DNAm Hierarchical Model - Fixed Effects Differential Methylation

Kevin Murgas

2020-09-28

Contents

1 Differential Methylation

Differential methylation is typically defined using a basic linear model over the data points between normal and tumor samples.

$$M_i = \mu + \Delta M * \text{tInd}(i)$$

Where tInd is a binary indicator for tumor samples.
This linear fit captures the normal mean (intercept) and shift in methylation in tumors (slope). The slope (ΔM) is used to define differentially methylated regions by first z-score normalizing and then determining statistical significance using a permutation based false discovery rate (FDR)

$$z = \text{mean}(\Delta M)/\text{s.d.}(\Delta M)$$

I am currently a bit unclear on the FDR step. Irizarry et al 2009 reference : Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes analysis of a microarray experiment. Journal of the American statistical association. 2001 Dec 1;96(456):1151-60.
Z-scores can convert to p-values, and Benjamini-Hochberg step-up can account for multiple hypothesis correction. The permutation could be from shuffling the normal/tumor labels on 1000 draws.

2 Model Description: Fixed Effects

2.1 Model equation

Let i=sample, j=patient

$$y_i = \mu + a_j + \text{tInd}(i) * (\beta_T + b_j + c_i)$$
$$a_j \sim N(0, \sigma_P), b_j \sim N(0, \sigma_{PT}), c_i \sim N(0, \sigma_T)$$

2.2 Differential Methylation using betaT

Effects μ and β_T describe the normal mean methylation and tumor shift in methylation, respectively. Therefore, differentially methylated CpG sites will have significant betaT values. A similar procedure is used as the ΔM , using a z-score normalization, and permutation based FDR rate.
This provides a way to identify CpG sites with significant differential methylation. These sites can then be directly compared to C-DMRs of Irizarry 2009 (provided as regions, i.e. base ranges on a given chromosome), or go up to identify genes with either any d.m. CpG sites or a certain fraction. Genes can also be compared to Irizarry C-DMRs (each region has a single gene listed)