

Rank Truncated Product of P -Values, With Application to Genomewide Association Scans

Frank Dudbridge^{1*} and Bobby P.C. Koeleman²

¹MRC Human Genome Mapping Project Resource Centre, Cambridge, UK

²Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands

Large exploratory studies are often characterized by a preponderance of true null hypotheses, with a small though multiple number of false hypotheses. Traditional multiple-test adjustments consider either each hypothesis separately, or all hypotheses simultaneously, but it may be more desirable to consider the combined evidence for subsets of hypotheses, in order to reduce the number of hypotheses to a manageable size. Previously, Zaykin et al. ([2002] *Genet. Epidemiol.* 22:170–185) proposed forming the product of all P -values at less than a preset threshold, in order to combine evidence from all significant tests. Here we consider a complementary strategy: form the product of the K most significant P -values. This has certain advantages for genomewide association scans: K can be chosen on the basis of a hypothesised disease model, and is independent of sample size. Furthermore, the alternative hypothesis corresponds more closely to the experimental situation where all loci have fixed effects. We give the distribution of the rank truncated product and suggest some methods to account for correlated tests in genomewide scans. We show that, under realistic scenarios, it provides increased power to detect genomewide association, while identifying a candidate set of good quality and fixed size for follow-up studies. *Genet Epidemiol* 25:360–366, 2003. © 2003 Wiley-Liss, Inc.

Key words: multiple tests; Bonferroni; false discovery rate

*Correspondence to: Frank Dudbridge, MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge CB10 1SB, UK.

E-mail: f.dudbridge@hgmp.mrc.ac.uk

Received 16 April 2003; Accepted 27 May 2003

Published online in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/gepi.10264

INTRODUCTION

Large exploratory studies such as genomewide association scans and microarray experiments are characterized by a preponderance of true null hypotheses, with a small though multiple number of false hypotheses. It may be more desirable to reduce the number of hypotheses to a manageable size, than to declare the truth of the individual hypotheses, because the combined evidence for multiple false hypotheses can be detected with greater power. Traditional multiple-test adjustments do not take this situation into account: for example, Bonferroni-like procedures [Hochberg, 1988] are based on individual hypothesis tests but do not consider combined evidence, whereas the Fisher product [Fisher, 1932] is motivated by a meta-analysis in which all hypotheses are simultaneously true or false. The false discovery rate is an alternative approach to the problem [Benjamini and Hochberg, 1995], but also does not consider combined evidence; it is equivalent to Simes' test [Simes, 1986] at the global level.

Some recent studies addressed these issues. Hoh et al. [2001] considered partial sums of order statistics. Though appropriate for genomewide scans, the approach has some practical drawbacks, including a reliance on permutation procedures, and difficulty in combining tests with different degrees of freedom. Zaykin et al. [2002] developed the truncated product method, which combines evidence from all tests whose significance exceeds some threshold. The method is applicable to a variety of situations, but here it can be difficult to choose the most appropriate threshold, as it should be chosen to maximize the evidence for the alternative, but is therefore very sensitive to the sample size and the number of hypotheses.

Here we propose a method complementary to the truncated product of Zaykin et al. [2002]. The *rank truncated product* is the product of the K most significant P -values from L hypothesis tests. We argue that this is more appropriate for the situation of genomewide scans, because the number of associated loci is fixed across studies, whereas their P -values are random. The

alternative hypothesis, that at least one of the K most significant tests is a true positive, corresponds more closely to the experimental situation where all loci have fixed effects. Furthermore, the choice of truncation point is rather less dependent on the experiment size. The approach combines the ideas of Hoh et al. [2001] with the advantages of Zaykin et al. [2002], and seems to provide increased power to detect genomewide association, while identifying a candidate set of good quality and fixed size for follow-up studies.

METHODS

RANK TRUNCATED PRODUCT

Suppose L tests are conducted, yielding P -values p_1, \dots, p_L . The order statistics of p_i are denoted by $p_{(i)}$. For a fixed integer $K, 1 \leq K < L$, the rank truncated product (RTP) is

$$W_R = \prod_{i=1}^K p_{(i)}.$$

Assume that the P -values are independent and uniformly distributed on $[0,1]$; situations where this is not the case are discussed later. Then when all null hypotheses are true, W_R has the exact distribution

$$\begin{aligned} \Pr(W_R \leq w) &= \binom{L}{K+1} (K+1) \int_t^1 (1-t)^{L-K-1} A(w, t) dt \end{aligned}$$

where

$$A(w, t) = \begin{cases} w \sum_{s=0}^{K-1} \frac{(K \ln t - \ln w)^s}{s!} & \text{when } w \leq t^K \\ t^K & \text{otherwise} \end{cases}.$$

The derivation of this distribution, with approximations for large L and K , is given in the Appendix.

