Biometrika (2015), **102**, 4, *pp.* 974–980
Printed in Great Britain
doi: 10.1093/biomet/asv031
Advance Access publication 24 July 2015

Changepoint estimation: another look at multiple testing problems

BY HONGYUAN CAO

Department of Statistics, University of Missouri-Columbia, Columbia, Missouri 65211, U.S.A.
caohong@missouri.edu

AND WEI BIAO WU

Department of Statistics, University of Chicago, Chicago, Illinois 60637, U.S.A.
wbwu@uchicago.edu

SUMMARY

We consider large scale multiple testing for data that have locally clustered signals. With this structure, we apply techniques from changepoint analysis and propose a boundary detection algorithm so that the clustering information can be utilized. Consequently the precision of the multiple testing procedure is substantially improved. We study tests with independent as well as dependent p-values. Monte Carlo simulations suggest that the methods perform well with realistic sample sizes and show improved detection ability compared with competing methods. Our procedure is applied to a genome-wide association dataset of blood lipids.

Some key words: Changepoint detection; Multiple hypothesis testing; p-value aggregation.

1. INTRODUCTION

In many applied areas, one wants to perform multiple tests for detecting clustered signals. Suppose we test $H_{0i} : p_i \sim \mathcal{U}(0, 1)$ vs $H_{1i} : p_i \not\sim \mathcal{U}(0, 1)$, where p_i is the p-value of the ith test performed at location $i\,(i = 1, \ldots, m)$, and $\mathcal{U}(0, 1)$ denotes the standard uniform(0,1) distribution. Our global null hypothesis is

$$p_1, \ldots, p_m \sim \mathcal{U}(0, 1). \tag{1}$$

We first study the case of independent p-values and then consider the dependent case. Let $\theta_i = 0$ if H_{0i} is true and $\theta_i = 1$ if H_{0i} is false. In the latter case, we say that a signal is present at location i. For data that have clustered signals, we expect that a location and its adjacent neighbours have similar values of θ_i. Multiple testing with clustered signals has not been extensively studied in the literature. Clarkes & Hall (2009) studied the clustering effect due to dependence from stochastic processes. Siegmund et al. (2011b) proposed a scan statistic and treated each cluster as a testing unit. Sun et al. (2015) explored the spatial testing problem in a decision theoretical framework. Zhang et al. (2011) employed a p-value smoothing approach and showed that their approach had higher power than methods based on individual p-values.

In this paper, we develop a multiple testing procedure for data with clustered signals. Specifically, our global alternative hypothesis is formulated as follows: there exist changepoints $1 < \tau_1 < \cdots < \tau_l < m$ such that

$$\begin{aligned} &p_1, \ldots, p_{\tau_1 - 1} \sim \mathcal{U}(0, 1),\ \ p_{\tau_1}, \ldots, p_{\tau_2 - 1} \not\sim \mathcal{U}(0, 1), \\ &p_{\tau_2}, \ldots, p_{\tau_3 - 1} \sim \mathcal{U}(0, 1),\ \ p_{\tau_3}, \ldots, p_{\tau_4 - 1} \not\sim \mathcal{U}(0, 1), \ldots. \end{aligned} \tag{2}$$

Let $\tau_0 = 1$ and $\tau_{l+1} = m$. Decompose $\{1, \ldots, m\} = \cup_{i=0}^{l} \mathcal{S}_i$, where $\mathcal{S}_i = \{\tau_i, \ldots, \tau_{i+1} - 1\}$ $(i = 0, \ldots, l-1)$, and $\mathcal{S}_l = \{\tau_l, \ldots, m\}$. We shall call $\tau_1, \ldots, \tau_l$ changepoints in the testing sense. They partition $\{1, \ldots, m\}$ into alternating blocks of signal and noise segments. In our setting, the distributions of $p_{\tau_1}, \ldots, p_{\tau_2-1}, p_{\tau_3}, \ldots, p_{\tau_4-1}, \ldots$, in the alternative hypothesis blocks $\mathcal{S}_1, \mathcal{S}_3, \ldots$, can be different.

