University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2008 *Paper* 239

FDR Controlling Procedure for Multi-stage Analyses

Catherine Tuglus* Mark J. van der Laan†

*Division of Biostatistics, University of California, Berkeley, ctuglus@berkeley.edu

†Division of Biostatistics and Department of Statistics University of California Berkeley

FDR Controlling Procedure for Multi-stage Analyses

Catherine Tuglus and Mark J. van der Laan

Abstract

Multiple testing has become an integral component in genomic analyses involving microarray experiments where large number of hypotheses are tested simultaneously. However before applying more computationally intensive methods, it is often desirable to complete an initial truncation of the variable set using a simpler and faster supervised method such as univariate regression. Once such a truncation is completed, multiple testing methods applied to any subsequent analysis no longer control the appropriate Type I error rates. Here we propose a modified marginal Benjamini \& Hochberg step-up FDR controlling procedure for multi-stage analyses (FDR-MSA), which correctly controls Type I error in terms of the entire variable set when only a subset of the initial set of variables is tested. The method is presented with respect to a variable importance application. As the initial subset size increases, we observe convergence to the standard Benjamini \& Hochberg step-up FDR controlling multiple testing procedure. We demonstrate the power and Type I error control through simulation and application to the Golub Leukemia data from 1999.

1 Introduction

Statistical analysis in genomics research often requires testing a large number of hypotheses simultaneously. This is especially true in microarray experiments where there are thousands of variables and often less than 100 observations. A common approach to determine which genes are significant is to apply univariate regression to all variables and test the significance of the coefficient β using a standard t-statistic with the null hypothesis $H_0 : \beta = 0$.

When there are many tests, multiple testing procedures are used to determine the rejection region of the null distribution corresponding to a specific Type I error rate, and, given this constraint, maximize power. One of the most common Type I error rates is the False Discovery Rate (FDR), which controls the expected proportion of Type I errors (False Positives) to total rejections.

When applying a more computationally intensive method, the researcher may want to initially reduce the dimensions of the data using a simple method, restricting it to a conservative set of potentially relevant genes. Often researchers will apply unsupervised methods, for instance restricting the set to genes that have a variance higher than a specified limit. When using these methods the potential of discounting relevant genes can be quite large since the truncation level is decided independent of the outcome. A supervised method which chooses genes based on their relationship to the outcome such as univariate regression or randomForest (Breiman et al., 1984; Breiman, 2001) would be preferred. However, once the initial variable set is restricted with respect to the outcome, multiple testing procedures on the results of secondary analyses are biased and no longer control the Type I and Type II error appropriately.

This is relevant for the newly developed targeted Maximum Likelihood based variable importance methodology (Bembom et al., 2007; Tuglus and van der Laan, 2008). Targeted Maximum Likelihood Estimation (tMLE) requires the estimation of an initial density $p(Y|W)$, with the observed data $O = (Y,W)$, where Y=outcome, W_j =variable of interest (i.e. a single gene), $W^* = W_{-j}$ =additional covariates (i.e. other genes). In practice this estimation procedure should be data-adaptive and must be completed for each variable of interest, W_j, since the appropriate covariate set, W^*, may vary from gene to gene. Data-adaptive methods can be very computationally intensive, and reduction of the initial gene set from 50,000 to 1,000 is very attractive.

In this paper we propose a modified marginal Benjamini & Hochberg step-up FDR controlling procedure for multi-stage analyses (FDR-MSA) which appropriately controls the FDR when applied to a reduced and data adaptively selected set of null hypotheses. We also show that if the restricted set contains all relevant variables, this procedure has equivalent control of Type I error and equivalent power to applying the standard Benjamini & Hochberg step-up FDR controlling procedure (BH-FDR) to the entire variable set.

This method is presented with respect to tMLE variable importance methodology, which we introduce in section two. In section three we summarize the FDR-MSA method, specifically as applied to tMLE variable importance methodology. In section four we provide simulation results demonstrating the Type I and Type II error control. In section five we present an application of the procedure to the commonly used Golub et al (1999) Leukemia data(Golub et al., 1999) , and we conclude with a discussion.

