An Evaluation of Statistical Methods for DNA Methylation Microarray Data Analysis

Dongmei Li, PhD

Clinical and Translational Science Institute
School of Medicine and Dentistry
University of Rochester
Email: *Dongmei_Li@urmc.rochester.edu*

Li, D; Xie, Z.; Le Pape, M.; Dye, T. "An evaluation of statistical methods for DNA methylation microarray data analysis". BMC Bioinformatics. 2015; 16(217).

November 16, 2015

Outline

DNA methylation

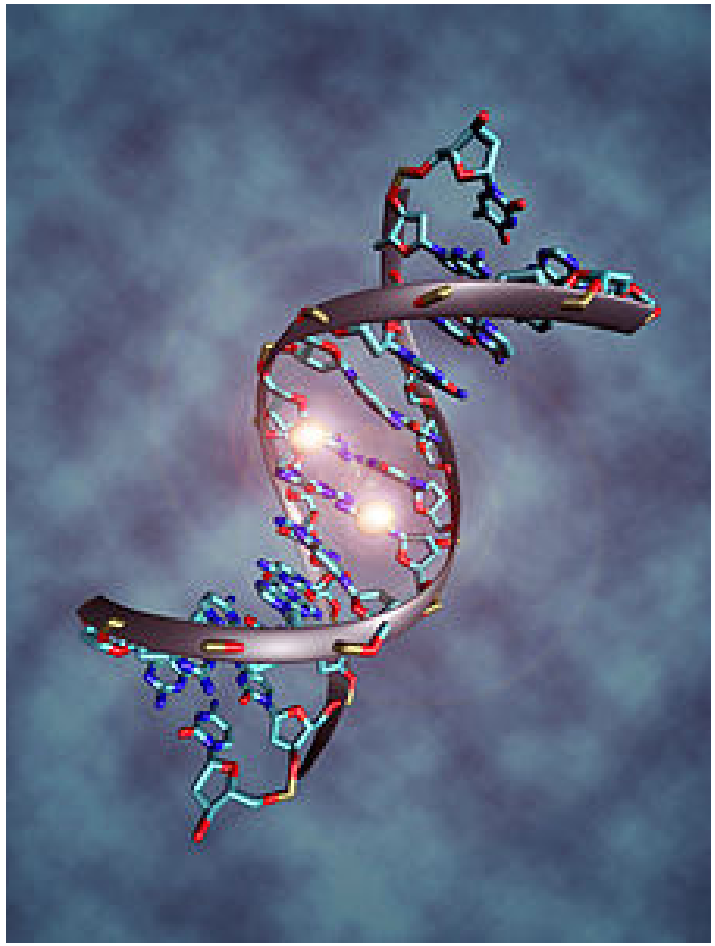

Figure: DNAmolecule that is methylated at the two center cytosines. Source: *http* : *//en.wikipedia.org/wiki/DNA*$_m$*ethylation*