Example 1. Let the test statistics T_i follow $N(\mu_i, 1)$ $(i = 1, \ldots, m)$. We test $H_{0i} : \mu_i = 0$ against $H_{1i} : \mu_i \neq 0$. The p-values p_i equal $2\{1 - \Phi(|T_i|)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Under H_{0i}, $p_i \sim \mathcal{U}(0, 1)$. Under H_{1i}, the cumulative distribution function of p_i is $F_i(x) = \Phi\{-\Phi^{-1}(1 - x/2) - \mu_i\} + \Phi\{-\Phi^{-1}(1 - x/2) + \mu_i\} > x$. Equation (2) is equivalent to $\mu_1 = \cdots = \mu_{\tau_1-1} = 0, \mu_{\tau_1}, \ldots, \mu_{\tau_2-1} \neq 0, \mu_{\tau_2} = \cdots = \mu_{\tau_3-1} = 0, \mu_{\tau_3}, \ldots, \mu_{\tau_4-1} \neq 0, \ldots$.

Given a sequence of ordered data, a changepoint is a position at which the structure of the sequence changes. The goal of changepoint analysis is to estimate the locations of changepoints and to provide an assessment of accuracy. There is a huge literature on changepoint detection, and applications in high-throughput genomics have led to recent developments. For example, Olshen et al. (2004) proposed a circular binary segmentation approach, Tibshirani & Wang (2007) used the fused lasso with a smoothing constraint on regression coefficients, Lai et al. (2008) developed a hidden Markov model approach, and a Bayesian approach is studied in Lai & Xing (2011). Detection of common and rare variants based on single and multiple sequences can be found in Zhang et al. (2010), Siegmund et al. (2011a) and Jeng et al. (2013). Niu & Zhang (2012) developed a screening and ranking algorithm for changepoint detection with applications in copy number variation; its properties were studied by Hao et al. (2013).

In this paper, we use a changepoint estimation approach for detection of clustered signals as in (2). Unlike in Niu & Zhang (2012), where a changepoint is the position where the distribution changes, we assume that the p-value distributions within an alternative hypothesis segment can be different. Unlike in Siegmund et al. (2011b), where a cluster is treated as one testing unit, our testing unit is each individual hypothesis, and we use the cluster structure to better delineate the boundary between null and alternative hypotheses. Unlike in Zhang et al. (2011), where smoothed p-values were used in a two-group mixture model, we use smoothed p-values to construct a test statistic. In this way, a cluster of hypotheses with low signal levels may appear significant if the cluster is assessed as a whole, whereas it might be completely missed by methods that evaluate each hypothesis individually. By aggregating p-values, our method becomes more sensitive to signals of low magnitude. Additionally, if sporadic small p-values appear within a null hypothesis segment, we can avoid such false positives from occurring by averaging over the neighbouring p-values. Furthermore, we make use of the fact that the p-values under the null hypothesis have mean $1/2$.

2. Testing and estimation procedures

2·1. *Connection with changepoint analysis*

In the simplest classical changepoint problem, one assumes that there is only one changepoint and the alternative hypothesis is $p_1, \ldots, p_{\tau-1} \sim \mathcal{F}$, $p_\tau, \ldots, p_m \sim \mathcal{G}$ for some τ, where $\mathcal{F}$ and $\mathcal{G}$ are two different distributions. In the framework of multiple testing, we let $\mathcal{F} = \mathcal{U}(0, 1)$ and assume that $\mathcal{G}$ is stochastically smaller than $\mathcal{U}(0, 1)$ in the sense that $\mathrm{pr}(p_\tau \leqslant u) > u$ for all $u \in (0, 1)$. Then one can apply a cumsum test statistic of type

$$\max_{i \leqslant m} \left| \sum_{j=1}^{i} p_j - i\bar{p}_m \right| = \frac{1}{m} \max_{i \leqslant m} \left| (m - i) \sum_{j=1}^{i} p_j - i \sum_{j=1+i}^{n} p_j \right|, \tag{3}$$

where $\bar{p}_m = \sum_{j=1}^{m} p_j / m$. We can also estimate τ by the maximizer i for (3).