2 Variable importance application

As an example of the general application of FDR-MSA, we apply the proposed method in the context of variable importance analysis. The variable importance methodology is described thoroughly in van der Laan and Rubin (2006) and van der Laan (2005), previously applied in Bembom et al. (2007)

and Tuglus and van der Laan (2008), and is outlined here.

We observe n i.i.d. $O_i = (W_i, Y_i) \sim P$ making no assumptions on the data generating distribution P. Here W is a M dimensional vector of variables.

The objective is to identify variables in W, which are significantly associated with outcome Y, therefore in this case the parameter of interest will be a measure of the effect of variable W_j on Y.

For this analysis, the marginal variable importance of a particular W_j on outcome Y controlling for covariates $W^* = W_{-j}$ can be defined generally as

$$\mu(w_j) = \mathbb{E}_{W^*}[m(w_j, W_{-j}|\beta(j))]$$

for a user supplied model m

$$m(W_j, W_{-j}|\beta) = \mathbb{E}_P[Y|W_j, W_{-j}] - \mathbb{E}_P[Y|W_j = 0, W_{-j}]$$

and satisfying $m(0, W_{-j} \mid \beta(j)) = 0$ for all $\beta(j)$ and W_{-j}.

Given an estimator $\beta_n(j)$ of $\beta(j)$, an estimate of this parameter of interest at a particular $W_j = w_j$ is defined as

$$\mu_n(w_j) = \frac{1}{n}\sum_{i=1}^{n}[m(w_j, W_{i,-j}|\beta_n(j))]$$

This working model based approach allows A to be continuous and include effect modifiers (i.e. $m(W_j, W_1|\beta(j)) = \beta_0 W_j + \beta_1 W_j W_1$). We note however that we define the parameter $\beta(j)$ and thus $\mu(w_j)$ non-parametrically.

In a typical application we define $m(W_j, W_{-j}|\beta)$ as $m(W_j, W_{-j}|\beta) = W_j\beta(j)$, so that the marginal importance of each W_j is represented by single coefficient value $\beta_n(j)$.

The benefit of this measure of importance is that under randomization assumptions, the importance level can be interpreted as a causal effect or, if the working model (i.e. a semi-parametric regression model) is misspecified, then it can be interpreted as a projection of the causal effect on a working model.

We estimate the parameter $\beta(j)$ with tMLE variable importance methodology for each W_j using the following algorithm.

1. Estimate initial density $Q_n^0(W_j, W_{-j}) = \mathbb{E}[Y|W_j, W_{-j}] = m(W_j, W_{-j}|\beta_n^0(j)) + g(W_{-j})$ using any software allowing specification of $m(W_j, W_{-j}|\beta_n^0(j)) = W_j\beta_n^0(j)$. We recommend polymars (Kooperberg et al., 1997; O'Connor), lars (Efron et al., 2004), or DSA (Sinisi and van der Laan, March 2004).

2. Estimate nuisance parameter $\mathbb{E}[W_j|W_{-j}]$ (a.k.a. "treatment mechanism") using any data-adaptive software package

3. Create clever covariate

$$r(W_j, W_{-j}) = \frac{d}{d\beta(j)}m(W_j, W_{-j}|\beta(j)) - \mathbb{E}\left[\frac{d}{d\beta(j)}m(W_j, W_{-j}|\beta(j))|W_{-j}\right]$$

   which for this case can be simplified to $r(W_j, W_{-j}) = W_j - \mathbb{E}[W_j|W_{-j}]$

4. Project Y onto $r(W_j, W_{-j})$ with $offset = Q_n^0(W_j, W_{-j})$, define the resulting coefficient as ε

5. update initial estimate $\beta_n^0 = \beta_n^0 + \varepsilon$ and overall density $Q_n^1(W_j, W_{-j}) = Q_n^0(W_j, W_{-j}) + \varepsilon r(W_j, W_{-j})$

6. Obtain standard error and inference for $\beta_n(j)$ using the influence curve as defined in Yu and van der Laan (2003), which corresponds to the double robust estimating function. This is possible because the tMLE solution also corresponds to the solution of the double robust estimating function (van der Laan, 2005; van der Laan and Rubin, 2006). Given scale factor $c = \mathbb{E}\left[\frac{d}{d\beta(j)} D(O|\beta_0(j), Q_0^1)\right]$, the empirical influence curve for a given W_j in W.

