomics signature for preoperative prediction of MGMT promoter methylation in astrocytomas with prognostic implication. Eur Radiol 2019;29:877-888.

15. Jiang S, Rui Q, Wang Y, Heo HY, Zou T, Yu H, Zhang Y, et al. Discriminating MGMT promoter methylation status in patients with glioblastoma employing amide proton transfer-weighted MRI metrics. Eur Radiol 2018;28:2115-2123.

16. Jiang D, He Z, Wang C, Zhou Y, Li F, Pu W, Zhang X, et al. Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma. Cell Death Dis 2018;10:1.

17. Lei Y, Liu L, Zhang S, Guo S, Li X, Wang J, Su B, et al. Hdac7 promotes lung tumorigenesis by inhibiting Stat3 activation. Mol Cancer 2017;16:170.

18. He Y, Cui Y, Wang W, Gu J, Guo S, Ma K, Luo X. Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma. Neoplasia 2011;13:841-853.

19. Lan Q, Liu PY, Haase J, Bell JL, Huttelmaier S, Liu T. The Critical Role of RNA m(6)A Methylation in Cancer. Cancer Res 2019.

20. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005;352:997-1003.

21. Chae YC, Kim JY, Park JW, Kim KB, Oh H, Lee KH, Seo SB. FOXO1 degradation via G9a-mediated methylation promotes cell proliferation in colon cancer. Nucleic Acids Res 2019;47:1692-1705.

22. Oshima G, Poli EC, Bolt MJ, Chlenski A, Forde M, Jutzy JMS, Biyani N, et al. DNA Methylation Controls Metastasis-Suppressive 14q32-Encoded miRNAs. Cancer Res 2019;79:650-662.

23. Chaudhry P, Srinivasan R, Patel FD. Utility of gene promoter methylation in prediction of response to platinum-based chemotherapy in epithelial ovarian cancer (EOC). Cancer Invest 2009;27:877-884.

24. Agirre X, Vizmanos JL, Calasanz MJ, Garcia-Delgado M, Larrayoz MJ, Novo FJ. Methylation of CpG dinucleotides and/or CCWGG motifs at the promoter of TP53 correlates with decreased gene expression in a subset of acute lymphoblastic leukemia patients. Oncogene 2003;22:1070-1072.

25. Weller M, Tabatabai G, Kastner B, Felsberg J, Steinbach JP, Wick A, Schnell O, et al. MGMT Promoter Methylation Is a Strong Prognostic Biomarker for Benefit from Dose-Intensified Temozolomide Rechallenge in Progressive Glioblastoma: The DIRECTOR Trial. Clin Cancer Res 2015;21:2057-2064.

26. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, Ostermann S, Otten P, et al. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. Clin Cancer Res 2004;10:1871-1874.

27. Yi JM, Tsai HC, Glockner SC, Lin S, Ohm JE, Easwaran H, James CD, et al. Abnormal DNA methylation of CD133 in colorectal and glioblastoma tumors. Cancer Res 2008;68:8094-8103.

28. Lopez-Gines C, Navarro L, Munoz-Hidalgo L, Buso E, Morales JM, Gil-Benso R, Gregori-Romero M, et al. Association between epidermal growth factor receptor amplification and ADP-ribosylation factor 1 methylation in human glioblastoma. Cell Oncol (Dordr) 2017;40:389-399.

29. Ghasemi A, Fallah S, Ansari M. MiR-153 as a Tumor Suppressor in Glioblastoma Multiforme is Downregulated by DNA Methylation. Clin Lab 2016;62:573-580.

30. Ayala-Ortega E, Arzate-Mejia R, Perez-Molina R, Gonzalez-Buendia E, Meier K, Guerrero G, Recillas-Targa F. Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines. BMC Cancer 2016;16:226.

31. Ma J, Hou X, Li M, Ren H, Fang S, Wang X, He C. Genome-wide methylation profiling reveals new biomarkers for prognosis prediction of glioblastoma. J Cancer Res Ther 2015;11 Suppl 2:C212-215.

32. Jha P, Pia Patric IR, Shukla S, Pathak P, Pal J, Sharma V, Thinagararanjan S, et al. Genome-wide methylation profiling identifies an essential role of reactive oxygen species in pediatric glioblastoma multiforme and validates a methylome specific for H3 histone family 3A with absence of G-CIMP/isocitrate dehydrogenase 1 mutation. Neuro Oncol 2014;16:1607-1617.