In our problem of estimating changepoints τ_i in (2), the approach based on (3) is not directly applicable. In (3), the p_i are identically distributed as $\mathcal{F}$ for $i \leqslant \tau$, and, for $i > \tau$, the p_i are also identically distributed as $\mathcal{G}$. In (2), however, the distributions of $p_{\tau_1}, \ldots, p_{\tau_2-1}$ under alternative hypotheses can be different. For example, there could exist $\tau' \in (\tau_1, \tau_2 - 1)$ such that $p_{\tau_1}, \ldots, p_{\tau'-1} \sim \mathcal{G}_1$ and $p_{\tau'}, \ldots, p_{\tau_2-1} \sim \mathcal{G}_2$ with $\mathcal{G}_1 \neq \mathcal{G}_2$. In this paper we are not interested in estimating such τ', the changepoint within alternative hypothesis segments. Additionally, due to the possibility of multiple changepoints, a test statistic of type (3) is not

directly applicable. Instead, we shall adopt a localized version. As a simple solution, one could consider the differences of local average of p_j over $j = i+1, \ldots, i+k_m$ and over $j = i-k_m+1, \ldots, i$:

$$d_i = \frac{1}{k_m}\left(\sum_{j=1}^{k_m} p_{i+j} - \sum_{j=0}^{k_m-1} p_{i-j}\right), \tag{4}$$

where $k_m \to \infty$ is a sequence of sliding window lengths for which $k_m/m \to 0$. In the sequel we write k for k_m for notational simplicity. Properties of the local discrepancy measure d_i and the global discrepancy $\max_{k \leqslant i \leqslant m-k} |d_i|$ have been widely studied; see for example Huskova & Slaby (2001), Zhao & Wu (2007), Niu & Zhang (2012) and a 2013 Karlsruhe Institute of Technology PhD thesis by M. Birte. In this paper we shall customize the local discrepancy measure d_i to multiple testing problems under a different setting. In a typical multiple testing problem, a majority of the p_i follow $\mathcal{U}(0, 1)$, which has mean $1/2$. Instead of (4), one can consider at location i the left and right differences

$$L_i = \left|\frac{1}{k}\sum_{j=i-k}^{i-1} p_j - \frac{1}{2}\right|, \quad R_i = \left|\frac{1}{k}\sum_{j=i}^{i+k-1} p_j - \frac{1}{2}\right|. \tag{5}$$

Clearly $L_{i+k} = R_i$. If there are no changepoints, so all $p_i \sim \mathcal{U}(0, 1)$, we expect that $\max_{1+k \leqslant i \leqslant m} L_i$ is small. An asymptotic Gumbel convergence result is presented in § 2·2. Large values of L_i or R_i can suggest that H_{0i} might be false. Section 2·3 provides an algorithm for identifying null and alternative hypothesis segments.

2·2. *Testing the global null hypothesis*

Theorem 1 provides the asymptotic distribution of a normalized version of $\Delta_m \equiv \max_{1+k \leqslant i \leqslant m} L_i$ under the global null hypothesis (1).

THEOREM 1. *Assume that $p_1, \ldots, p_m$ are independent identically distributed as $\mathcal{U}(0, 1)$ and*

$$k^{-1}(\log m)^3 + m^{-1}k \to 0, \quad m \to \infty. \tag{6}$$

Let $A_T = 2\log T + 2^{-1}\log\log T + \log(\pi^{-1/2})$ and $\Delta_m = \max_{1+k \leqslant i \leqslant m} L_i$. Then

$$(1/12)^{-1/2}\{2k\log(g_m)\}^{1/2}\Delta_m - A_{g_m} \to E \tag{7}$$

in distribution, where $g_m = m/k - 1$ and $\mathrm{pr}(E \leqslant x) = \exp(-2e^{-x})$, $x \in \mathbb{R}$.

Theorem 1 provides a criterion for testing the global null hypothesis (1) for large m. We reject the global null hypothesis at level $\alpha \in (0, 1)$ if

$$\max_{1+k \leqslant i \leqslant m} L_i \geqslant (24k\log g_m)^{-1/2}\{A_{g_m} - \log\log(1-\alpha)^{-1/2}\}. \tag{8}$$

Let $\gamma_{m,\alpha}$ be the $(1-\alpha)$th quantile of Δ_m under the global null hypothesis that $p_1, \ldots, p_m$ are independent identically distributed $\mathcal{U}(0, 1)$ random variables. Theorem 1 suggests that for very large m, $\gamma_{m,\alpha}$ can be approximated by the right-hand side of (8). For relatively small m, such as in the order of tens of thousands, one can approximate $\gamma_{m,\alpha}$ by sample quantiles of realizations of Δ_m via extensive Monte Carlo simulations.