$$IC_j(O) = c^{-1} D(O|\beta_0(j), Q_0^1)$$

where,

$$D(p_0)(O) \equiv r(W_j, W_{-j})(Y - m(W_j, W_{-j}|\beta_0(j)) - Q_0(0, W_{-j}))$$

The covariance of $\beta_0(j)$ is asymptotically equivalent to the covariance of $IC_j(O)$.

$$\Sigma_n(j) = \frac{1}{n} \sum \hat{IC_j}(O) \hat{IC_j}(O)^T$$

where

$$\sqrt{n}(\beta_n(j) - \beta_0(j)) \sim N(0, \Sigma_n(j))$$

Covariance can also be estimated by bootstrap estimates of $\beta(j)$, but this would requiring extra computational time. If $\mathbb{E}(W_j \mid W_{-j})$ is estimated consistently, then the variance estimates based on the influence curve are consistent or asymptotically conservative.

3 Modified Marginal Benjamini & Hochberg Step-up FDR Controlling Procedure for Multi-stage Analyses

Given a multivariate parameter $\Psi(P) = (\psi(m) : m = 1, ...M)$, we can define the null hypotheses and alternative hypotheses in terms of the parameter null value ψ_0 which typically equals 0. For the two-sided hypothesis test, the null hypothesis is $H_0(m) = I(\psi(m) = \psi_0(m))$ and the alternative hypothesis is $H_1(m) = I(\psi(m) \neq \psi_0(m))$.

Whether or not we reject the null hypothesis is determined by the value of the test statistic $T_n = (T_n(m) : m = 1, ..., M)$. The parameter of interest is tested using a standard t-statistic defined as

$$T_n(m) = \sqrt{(n)} \frac{\psi_n(m) - \psi_0(m)}{\sigma_n(m)}$$

where $\psi_n(m)$ is an asymptotically linear estimator of $\psi(m)$ with specified influence curve $IC_m(P)(O)$, and $\sigma_n^2(m)$ is an estimate of the variance $\sigma^2(m) = \mathbb{E}[IC_m(P)(O)^2]$ of the influence curve.

Specifically for variable importance measures, testing $H_0 : \beta_n(j) = 0$, p-values can be determined using test statistic

$$T_n(j) = \frac{\sqrt{n}\beta_n(j)}{\sqrt{\Sigma_n(j,j)}} \sim_{n\to\infty} N(0,1)$$

where Σ_n is the covariance matrix of the vector influence curve $IC = (IC_m : m = 1 \ldots, M)$.

3.1 FDR-MSA Method

According to the standard Marginal Benjamini & Hochberg Step-up FDR controlling procedure (Benjamini and Hochberg, 1995) we define

$$FDR = \mathbb{E}\left[\frac{V}{R}\right] = \mathbb{E}\left(\frac{V}{R}|R>0\right)P(R>0)$$

where $V =$ number of false positives and $R =$ number of total rejections of the null.

For a set of variables $W = \{W_m, m = 1,...M\}$ Given a set of M test statistics $T_n = (T_n(m) : m = 1,...,M)$ and their associated p-values $p_n = (p_n(m) : m = 1,...,M)$, define the ordered set of p-values as $p_{n(1)} \leq p_{n(2)} \cdots \leq p_{n(M)}$. According to Benjamini & Hochberg (Benjamini and Hochberg, 1995), to control FDR at level α find $\hat{k}$ such that

$$\hat{k} = max\left\{k : p_{(k)} \leq \frac{k}{M}\alpha\right\}$$

and reject $p_{n(1)} \leq ... \leq p_{n(\hat{k})}$. We define the set of rejected null hypotheses as $\mathcal{R}$.

This multiple testing has the following *monotonicity property* in the p-values which we will exploit in our proposed two stage FDR procedure: If we replace p_k by a q_k for all k and $p_k \leq q_k$, then the set of rejections of FDR applied to q_k is included in the set of rejections of FDR applied to the original p_k. That is, the validity of our two stage proposal is simply based on applying this monotonicity property to q_k equal to p_k for a supervised/data adaptively selected subset of the null hypotheses and setting $q_k = 1$ for all other null hypotheses.

