Detecting Differential Methylation in PTSD Patients with Dirichlet Mixtures

Ilyas Patanam Center for Human Diseases
Duke Unviersity University
Durham, NC 27705
imp9@duke.edu

Abstract

The underlying molecular mechanisms of PTSD are largely unknown. We hypothesize DNA methylation may be significant in the pathophysiology of PTSD since DNA methylation is recognized as an epigenetic mechanism for environmental regulation of genes. The high-throughput Illumina HumanMethylation450 Beadchips provides quantitative methylation measurements at 482,421 CpG sites allowing for epigenetic wide association studies. The distribution of sample measurements is bounded between 0 (unmethylated) and 1 (fully methylated) and is multimodal. We utilize a nonparametric Bayes test to identify differentially methylated sites between 2 groups, veterans diagnosed with PTSD and a control group of veterans. Compared to the standard 2 sample t-test used to detect sites, our model incorporates the interpretation that the 2 cases arise from the same discrete process but in potentially different proportions.

1 Introduction

Postraumatic stress disorder (PTSD) is an atypical psychological nd physiological response that can occur among persons exposed to a potentially traumatic event involving life threat, serious injury, or death. To obtain a diagnosis of PTSD reactions to the traumatic stressor must include fear, helplessness, or horror. Additionally, persons must experience symptoms from intrusive recollections, persistent avoidance of thoughts associated with the trauma, and increased arousal. When these cooccurring symptoms are present for a minimum of 1 month, a patient may be diagnosed with PTSD.

Recent work suggests that epigenetic DNA methylation changes may accompany lifetime experiences and alter gene expression profiles. Methylation of cytosine bases in DNA CpG islands is an important epigenetic regulation mechanism in organ development, aging, and different disease statuses. Hypermethylation of CpG islands located in the promoter regions can repress transcription of the associated genes. Thus, high throughput profiling of DNA methylation status of CpG islands is crucial for understanding the biological mechanism causing PTSD.

2 Methods

In our study of 192 African American veterans, 91 were diagnosed with PTSD and 101 were non-PTSD veterans. We used the DNA derived from their whole blood and used the Infinium Human Methylation 450 BeadChip by Illumina to asses the methylation profiles of more than 482,421 CpG sites covering 99% of RefSeq genes including those in regions of low CpG island density.

We introduce a Bayesian approach for screening using shared kernels. The population distribution for each variable is approximated using a mixture of kernels $F_k{}_{k=1}^{K}$. For this case-control study,

we test whether the groups have different kernel wights. The generative Model is $x_{mn} \sim F_k$ with probability π_{mk}. For group distributions $F_m(0)$ and $F_m(1)$ at variable m,

$$F_m^{(0)} = \Sigma_{(k=1)}^{K} \pi_{mk}^{(0)} F_k \text{ and } F_m^{(1)} = \Sigma_{k=1}^{K} \pi_{mk}^{(1)} F_k$$

and the competing hypotheses are

$$H_{0m} : \pi_{mk}^{(0)} = \pi_{mk}^{(1)} \; \forall \; k$$
$$H_{0m} : \pi_{mk}^{(0)} \neq \pi_{mk}^{(1)} \text{ for some } k$$

In practice $F_1, \ldots, F_K$ and a shared Dirichlet prior distribution for the weights $\Pi_m^{(0)}, \Pi_m^{(1)}$ are estimated empirically. A Gibbs sampling procedures is then used to estimate the posterior probability of H_{0m} for each variable.

Under H_{0m} the distribution for the component memberships $C_m^{(0)}$ and $C_m^{(1)}$ is

$$pr(C_m^{(0)}, C_m^{(1)} | H_{(0m)} = \beta(\vec{n}_m + \alpha)/\beta(\alpha)$$

where β is the multivariate beta function.