2·3. *Locating null and alternative hypothesis clusters*

If the global null hypothesis is rejected, we are interested in finding the changepoints. This can be accomplished by using local discrepancy measure (5) through the following algorithm.

Algorithm 1. Estimating changepoints and locating null and alternative hypothesis segments.

Step 1. For a chosen window size k, calculate L_i and R_i in (5).

Step 2. For a prespecified cut-off value $\gamma > 0$, let $Q_i = 1(L_i > \gamma) + 1(R_i > \gamma)$.

Step 3. Decompose $\{1, \ldots, m\} = W_0 \cup W_1 \cup W_2$, where $i \in W_0$ if $Q_i = 0$, $i \in W_1$ if $Q_i = 1$ and $i \in W_2$ if $Q_i = 2$.

Step 4. Let $\mathcal{M}_1, \ldots, \mathcal{M}_{\hat{l}}$ be connected components of W_1 whose lengths are larger than $k/2$.

Step 5. For each $\mathcal{M}_i$, let $\hat{\tau}_i = \text{argmax}_{j \in \mathcal{M}_i}\{\max(L_j, R_j)\}$ be estimates of τ_i.

The rationale for our procedure is as follows. Recall that $\mathcal{S}_0 = \{\tau_0, \ldots, \tau_1 - 1\}$, $\mathcal{S}_2 = \{\tau_2, \ldots, \tau_3 - 1\}, \ldots$, correspond to null hypotheses, while the odd indexed sets $\mathcal{S}_1 = \{\tau_1, \ldots, \tau_2 - 1\}, \ldots$, are alternative hypotheses. For $i \in [\tau_0 + k, \tau_1 - k] \subset \mathcal{S}_0$, both L_i and R_i are close to zero, so with high probability $[\tau_0 + k, \tau_1 - k] \subset W_0$. For an appropriate interior interval of $\mathcal{S}_2$, if both L_i and R_i are larger than γ, then that interior interval is a subset of W_2. All the changepoints $\tau_1, \tau_2, \ldots$ are located in W_1. One can estimate the changepoints by analysing major connected components of W_1, namely whose lengths are larger than $k/2$, and maximizing R_j over j with $L_j < \gamma$, or vice versa. In other words, the changepoints occur when $Q_i = 1$. By using the mean of p-values under the global null hypothesis (1), we reduce the variance of the estimate in (4). We require that the cluster lengths of null and alternative hypotheses are larger than the sliding window length k.

The sequences $\{L_i\}$, $\{R_i\}$ and $\{Q_i\}$ can be computed within $O(m)$ steps. Thus the computational complexity of our algorithm scales linearly in m, so it can be executed very quickly, and it is quite attractive for dealing with testing problems with very large m.

In Algorithm 1 we require $\gamma < 1/2$, since otherwise both $L_i < \gamma$ and $R_i < \gamma$ hold trivially. We can choose γ as $\gamma_{m,\alpha}$ with $\alpha = 0{\cdot}05$, which plays the role of controlling Type I errors. Theorem 2 is a consistency result. It asserts that with probability approaching unity, the number of detected changepoints is the same as the number of true ones, and the detected changepoints and the true ones are uniformly close. The following technical conditions are needed. For $x \in \mathbb{R}$ let $\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \leqslant x\}$.

Condition 1. The changepoints are $\tau_i = \lfloor \eta_i m \rfloor (i = 1, \ldots, l)$, where $0 < \eta_1 < \cdots < \eta_l < 1$. In addition, there exists a constant $c > 0$ such that $\eta_i - \eta_{i-1} \geqslant c (i = 1, \ldots, l + 1)$, where $\tau_0 = 0$ and $\tau_{l+1} = 1$.

Condition 2. There exists a constant $0 < \rho < 1/2$ such that, under H_{1j}, $E(p_j) \leqslant \rho$.

Condition 1 is the condition that the changepoints are well separated. We assume that the number of changepoints is finite. Condition 2 makes it possible to distinguish alternative hypotheses from the null hypotheses. For two positive sequences (a_m) and (b_m), we say that $a_m \asymp b_m$ if there exists a constant $w > 0$ such that, for all large m, $w a_m \leqslant b_m \leqslant a_m / w$.