DNA methylation and disease

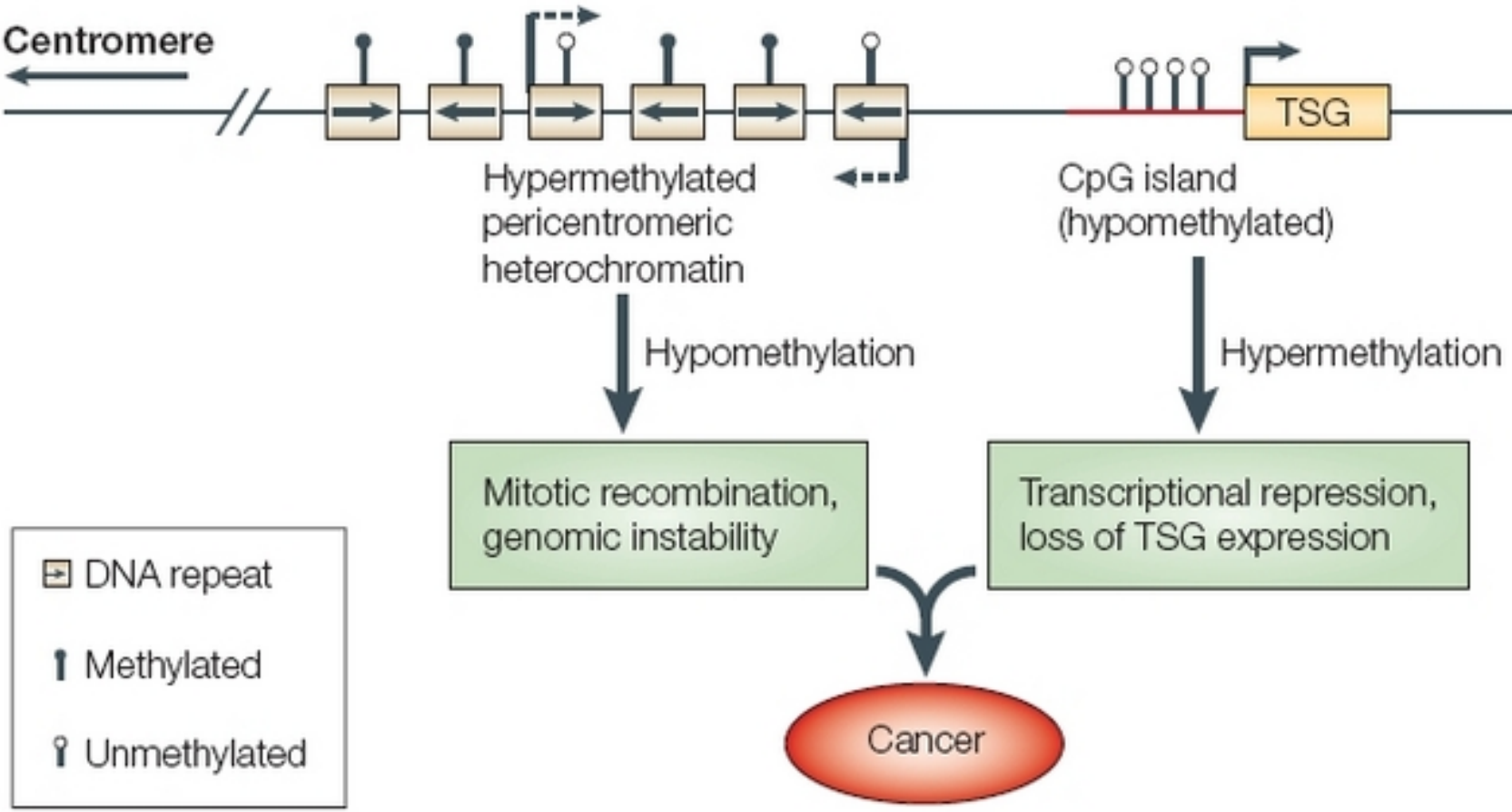

Figure: Diagram for DNA methylation and cancer. Source: 2005Nature Publishing GroupRobertson, K. DNA methylation and human disease. *Nature Reviews Genetics*6,598.

Epigenome-wide association studies (EWAS)

- Illumina Methylation Assay
- Three platforms for DNA methylation assay
  - GoldenGate ($1,563$ methylation site per sample)
  - Infinium Human Methylation27 ($> 27,000$ methylation sites per sample)
  - Infinium HumanMethylation450 BeadChip ($> 485,000$ methylation sites per sample)

Work flow of Illumina Assay

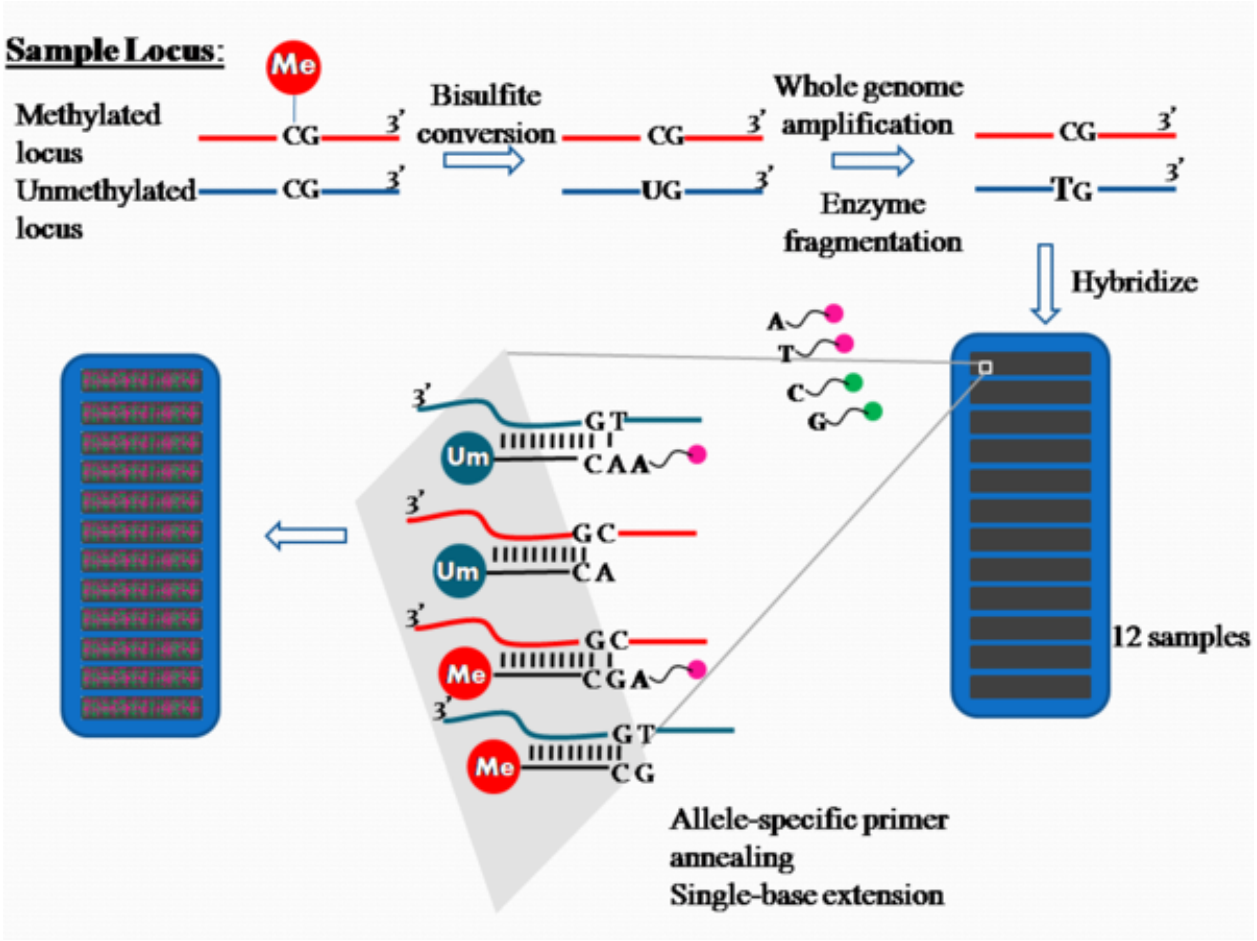

Figure: Source: *http://en.wikipedia.org/wiki/Illumina_Methylation_Assay*

Methylation array downstream analysis

- Locus-by-Locus analyses are commonly used for EWAS
- Average β value denote the level or percentage of methylation for a locus
- M value, or log ratio of percentage of methylation, is also commonly used to measure methylation
- Relationship between the β-value and the M-value

$$M = log_2 \frac{\beta}{1-\beta}$$

Methods implemented in Bioconductor/R

- Wilcoxon rank sum test (methyAnalysis)
- t-test (methyAnalysis, CpGAssoc, RnBeads, and IMA package)
- Kolmogorov-Smirnov Tests (Used in some papers: Price et al. Epigenetics & Chromatin 2013, 6:4)
- Permutation test (CpGAssoc package)
- Empirical Bayes method (RnBeads, IMA and minfi package)
- Bump hunting method (minfi package)