The choice of the truncation point K depends on the experimental conditions. Certainly, if the number of false hypotheses were known in advance, and the sample size were sufficiently large, then K could be chosen to almost surely identify only the false hypotheses. In practice, we might reason an appropriate truncation point using biological arguments and economic considerations, such as the number of loci we wish to follow up. When the sample is finite and the effects are small, we expect some true hypotheses to be among the more significant tests, so K should be modified accordingly.

It is possible to optimize the significance of W_R over all values of K . This approach was proposed by Hoh et al. [2001] in a similar context, but we find that the cost of optimizing the significance tends to offset the power gained by truncation (data not shown). Furthermore, there may not be a locally smooth optimum, because the significances at different truncation points are random; and it is not obvious what this optimum actually represents. We prefer to fix K a priori, using arguments and simulations based on the experimental conditions.

RANK-AND-THRESHOLD TRUNCATED PRODUCT

The threshold truncated product method (TPM) proposed by Zaykin et al. [2002] is formed from all P -values smaller than a threshold τ :

$$W_T = \prod_{i=1}^L p_i^{I(p_i < \tau)}.$$

This approach can be combined with the RTP to give a dual truncated product. We form the product of at most K P -values which are less than τ :

$$W_{RT} = \prod_{i=1}^K p_{(i)}^{I(p_{(i)} < \tau)} = \max(W_R, W_T).$$

The distribution of W_{RT} is given in the Appendix. An advantage of this approach is that it could defend against either truncation point being too conservative, so there is greater margin for error in their choice. It is also appropriate when there is some natural truncation applied to the P -values, such as when the values are only reported when they reach some nominal significance.

CORRELATED TESTS

Correlations due to linkage disequilibrium are a particular problem for association scans, since the design relies upon such correlations to be successful. The RTP is affected by problems of redundancy, both in the ensemble of tests and in the choice of truncation rank. Zaykin et al. [2002] presented some methods for general correlation structure. Here we suggest some further methods for the situation of genomewide association. We note that linkage disequilibrium is a local phenomenon which decays over distance, but its magnitude and extent seem highly variable across the genome [Daly et al., 2001; Phillips et al., 2003], and a suitable characterization of the correlation

structure remains problematic [Maniatis et al., 2002].

Permutation tests are usually available to simulate the null hypothesis of interest, conditioning on the empirical correlation structure. These procedures could be used to estimate the effective number of independent tests, by determining what factor of Bonferroni correction coincides with the empirical significance of the smallest P -value. That is, we find L_e such that $L_e p_{(1)} \approx p_{perm}$. Let $L_e = \lambda L$, and the effective truncation point be $K_e = g(\lambda)K$. Assuming that g is linear and satisfies $g(1) = 1$ and $g(L^{-1}) = K^{-1}$, we have

$$K_e = \left\lceil \frac{L_e(K-1) + L - K}{L-1} \right\rceil$$

where $\lceil \cdot \rceil$ denotes nearest integer. The truncated product is then adjusted to $W^{K_e K^{-1}}$, and the analytic distribution can be used. In practice, we may prefer to specify K_e and then compute the RTP with

$$K = \left\lceil \frac{K_e(L-1) + L_e - L}{L_e - 1} \right\rceil.$$

Correlations may also be reduced by blocking analysis, in which we group the marker loci into LD blocks for which global tests of association are performed, e.g., using Hotelling's T^2 [Xiong et al., 2002; Fan and Knapp, 2003]. The blockwise tests are then combined in the RTP to identify the most associated blocks, with residual correlation accounted for as above.

RESULTS

We performed proof-of-concept simulations to assess the power of the RTP relative to the TPM of Zaykin et al. [2002], Simes' global test, the Fisher product, and the Bonferroni adjustment. We did not consider improved Bonferroni procedures [Hochberg, 1988], because they remain less well-known and the improvement is small when the number of tests is large. The dual truncated product was found to have approximately the same power as either the RTP or the TPM, depending on which threshold was the more strict in the particular model, though not necessarily the more powerful. For brevity, we do not show those results.

We simulated P -values from a two-tailed test of the log odds ratio in an unmatched case control design, using given formulas [Johnson et al., 2001]. We initially set all odds ratios to 1 to confirm the type I error rate of the various

TABLE I. Type-I error rates at $\alpha=0.05$ for L true hypotheses (10,000 simulations, exact test)

	$L=100$	$L=1,000$	$L=10,000$
RTP, $K=100$	0.0499	0.0473	0.0519
RTP, $K=10$	0.0482	0.0493	0.0509
TPM, $\tau=0.05$	0.0521	0.0475	0.0526
DTP, $K=100, \tau=0.05$	0.0521	0.0475	0.0519
DTP, $K=10, \tau=0.05$	0.05	0.0493	0.0509
Simes	0.0528	0.0523	0.0495
Bonferroni	0.0516	0.0507	0.0486
Fisher	0.0501	0.0436	0.0468

methods. Table I gives results for different truncation points and numbers of hypotheses. All values are close to the nominal level.