For example, the supervised subset of the null hypotheses can be selected to be all variables for which their univariate regression p-value is smaller than 0.1. In the special case that our supervised subset does include the FDR selected set $\mathcal{R}$ when applied to the p-values p_k, then the two stage FDR procedure applied to q_k is equivalent with FDR applied to p_k.

Thus, in this case the two stage FDR is equivalent to applying BH FDR to all variables. To conclude, loss of Power will only occur if the initial restriction excludes variables which would have been rejected by the BH-FDR procedure when applied to the original p_k. As a consequence, the initial restriction should be generous to avoid loss of power.

We call this procedure the Modified Marginal Benjamini & Hochberg Step-up FDR controlling procedure for multi-stage analyses or simply FDR-MSA. This procedure is outlined below.

1. Given set of M variables, select a subset of U variables based on an initial supervised analysis.
2. Complete the wished test statistics for the null hypotheses of interest for these U variables only, and calculate their raw p-values.
3. Assign the p-value 1 to all M-U not selected variables.

   Thus, with regard to construction of the ordered list of p-values as needed for BH-FDR, to the end of the list of these U p-values add K=(M-U) 1's
4. Apply standard Marginal Benjamini & Hochberg Step-up FDR controlling procedure (Benjamini and Hochberg, 1995)

4 Simulations

We compare the performance of FDR-MSA under different levels of initial data screening. In this case the initial screen is determined by ranked p-values from univariate linear regression results. The different levels of screening correspond to different (increased) p-value cut-off values. The comparison is completed in terms of Type I error and Power.

4.1 Simulated Data

Covariate matrix W consists of 100 independent variables simulated from a multivariate normal distribution with variance 1 and mean vector created by randomly sampling mean values from $\{0.1, 0.2, ..., 9.9, 10.0, 10.1, ..., 50\}$.

Outcome Y is simulated from a main effect linear model using ten variables each with coefficient 4. These ten variables are designated as "true effects." A normal error with mean zero and variance $\sigma_Y = 10$ is added as noise. We use $\sigma_Y = 10$ to simulate a realistic noise scenario and provide enough variation to motivate false positive findings.

4.2 Analysis

Univariate linear regressions are applied to all 100 independent variables. Of these 100 variables we define subsets according to their ranked raw p-values from the univariate tests. We compare five different levels of screening corresponding to p-value cut-off values of $k_s = \{0.05, 0.1, 0.2, 0.3, \text{ and } 1\}$, where a subset is defined as all variables with raw univariate p-value less than or equal to a specific cut-off. A p-value cut-off of 1 corresponds to no initial screening of the data at which point FDR-MSA is equivalent to standard BH-FDR.

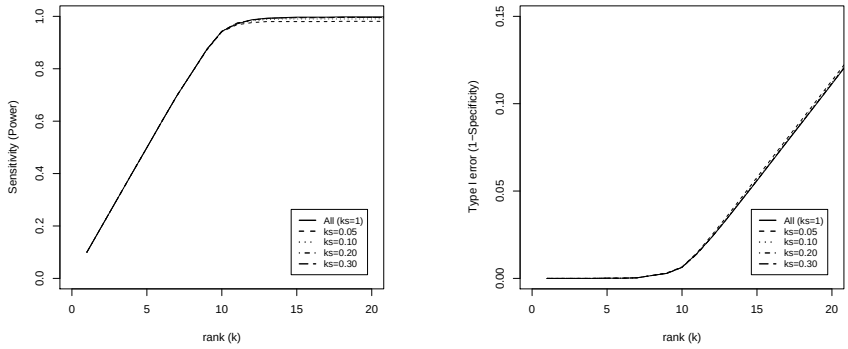
For each subset, we apply tMLE variable importance methodology and obtain measures (β) and associated inference under a null hypothesis $H_0 : \beta = 0$. Initial density estimate for $\mathbb{E}[Y|A,W]$ and $\mathbb{E}[A|W]$ are estimated using lasso regression (Tibshirani, 1996), applied by the lars() R package (Efron and Hastie). FDR-MSA is applied to each set of variable importance p-values.