Motivation for evaluation methylation data analysis methods

- Finding the most appropriate one to use for a specific data set is challenging
- Different methods have different assumptions to get validate results
- Multiple methods could provide inconsistent results for the same data set
- Exploring power and stability differences across different methods for the same data set
- Proving advice for investigators choosing appropriate method for their methylation data

Definition of power and stability

	number not rejected	number rejected	
true null hypotheses	U	V	m_0
non-true null hypotheses	T	S	m_1
total	$m - R$	R	m

Table: Possible outcomes from m hypotheses tests

$$Power = E(\frac{S}{m_1}|m > m_0),$$

$$Stability = Var(R) = Var(S + V) = Var(S) + Var(V) + 2Cov(S, V).$$

Wilcoxon rank sum test

$$H_0 : Median_1 = Median_2$$

$$z = \frac{R - \mu_R}{\sigma_R}$$

where

$$\begin{aligned}
\mu_R &= \frac{n_1(n_1 + n_2 + 1)}{2} \\
\sigma_R &= \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} \\
R &= sum \quad of \quad ranks \quad for \quad smaller \quad sample \quad size \quad (n_1) \\
n_1 &= smalle \quad of \quad sample \quad sizes \\
n_2 &= larger \quad of \quad sample \quad sizes \\
n_1 &\geq 10 \quad and \quad n_2 \geq 10
\end{aligned}$$

t-test

$$H_0 : \mu_1 = \mu_2$$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}_1 - \bar{y}_2}}$$

where

$$s_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{(\frac{s_{i1}^2}{n_1} + \frac{s_{i2}^2}{n_2})^2}{\frac{(\frac{s_{i1}^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_{i2}^2}{n_2})^2}{n_2 - 1}}.$$

Kolmogorov-Smirnov Tests

$$H_0 : F_{1,n_1}(y) = F_{2,n_2}(y)$$

$$D_{n_1,n_2} = \sup_y |F_{1,n_1}(y) - F_{2,n_2}(y)|$$

The null hypothesis is rejected at level α if

$$D_{n_1,n_2} > c(\alpha)\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

where $c(\alpha) = 1.36$ for $\alpha = 0.05$

Permutation test

$$H_0 : F_{1,n_1}(y) = F_{2,n_2}(y)$$

- Compute the test statistic for the observed data set
- Permute the original data in a way that matches the null hypothesis
- Calculate the critical value of a level α test based on the upper α percentile of the reference distribution
- Obtain the raw p-value from the reference distribution.

Empirical Bayes method

$$H_0 : \beta^*_{gj} = 0$$

The moderated t-statistic, based on a hybrid classical/Bayes approach, is defined by:

$$\tilde{t}_{gj} = \frac{\hat{\beta}^*_{gj}}{\tilde{s}_g \sqrt{v_{gj}}}$$

The posterior mean of σ^2_g given s^2_g is

$$\tilde{s}^2_g = E(\sigma^2_g | s^2_g) = \frac{d_0 s^2_0 + d_g s^2_g}{d_0 + d_g}$$

The prior estimator s^2_0 and d_0 degrees of freedom is estimated from the data by equating empirical to expected values for the first two moments of $logs^2_g$

Bump hunting method

$$H_0 : \beta^*(t_j) = 0$$

Fit a linear model between methylation and disease type, covariates, and potential confounding variables

$$Y_{ij} = \mu(t_j) + \beta^*(t_j)X_i + \sum_{k=1}^{p} \gamma_k(t_j)Z_{i,k} + \sum_{l=1}^{q} a_{l,j}W_{i,l} + \epsilon_{i,j}$$

where i is ith subject and j is $j-$th genomic locus

1. Estimate $\beta(t_j)$ for each t_j
2. Use these to estimate the smooth function $\beta(t)$
3. Use this to estimate the regions R_n, $n = 1, \ldots, N$ for which $\beta(t) \neq 0$ for all $t \in R_n$
4. Use permutation tests to assign statistical uncertainty to each estimated region

Simulation Set up

- Methylation data are generated from mixed beta distributions to mimic real methylation data
- Proportion of methylated loci are 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90 to cover all possible scenarios
- Mean β value differences were set to be between 0.1 and 0.4 with steps equaling $1/(m - m_0)$ such as 1/10, 1/50, 1/100, 1/250, 1/500, 1/750, 1/900
- 1000 loci and 1000 independent simulations
- Sample sizes are 3, 6, 12, and 24 in each group with two-group comparisons
- Both β values and M values are compared

Methylation Results for sample size 3 in each group (Left: β values and Right: M values)