THEOREM 2. *Assume Conditions 1 and 2 hold. Let* $\gamma \asymp (k^{-1} \log m)^{1/2}$ *and assume that* $\gamma + \rho < 1/2$. *Then*

$$\text{pr}\left\{\hat{l} = l, \max_{i \leqslant l} |\hat{\tau}_i - \tau_i| \leqslant 2k\gamma(1/2 - \rho)^{-1/2}\right\} \geqslant 1 - 4k^{-1} m e^{-3k\gamma^2/(4+4\gamma)}. \tag{9}$$

Theorem 2 asserts the uniform closeness of our estimated changepoints to the real ones. As an immediate consequence, we have uniform consistency of the detected changepoints. Let $k = \lambda \log m$ in (9), where λ is a large constant, $\max_{i \leqslant l} |\hat{\tau}_i - \tau_i| = O_p(\log m)$. We do not assume that the p-value distributions are the same within the alternative hypothesis segment. In the context of multiple changepoint detection via the wild binary segmentation technique, using Theorem 3.1 in Fryzlewicz (2014), we can obtain the bound $O_p(\log m)$ under similar conditions.

Remark 1. A careful check of the proof indicates that Theorem 2 is still valid if Conditions 1 and 2 are relaxed in the following way: we can allow $c = c_m$ close to 0 and $\rho = \rho_m$ close to 1/2 such that $k^{-1} \log m = o(1/2 - \rho_m)$ and $k = o(c_m m)$.

Both our algorithm and theoretical results depend on the window size k. Our simulation studies in the Supplementary Material show that the performance of Algorithm 1 is quite robust to the choice of k. For large m it works reasonably well for a wide range of k that satisfies $(\log m)^2 \leqslant k \leqslant m^{1/2}$. In practice, we suggest the rule-of-thumb choice $k = \lfloor (\log m)^2 \rfloor$.

With the clustering property of the alternative hypotheses, by aggregating p-values, our procedure is better able to detect signals than procedures that evaluate p-values individually. Donoho & Jin (2004) considered testing the mixture model under independence:

$$H_0 : X_i \sim N(0, 1) \quad (i = 1, \ldots, m),$$
$$H_1^{(m)} : X_i \sim (1 - \epsilon)N(0, 1) + \epsilon N(\mu, 1) \quad (i = 1, \ldots, m).$$

Assume that, under $H_1^{(m)}$, a fraction $\epsilon = m^{-\beta}$, $\beta \in (1/2, 1)$, of the data come from $N(\mu, 1)$ with $\mu \neq 0$. By Donoho & Jin (2004), signal strength $\mu = \mu_m$ has to be at least $c\{\log(m)\}^{1/2}$ for some constant c in order to be detected. If signals are clustered, for example, letting $X_i \sim N(\mu_i, 1)$ with $\mu_i = \mu$ if $i \leqslant \tau = \lfloor m(1 - \epsilon) \rfloor$ and $\mu_i = 0$ if $i > \tau$. Choosing $k \asymp \lfloor (\log m)^2 \rfloor$, we can reject H_0 under the weaker condition $\{\log(m)/k\}^{1/2} = o(\mu_m)$.

2·4. *Adjustments with dependent p-values*

In (7) of Theorem 1, the quantity $1/12$ is the variance of p_i with independent $\mathcal{U}(0, 1)$ distribution. If $(p_1, \ldots, p_m)$ is a stationary process, we define the long-run variance $\sigma^2 = \sum_{k \in Z} \mathrm{cov}(p_0, p_k)$. The primary impact that dependence has on our testing procedure is that instead of using the marginal variance of p-values that follow $\mathcal{U}(0, 1)$, we need to use the long-run variance σ^2 to account for the dependence. In the Supplementary Material, we provide a framework of dependence and develop a related asymptotic theory.

3. APPLICATION

We apply our procedure to a genome-wide association dataset featured in Teslovich et al. (2010), which is freely available at http://www.sph.umich.edu/csg/abecasis/public/lipids2010/. We study the association between single nucleotide polymorphisms on chromosome 8, and HDL cholesterol. There are $m = 353\,488$ single nucleotide polymorphisms on chromosome 8, and we obtain p-values through marginal regression. We aim to identify regions of single nucleotide polymorphisms that are statistically significantly associated with HDL cholesterol. These identified single nucleotide polymorphisms can suggest follow-up studies and intervention strategies.