We compare the performance of FDR-MSA under different levels of screening in terms of Type I error and Power. Type I error (or 1-Specificity) is defined as the probability of rejecting the null hypothesis ($\beta = 0$) when the null hypothesis is true and power (or Sensitivity) is defined as the probability of rejecting the null hypothesis ($\beta = 0$) when the alternative hypothesis ($\beta \neq 0$) is true.

Results are compared using plots representing levels of Power and Type I error with respect to p-value rank and p-value cut-off. We select the top k ranked p-values or all variables with p-value less than the specified cut-off (α) and assess the Power and Type I error among the variables in that group. Results are shown using the following four plots.

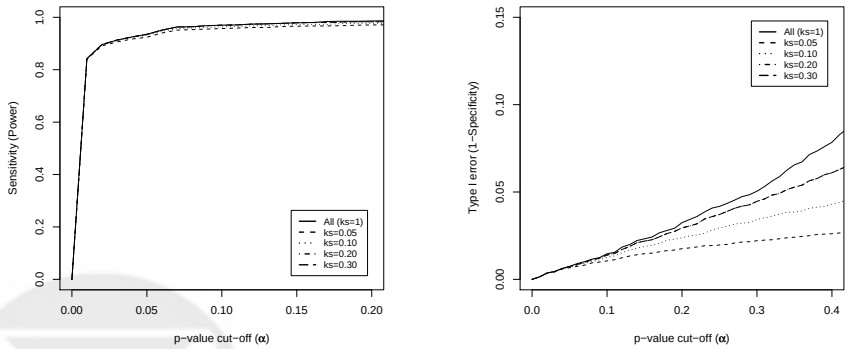
1. Sensitivity (Power) versus p-value rank (k)
2. Type I error (1-Specificity) versus p-value rank (k)
3. Sensitivity (Power) versus p-value cut-off (α)
4. Type I error (1-Specificity) versus p-value cut-off (α)

4.3 Results



(a) (b)

Figure 1: (a) Sensitivity (Power) and (b) Type I error (1-Specificity) versus p-value rank (k)



(a) (b)

Figure 2: (a) Sensitivity (Power) and (b) Type I error (1-Specificity) versus p-value cut-off α

Overall, as the size of the initial subset of variables increases (k_s increases), the performance of FDR-MSA in terms of both power and type I error converge to standard BH-FDR (under no truncation).

From the slight loss of power in Figure 1a and 2a it is evident that initially truncating the set of variables according to $k_s = 0.05$ was too harsh and did not allow all truly significant variables into the subset. As k_s is increased we see power converging to the power of BH-FDR applied to the full data (under no truncation).

In terms of ranked p-value, at k_s values of 0.1 and above the control of Type I error is equivalent to BH-FDR under no truncation. (Figure 1b). When Type I error is compared with respect to p-value

cut-off (Figure 2b) when controlling at a level of $\alpha = 0.05$ or below, the methods are equivalent for k_s values as high as 0.2. Above $\alpha = 0.05$, the FDR-MSA method slowly converges to the Type I error of BH-FDR on the entire data as the raw p-value cut-off is increased, becoming equivalent at higher α values.

5 Application - Leukemia

To illustrate its application in practice, the FDR-MSA method is applied to the Golub et al (1999) Leukemia data in conjunction with targeted variable importance (Tuglus and van der Laan, 2008). Targeted variable importance is applied to the full data and subsets of the data defined by an initial univariate raw p-value cut-off. The resulting p-values from each case will be adjusted with FDR-MSA. The resulting ranked lists will be compared.

5.1 Data

Variable importance methods (Bembom et al., 2007; Tuglus and van der Laan, 2008) are used to identify genes which distinguish patients with acute lymphoblastic leukemia (ALL) from patients with acute myeloid leukemia (AML). For the study presented in Golub et al (1999), the gene expression levels were measured using Affymetric oligonucleotide arrays with 6,817 human genes for n=38 patients (27 ALL, 11 AML). The gene expression set was pre-processed using unsupervised methods and reduced to 3,051 genes according to methods described in Dudoit et al. (2002). This dataset was obtained from the R package *multtest*, dataset golub (Pollard et al., 2005).