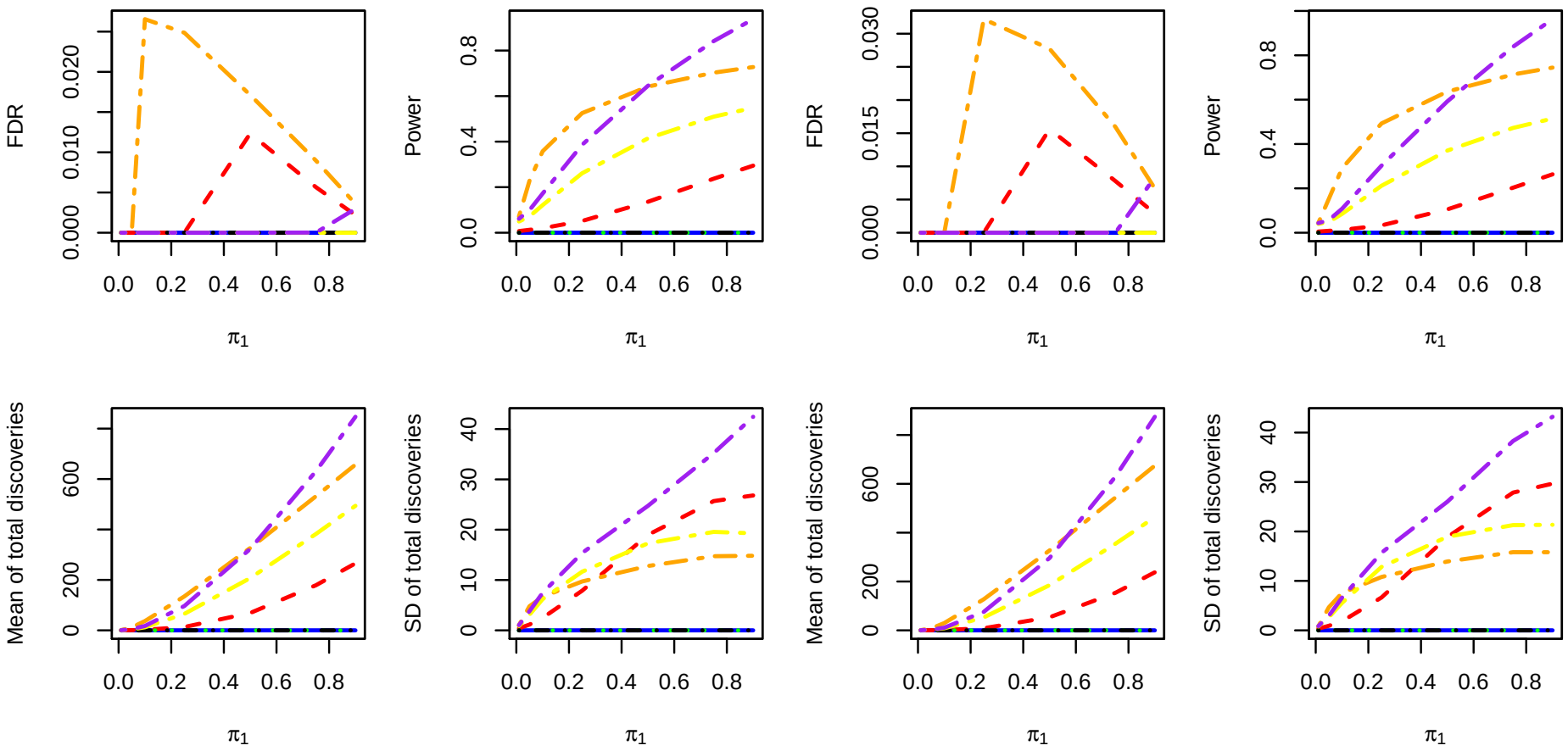

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH; Purple: Bump Hunting q-value.

Methylation Results for sample size 6 in each group (Left: β values and Right: M values)