An examination of the dataset through the autocorrelation function calculation shows that there is substantial serial dependence among the p-values. To account for dependence, we need to estimate the long-run variance as indicated in § 2·4. We apply the batched mean estimate (Brockwell & Davis, 2009)

$$\hat{\sigma}^2(l_m) = \frac{l_m}{m - l_m + 1} \sum_{j=1}^{m-l_m+1} \left(l_m^{-1} \sum_{i=j}^{j+l_m-1} p_i - \bar{p}_m \right)^2,$$

where $\bar{p}_m = \sum_{i=1}^m p_i/m$, and l_m is the window size satisfying $l_m \to \infty$ and $l_m/m \to 0$. We use $l_m = 28$, which is between $m^{1/3}$ and $m^{1/4}$ and obtain $\hat{\sigma}^2 = 0{\cdot}8736$.

We use a simulation-assisted approach to compute the dependence-adjusted critical value $\gamma^*_{m,0\cdot05}$. For simplicity we write γ^* for $\gamma^*_{m,0\cdot05}$. Our simulation shows that the results are robust when k ranges between $\lfloor \{\log(m)\}^2 \rfloor$ and $\lfloor m^{1/2} \rfloor$; we use $k = \lfloor \{\log(m)\}^2 \rfloor = 163$. Specifically, we obtain 10^4 independent realizations of $A_m = 12^{-1} \max_{k+1 \leqslant i \leqslant m} L_i$ based on m independent $\mathcal{U}(0, 1)$ random variables. We then compute the empirical 95% quantile of these 10^4 realizations of A_m. Next we multiply this 95% empirical quantile by $\hat{\sigma}$ to obtain the critical value γ^*, which is 0·3422. For each single nucleotide polymorphism, we compute $Q_i = 1(L_i > \gamma^*) + 1(R_i > \gamma^*)(i = 1, \ldots, m)$. Then we apply Algorithm 1 to look for $\max(L_i, R_i)$ within the connected components where $Q_i = 1$ and to compute the changepoints.

Based on γ^*, loci 42 845, 43 046, 83 839, 84 431, 2 82 143, 2 82 345, 2 97 063 and 2 97 226 are identified as changepoints. The alternative hypothesis blocks are $\{42\,845, \ldots, 43\,045\}$, $\{83\,839, \ldots, 84\,430\}$, $\{2\,82\,143, \ldots, 2\,82\,344\}$ and $\{2\,97\,063, \ldots, 2\,97\,225\}$. The total number of rejections is 1158, which is