Similarly, under H_{1m},

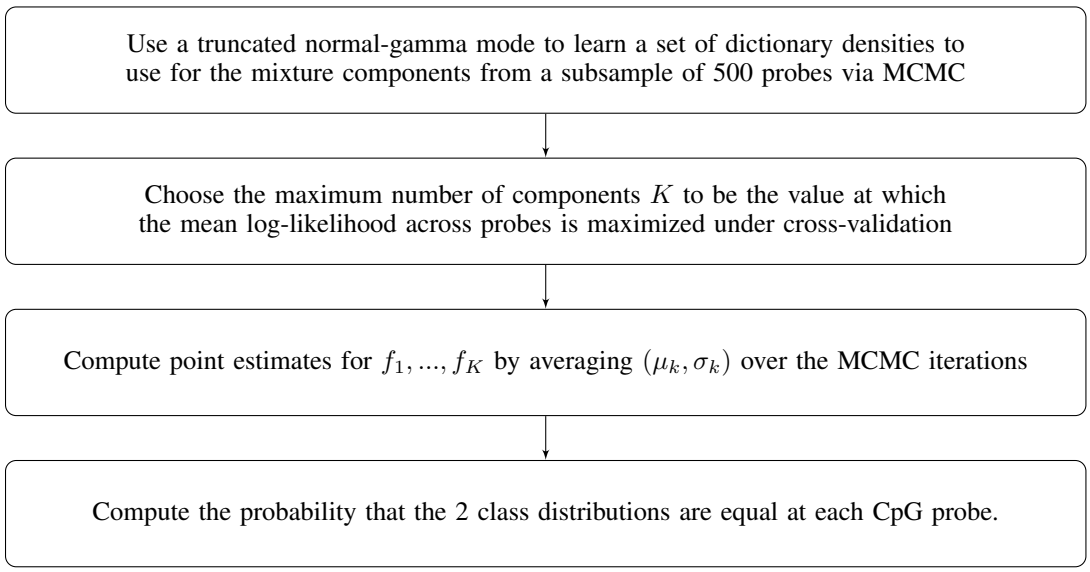
$$pr(C_m^{(0)}, C_m^{(1)} | H_{1m}) = \frac{\beta(\vec{n}_m^{(0)} + \alpha)\beta(\vec{n}_m^{(1)} + \alpha)}{\beta(\alpha)^2}$$

Let $P_0 = pr(H_{(0m)})$ be a global prior probability of no prior difference. The posterior probability of $H_{(0m)}$ given $C_m^{(0)} and C_m^{(1)}$ has the closed form

$$P(H_{0m} | C_m^{(0)}, C_m^{(1)}) = \frac{P_0 \beta(\alpha)\beta(\vec{n}_m + \alpha)}{P_0 \beta(\alpha)\beta(\vec{n}_m + \alpha) + (1 - P_0)\beta(\vec{n}_m^{(0)} + \alpha)\beta(\vec{n}_m^{(1)} + \alpha)}$$

3 Discussion

We use a Bayesian approach with Dirichlet prior distributions for the mixing weights Π^0, Π^1 to estimate the posterior probability $P(H_0 | X^{(0)}, X^{(1)})$. Our model assumes $F^{(0)}$ and $F^{(1)}$ are drawn from the same mixture components but with potentially different mixing weights.



The number of kernels $K = 9$ is chosen by cross validation based on the mean log-likelihood for held out observations. For fixed $f_1, \ldots, f_9$, we compute the posterior for the two-group model at each CpG site using a simple and efficient Gibbs sampler and a uniform prior for P_0. We calculate the component likelihood $f_k(x_{mn})$ for all sites m, samples n, and components k in advance to reduce the computational burden.

4 Future Directions

Although quality control probes were checked, certain probes gave contradictory information. We need to clarify if the readings from certain quality control probes necessitate discarding the sample or controlling for the quality control effects during analysis. Then, we will need to look into the biological significance of the CpG sites that we discovered.

Acknowledgments

I want to thank my mentor Allison Ashley Koch for helping guide me through this analysis, and David Dunson and Eric Lock for their method which I implemented, I also appreciate Sayan's guidance throughout the course.

References

[1] Dunson, D and Lock, E. (2014) *Shared kernel Bayesian Screening.*

[2] Uddin, M., et. al. (2010) *Epigenetic and immune function profiles associated with post traumatic stress disorder. Proceedings of National Academy of Sciences* **107**(20):9470-9475