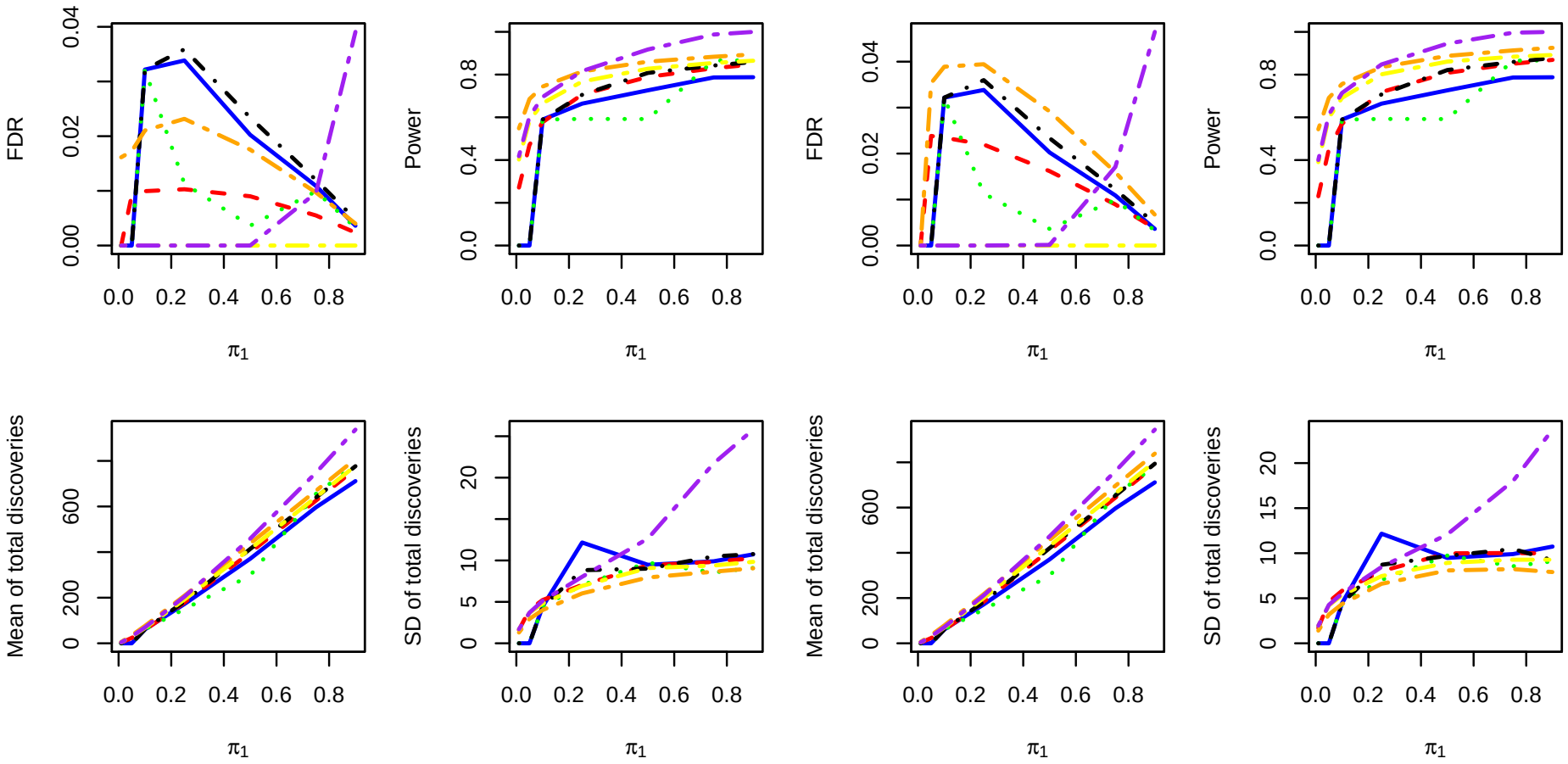

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH; Purple: Bump Hunting q-value.

Methylation Results for sample size 12 in each group (Left: β values and Right: M values)

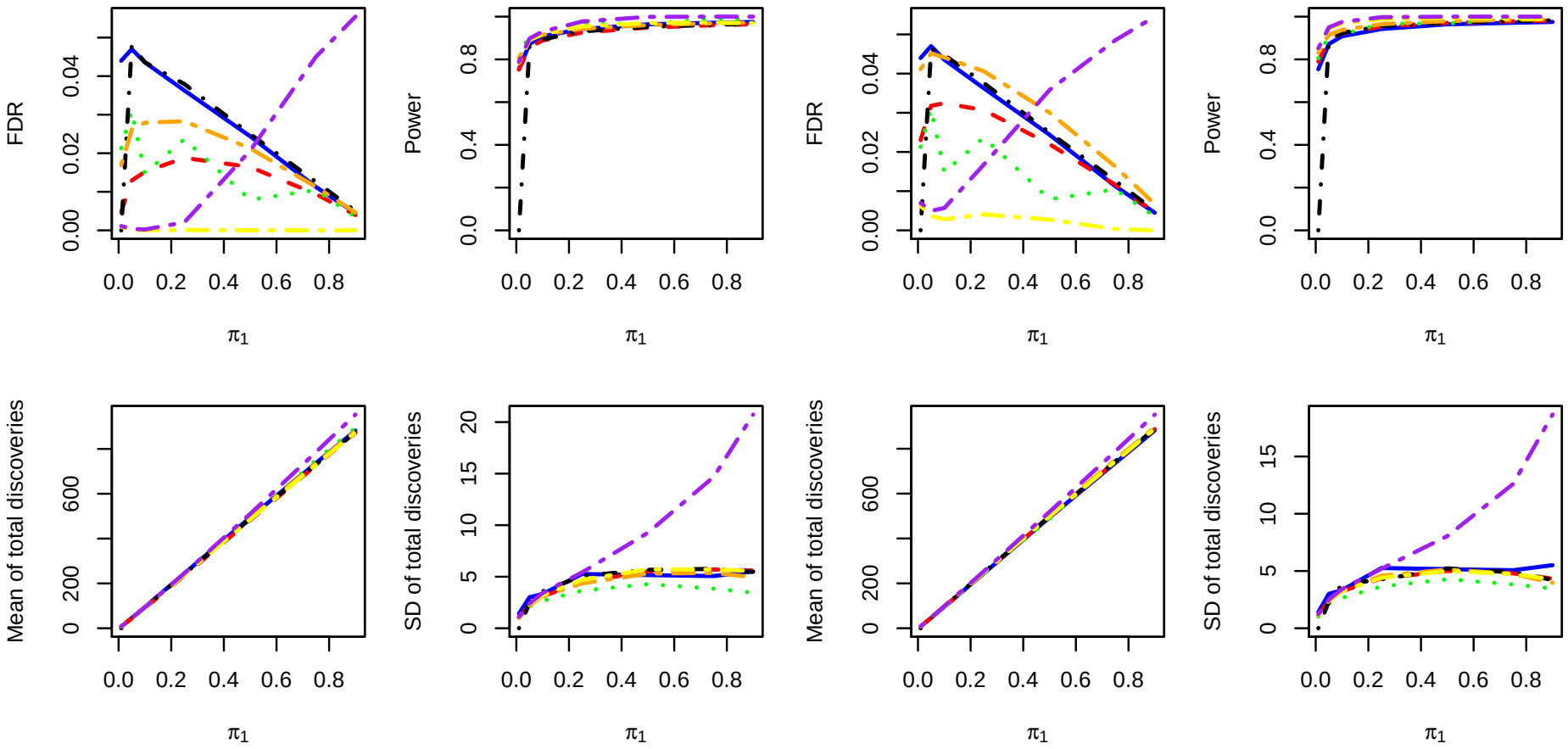

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH; Purple: Bump Hunting q-value.