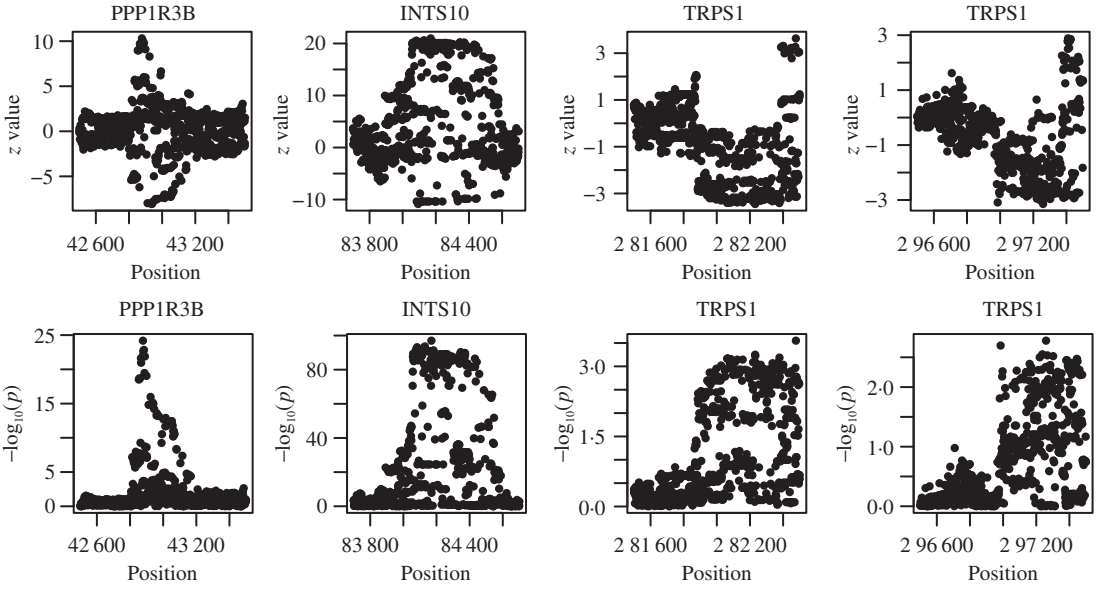

Fig. 1. Identified gene sets. Top panel: z-values; bottom panel: p-values.

roughly 0·3% of the total number of tests. The last two clusters belong to the same gene set. Our result agrees well with others in the literature: the closest genes to these regions are PPP1R3B (Teslovich et al., 2010), INTS10 (Wilke , 2011) and TRPS1 (Teslovich et al., 2010). Figure 1 plots z-values and p-values for regions around these estimated changepoints.

Acknowledgement

The authors thank Yunda Zhong and Ziwei Zhu for help in simulation studies and Xiang Zhou for providing the dataset and helping with the interpretation of our analysis results. We are indebted to the editor and reviewers for help in improving this paper. This research was supported in part by the University of Missouri Research Board and National Science Foundation.

Supplementary material

Supplementary material available at *Biometrika* online includes the proofs of Theorems 1–2, testing and estimation of changepoints under dependence and simulation studies.

References

Brockwell, P. J. & Davis, R. A. (2009). *Time Series: Theory and Methods*. New York: Springer.
Clarkes, S. & Hall, P. (2009). Robustness of multiple testing procedures against dependence. *Ann. Statist.* **37**, 332–58.
Donoho, D. L. & Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962–94.
Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42**, 2243–81.
Hao, N., Niu, Y. & Zhang, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist. Sinica.* **23**, 1553–72.
Huskova, M. & Slaby, A. (2001). Permutation tests for multiple changes. *Kybernetika* **37**, 605–22.
Jeng, X. J., Cai, T. T. & Li, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100**, 157–72.
Lai, T. & Xing, H. (2011). A simple Bayesian approach to multiple change-points. *Statist. Sinica* **21**, 539–69.

LAI, T., XING, H. & ZHANG, N. (2008). Stochastic segmentation models for array-based comparative genomic hybridization data analysis. *Biostatistics* **9**, 290–307.

NIU, Y. & ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Statist.* **6**, 1306–26.

OLSHEN, A., VENKATRAMAN, E., LUCITO, R. & WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–72.

SIEGMUND, D. O., YAKIR, B. & ZHANG, N. R. (2011a). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Statist.* **5**, 645–68.

SIEGMUND, D. O., ZHANG, N. R. & YAKIR, B. (2011b). False discovery rate for scanning statistics. *Biometrika* **98**, 979–85.

SUN, W., REICH, B., CAI, T., GUINDANI, M. & SCHWARTZMAN, A. (2015). False discovery control in large-scale spatial multiple testing. *J. R. Statist. Soc.* B **77**, 59–83.

TESLOVICH, T. M., MUSUNURU, K., SMITH, A. V., EDMONDSON, A. C., STYLIANOU, I. M., KOSEKI, M., PIRRUCCELLO, J. P., RIPATTI, S., CHASMAN, D. I., WILLER, C. J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–13.

TIBSHIRANI, R. J. & WANG, P. (2007). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9**, 18–29.

WILKE, R. A. (2011). High density lipoprotein cholesterol: leveraging practice-based biobank cohorts to characterize clinical and genetic predators of treatment outcome. *Pharmacogenomics* **11**, 162–73.

ZHANG, C., FAN, J. & YU, T. (2011). Multiple testing via FDR_L for large-scale imaging data. *Ann. Statist.* **39**, 613–42.

ZHANG, N. R., SIEGMUND, D. O., JI, H. & LI, J. (2010). Detecting simultaneous change-point in multiple sequences. *Biometrika* **97**, 631–45.

ZHAO, Z. & WU, W. B. (2007). Inference of trends in time series. *J. R. Statist. Soc.* B **69**, 391–410.

[*Received August* 2014. *Revised May* 2015]