5.2 Analysis

Univariate logistic regressions are applied to all 3,051 variables. Of these variables we defined subsets according to their raw p-values from the univariate regressions. We restrict the data to all variables with raw p-values less than 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, and 1. We obtain tMLE variable importance measures and associated p-values for all subsets. The initial density estimate for $\mathbb{E}[A|W]$ is estimated using polymars regression, applied by the polspline() R package (O'Connor). To estimate $Q(A,W) = E[Y|A,W]$, we use lasso regression using the *lars* R package(Efron and Hastie). There are more powerful methods to data-adaptively select $Q(A.W)$, such as DSA(Sinisi and van der Laan, March 2004), and super Learner (van der Laan et al., 2007). Using a less powerful method does cost us consistency and efficiency with respect to our variable importance estimate. However *lars* provides a quick implementation of lasso regression making it convenient for this particular demonstration. Future work on the variable importance method will use more data-adaptive estimates for $Q(A,W)$ estimates.

We apply FDR-MSA to the resulting sets of variable importance p-values. Results are compared plotting the rank of the p-value versus its value.

5.3 Results

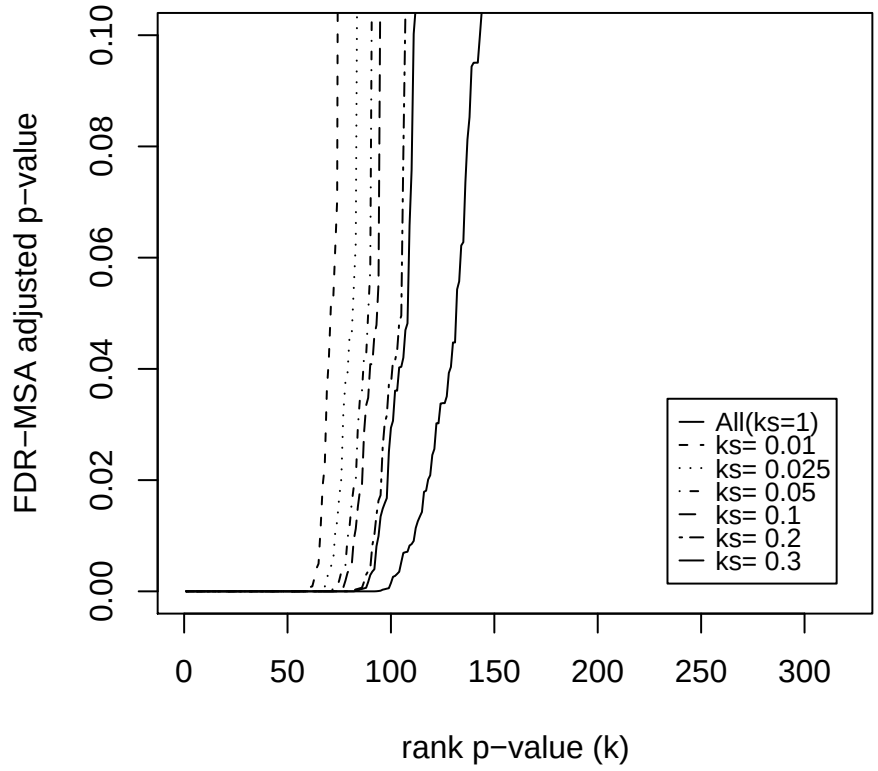


Figure 3: (a) FDR adjusted p-value versus p-value rank for FDR-MSA for raw p-value cut-offs $k_s = \{0.01, 0.025, 0.05, 0.1, 0.2, 0.3, \text{and} 1\}$

As we weaken the restriction on the initial cut and become more generous, we find that the results for FDR-MSA converge to the results when BH-FDR is applied to all the data.

The apparent loss in power from truncating the data is due to the initial screening process discounting important and significant variables. Ideally screening the data would result in an initial subset of variables that contain all true variables. The fact that univariate regression does not accomplish this suggests that more sophisticated screens are necessary.