Methylation Results for sample size 24 in each group (Left: β values and Right: M values)

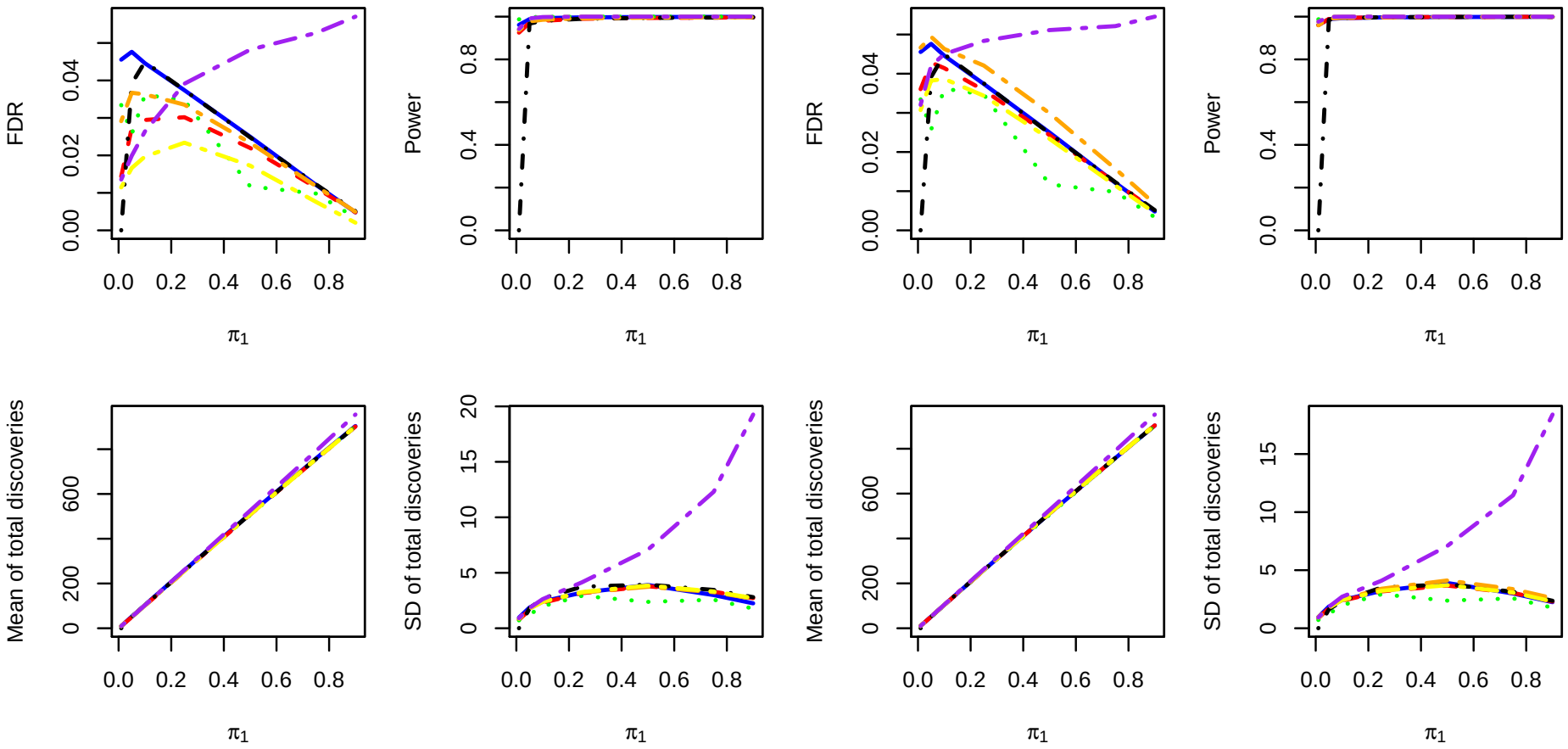

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH; Purple: Bump Hunting q-value.

Real data example

- Genome wide DNA methylation profiling of United Kingdom Ovarian Cancer Population Study (UKOPS) with GEO accession number GSE19711
- Illumina Infinium 27k Human DNA methylation Beadchip v1.2 with 27578 CpGs in whole blood sample from 3, 6, or 12 cases and 3, 6, or 12 controls
- Total number of rejections at 10 significance levels were recorded using raw p-values
- Both β values and M values are compared
- Raws p-values are used for comparisons and no rejections for Bump Hunting method using Storey's q-value adjustment