33. Lai RK, Chen Y, Guan X, Nousome D, Sharma C, Canoll P, Bruce J, et al. Genome-wide methylation analyses in glioblastoma multiforme. PLoS One 2014;9:e89376.

34. Zhang W, Yan W, You G, Bao Z, Wang Y, Liu Y, You Y, et al. Genome-wide DNA methylation profiling identifies ALDH1A3 promoter methylation as a prognostic predictor in G-CIMP- primary glioblastoma. Cancer Lett 2013;328:120-125.

35. Sun X, St John JC. Modulation of mitochondrial DNA copy number in a model of glioblastoma induces changes to DNA methylation and gene expression of the nuclear genome in tumours. Epigenetics Chromatin 2018;11:53.

36. Serrano J, Snuderl M. Whole Genome DNA Methylation Analysis of Human Glioblastoma Using Illumina BeadArrays. Methods Mol Biol 2018;1741:31-51.

37. Tran A, Escovedo C, Migdall-Wilson J, Chou AP, Chen W, Cloughesy T, Nelson S, et al. In Silico Enhanced Restriction Enzyme Based Methylation Analysis of the Human Glioblastoma Genome Using Agilent 244K CpG Island Microarrays. Front Neurosci 2009;3:57.

38. Jovcevska I. Sequencing the next generation of glioblastomas. Crit Rev Clin Lab Sci 2018;55:264-282.

39. Barault L, Amatu A, Bleeker FE, Moutinho C, Falcomata C, Fiano V, Cassingena A, et al. Digital PCR quantification of MGMT methylation refines prediction of clinical benefit from alkylating agents in glioblastoma and metastatic colorectal cancer. Ann Oncol 2015;26:1994-1999.

40. Ramirez JL, Taron M, Balana C, Sarries C, Mendez P, de Aguirre I, Nunez L, et al. Serum DNA as a tool for cancer patient management. Rocz Akad Med Bialymst 2003;48:34-41.

41. Bloch O, Lim M, Sughrue ME, Komotar RJ, Abrahams JM, O'Rourke DM, D'Ambrosio A, et al. Autologous Heat Shock Protein Peptide Vaccination for Newly Diagnosed Glioblastoma: Impact of Peripheral PD-L1 Expression on Response to Therapy. Clin Cancer Res 2017;23:3575-3584.

42. Reynes G, Vila V, Fleitas T, Reganon E, Font de Mora J, Jorda M, Martinez-Sales V. Circulating endothelial cells and procoagulant microparticles in patients with glioblastoma: prognostic value. PLoS One 2013;8:e69034.

43. Rosell R, de Las Penas R, Balana C, Santarpia M, Salazar F, de Aguirre I, Reguart N, et al. Translational research in glioblastoma multiforme: molecular criteria for patient selection. Future Oncol 2008;4:219-228.

44. Faria G, Silva E, Da Fonseca C, Quirico-Santos T. Circulating Cell-Free DNA as a Prognostic and Molecular Marker for Patients with Brain Tumors under Perillyl Alcohol-Based Therapy. Int J Mol Sci 2018;19.

45. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-210.

46. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, et al. ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003;31:68-71.

47. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317-330.

48. Shankavaram UT, Varma S, Kane D, Sunshine M, Chary KK, Reinhold WC, Pommier Y, et al. CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. BMC Genomics 2009;10:277.

49. Kong J, Cooper LA, Wang F, Gutman DA, Gao J, Chisolm C, Sharma A, et al. Integrative, multimodal analysis of glioblastoma using TCGA molecular data, pathology images, and clinical outcomes. IEEE Trans Biomed Eng 2011;58:3469-3474.

50. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 2010;8:e1000533.

51. Guo S, Zhu Q, Jiang T, Wang R, Shen Y, Zhu X, Wang Y, et al. Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis. Mod Rheumatol 2017;27:441-447.

52. Ding W, Pu W, Wang L, Jiang S, Zhou X, Tu W, Yu L, et al. Genome-Wide DNA Methylation Analysis in Systemic Sclerosis Reveals Hypomethylation of IFN-Associated Genes in CD4(+) and CD8(+) T Cells. J Invest Dermatol 2018;138:1069-1077.

53. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004;306:636-640.

54. Pu W, Wang C, Chen S, Zhao D, Zhou Y, Ma Y, Wang Y, et al. Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC). Clin Epigenetics 2017;9:129.

55. Guo S, Yan F, Xu J, Bao Y, Zhu J, Wang X, Wu J, et al. Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC). Clin Epigenetics 2015;7:3.