6 Discussion

The FDR-MSA multiple testing procedure applied to a restricted subset of variables has equivalent power and type I error control to FDR applied to all variables when all BH-FDR-significant variables are present in the restricted set. Thus it conservatively controls FDR while only requiring calculation of the test statistics for the restricted and data adaptively selected set of null hypotheses among the complete set of null hypotheses.

Both in simulation and in practice we see that restricting solely based on raw p-values from univariate regression may not provide the desired subset of variables for subsequent analysis. We recommend to be thoroughly generous, to apply multiple supervised learning methods and take the union of the selected variables as the restricted set. For instance applying randomForest to the full dataset and taking all variables with non-zero importance or univariate regression p-values less than a particular cut-off (0.1 for instance). Alternatively, one can apply the tMLE-variable importance analysis with a simple and less computationally intensive initial regression estimator as a first stage analysis, selecting the restricted set based on a p-value cut-off. Approaches like this one will be investigated in more detail in order to improve the power of the FDR-MSA method while still maintaining the reduction in computation time.

Finally, we remark that the two stage method for multiple testing applies to any multiple testing procedure based on marginal p values which has the monotonicity property in the p-values. In particular, one could carry out a MSA modified method for controlling the generalized family-wise error (FWER-MSA) based on the appropriate multiple testing procedure controlling this type I error.

References

O. Bembom, M.L. Petersen, S. Rhee, W.J. Fessel, R.W. Shafer S.E. Sinisi, and M.J. van der Laan. Biomarker discovery using targeted maximum likelihood estimation: Application to the treatment of antiretroviral resistant hiv infection. *U.C. Berkeley Division of Biostatistics Working Paper Series*, Working Paper 221, 2007. URL `http://www.bepress.com/ucbbiostat/paper221`.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.*, 57:289–300, 1995.

L. Breiman, J. H. Freidman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth, 1984.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. URL `http://dx.doi.org/10.1023/A:1010933404324`.

S. Dudoit, J. Fridlyand, and T. P. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457): 77–87, 2002.

Brad Efron and Trevor Hastie. lars. R package.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *Annals of Statistics (with discussion)*, 32(2):407–499, 2004.

T.R. Golub, D.K. Slonim, P Tamayo, C Huard, M Gaasenbeek, J.P. Mesirov, H Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(531-537), 1999.

Charles Kooperberg, Smarajit Bose, and Charles J. Ston. Polychotomous regression. *Journal of the American Statistical Association*, 92:117–127, 1997.

Martin O'Connor. polymars. R package polspline.

K.S. Pollard, S. Dudoit, and M.J. van der Laan. *Multiple Testing Procedures: R multtest Package and Applications to Genomics in Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Number 209-229. Springer (Statistics for Biology and Health Series), 2005.

Sandra E. Sinisi and Mark J. van der Laan. Loss-based cross-validated deletion/substitution/addition algorithms in estimation. Working paper 143, U.C. Berkeley Division of Biostatistics Working Paper Series, March 2004. URL http://www.bepress.com/ucbbiostat/paper143.

R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Statist. Soc B.*, 58(1): 267–288, 1996.

C Tuglus and M.J. van der Laan. Targeted methods for biomarker discovery: The search for a standard. *U.C. Berkeley Division of Biostatistics Working Paper Series, submitted*, 2008.

Mark J. van der Laan. Statistical inference for variable importance. Technical Report Working Paper 188, U.C. Berkeley Division of Biostatistics Working Paper Series, 2005. URL http://www.bepress.com/ucbbiostat/paper188.

Mark J. van der Laan and Daniel Rubin. Targeted maximum likelihood learning. Working paper 213, U.C. Berkeley Division of Biostatistics Working Paper Series, 2006. URL http://www.bepress.com/ucbbiostat/paper213.

M.J. van der Laan, E.C. Polley, and A.E. Hubbard. Super learner. *U.C. Berkeley Division of Biostatistics Working Paper Series*, (Working Paper 222), 2007. URL http://www.bepress.com/ucbbiostat/paper222.

Zhuo Yu and Mark J. van der Laan. Measuring treatment effects using semiparametric models. Technical Report Working Paper 136, U.C. Berkeley Division of Biostatistics Working Paper Series, 2003. URL http://www.bepress.com/ucbbiostat/paper136.