Apparent test power comparisons for $n = 3$ in each group

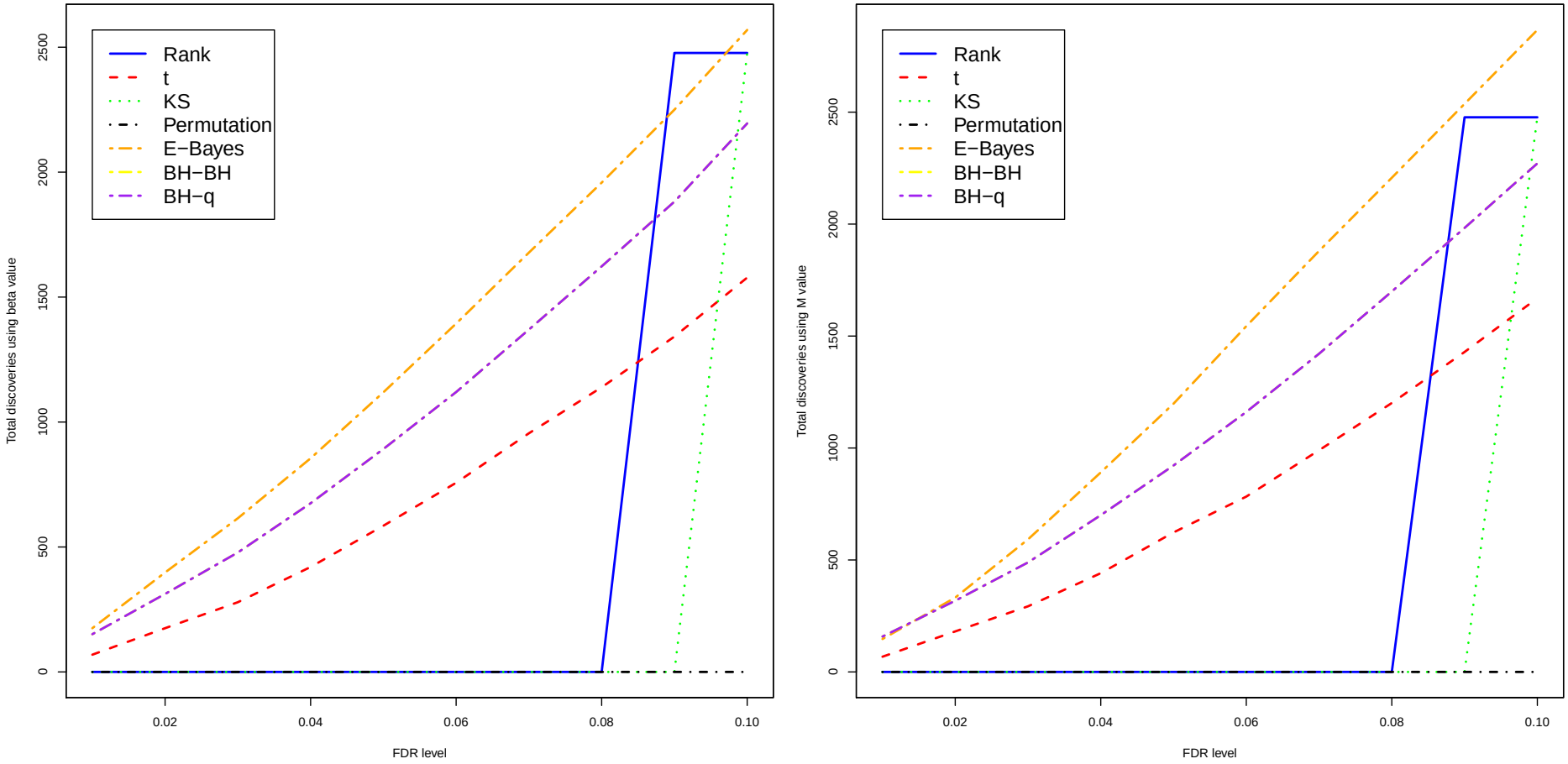

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH.

Apparent test power comparisons for $n = 6$ in each group

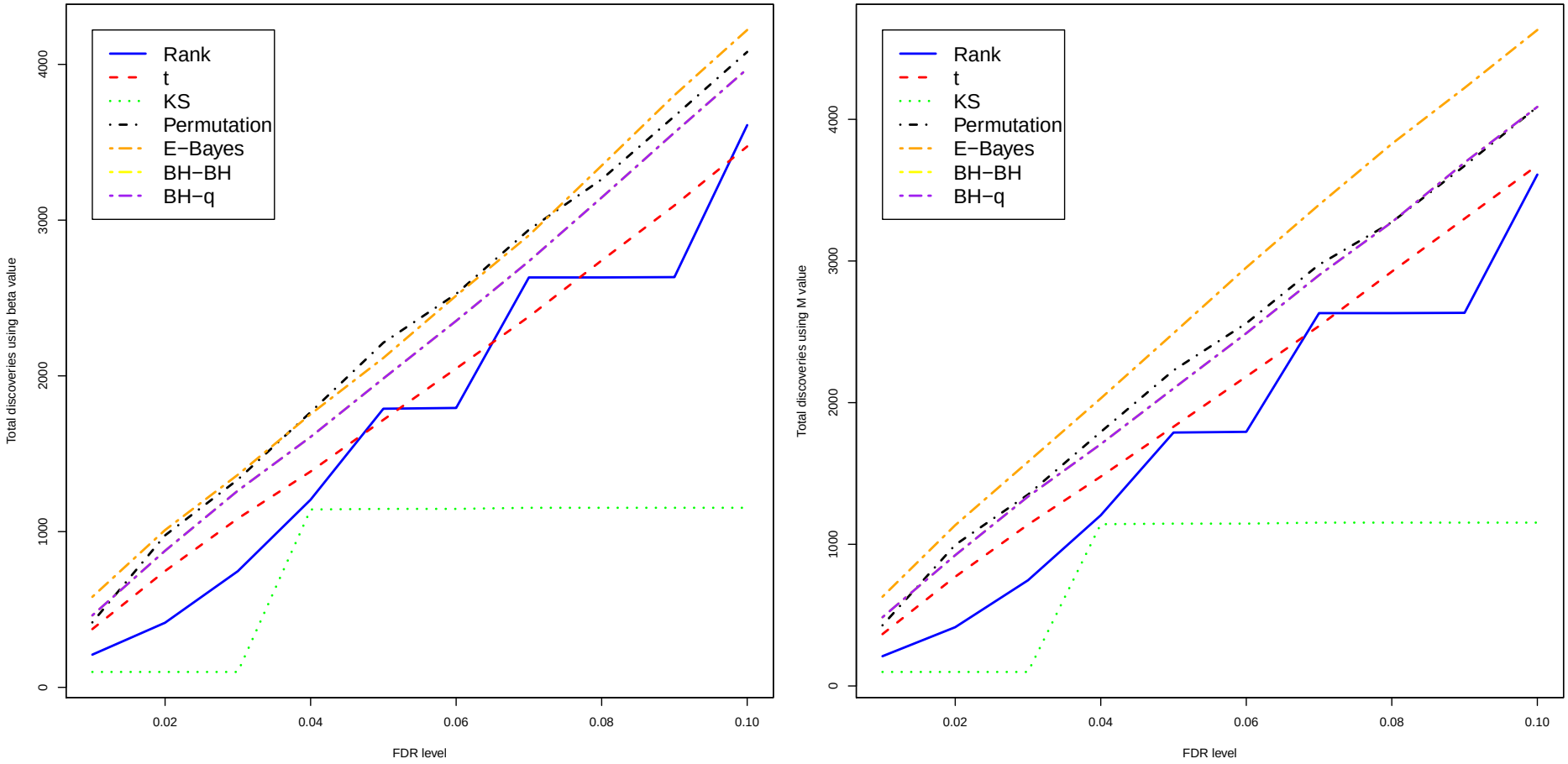

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH.

Apparent test power comparisons for $n = 12$ in each group

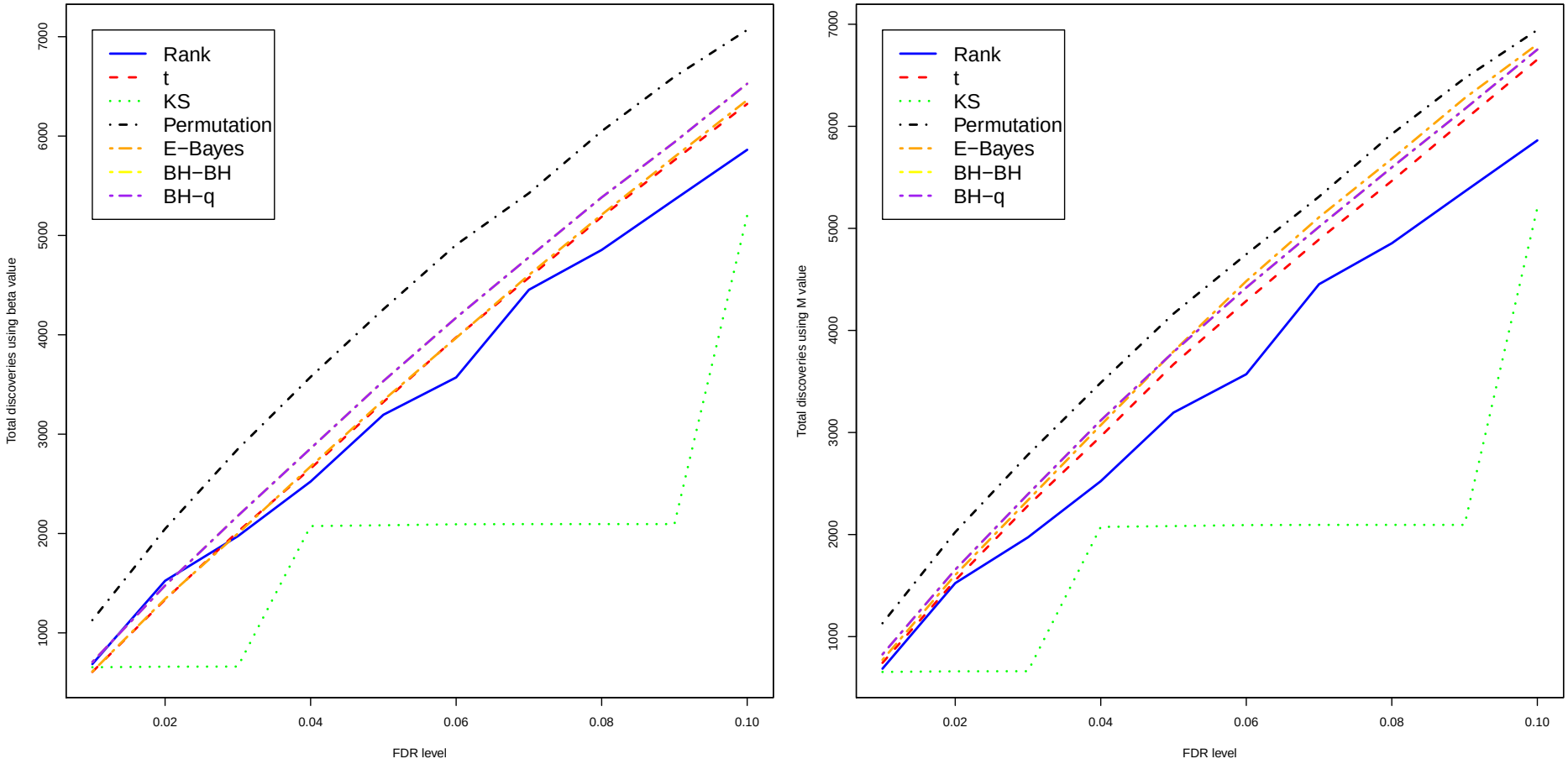

Figure: Blue: rank test; Red: t-test; Green: KS test; Black: permutation test; Orange: Empirical Bayes; Yellow: Bump Hunting BH.

Discussion

- No significant differences were detected in terms of FDR control, power, and stability between β values and M values
- For small sample size, both empirical Bayes method and bump hunting method showed good FDR control and much larger power than all other methods compared
- For medium to large sample size, all methods compared have good FDR control except the bump hunting method with large proportion of differentially methylated loci
- For medium to large sample size, all methods compared have almost equivalent power except permutation test with very low proportion of differentially methylated loci
- For all sample sizes, bump hunting method has lowest stability in terms of variance of total discoveries

Conclusion

- Either β values or M values are good to use for methylation data analysis
- Empirical Bayes method is recommended for methylation studies with small sample size
- For medium to large sample size, all methods except the bump hunting method are good for differentially methylation data analysis

Acknowledgement

- Clinical and Translational Science Institute