We then considered disease models consisting of 2, 5, 10, and 20 loci with allelic odds ratio 1.5. We assumed fixed marginal effects on disease risk, with no dominance. We used a single truncation point of $K=10$ in the RTP. For the TPM, we used both $\tau=\alpha$ and $\tau=K/2L$, an ad hoc choice to include roughly the same number of P -values as the RTP. We fixed the sample size such that the RTP achieved about 80% power at $\alpha=0.05$.

We compared the power with the total number of loci at 100, 1,000, and 10,000. All loci were diallelic with allele frequency 0.2, and in linkage equilibrium. Results are given in Tables II–IV. When there were two disease loci, the RTP had lower power than the Bonferroni and Simes tests, due to the overestimation of the best truncation point. However, for 5 and 20 disease loci, the RTP performed well despite the disparity between the truncation rank and the number of loci. Overall there was little difference in power between Bonferroni and Simes, in agreement with previous reports [Simes, 1986]. The power of the TPM improved as the proportion of disease loci increased, and the Fisher product had the highest power when the proportion of disease loci was very high. In the intermediate, more realistic situations, the RTP was at a higher power than the other methods, with a fair margin for error in choosing the truncation point.

Tables II–IV also show the RTP power when K was chosen to match the number of true loci. In these cases it was always the most powerful method, which shows that the cases where $K=10$ was not the most powerful were due to the choice of truncation point, rather than a general loss of power for those models.

TABLE II. Power to reject global null hypothesis at $\alpha=0.05$ for $L=100$ markers (10,000 simulations)

No. of true loci	2	5	10	20
No. of cases/controls	470	230	135	75
RTP, $K=10$	0.7905	0.8198	0.8191	0.7888
TPM, $\tau=0.05$	0.54	0.7053	0.7684	0.7856
TPM, $\tau=K/2L$	0.54	0.7053	0.7684	0.7856
Simes	0.8555	0.7041	0.574	0.4341
Bonferroni	0.8477	0.683	0.5472	0.4078
Fisher	0.4492	0.6192	0.7475	0.8287
RTP, $K=\text{no. of true loci}$	0.9019	0.8257	0.8191	0.8362

TABLE III. Power to reject global null hypothesis for $L=1,000$ markers (10,000 simulations)

No. of true loci	2	5	10	20
No. of cases/controls	675	375	250	165
RTP, $K=10$	0.8346	0.8332	0.8349	0.8168
TPM, $\tau=0.05$	0.2055	0.3361	0.503	0.6879
TPM, $\tau=K/2L$	0.4584	0.6388	0.7315	0.7738
Simes	0.8976	0.7687	0.6722	0.5483
Bonferroni	0.8927	0.7526	0.6494	0.5217
Fisher	0.1713	0.2732	0.4102	0.6084
RTP, $K=\text{no. of true loci}$	0.9241	0.8485	0.8349	0.8486

The alternative hypothesis for both truncation methods is that there is at least one false hypothesis among those whose P -values contribute to the product. This reduces the exploratory hypotheses to a small candidate set, although it does not give decisions on individual hypotheses. We are interested in the quality of this candidate set in terms of the numbers of true and false loci included, as an additional issue to the power, as this is important for designing follow-up studies. We defined the expected number of true loci (i.e., false hypotheses) identified, conditional on rejection of the global null, as

$$(1 - \beta)^{-1} \sum_{i=0}^L i\gamma(i)$$

where $\gamma(i)$ is the probability of the global null being rejected with exactly i true loci in the truncated product, and $1 - \beta = \sum \gamma(i)$. For the RTP, $\gamma(i)=0$ when $i > K$. A similar definition holds for the expected number of false loci identified. These measures vary with the sample size and significance level; for a realistic situation, we again fixed it such that the RTP achieved 80% power at $\alpha=0.05$.

We compared these measures with similar estimates for a single-step procedure using the Bonferroni adjustment, and by a false discovery rate step-up procedure (BH) [Benjamini and

TABLE IV. Power to reject global null hypothesis for $L=10,000$ markers (10,000 simulations)

No. of true loci	2	5	10	20
No. of cases/controls	825	500	375	275
RTP, $K=10$	0.8165	0.8	0.8418	0.8448
TPM, $\tau=0.05$	0.0947	0.1379	0.2042	0.3306
TPM, $\tau=K/2L$	0.3929	0.5627	0.7016	0.767
Simes	0.8854	0.7666	0.7262	0.6467
Bonferroni	0.88	0.7525	0.7085	0.6225
Fisher	0.0832	0.1141	0.1625	0.2525
RTP, $K=\text{no. of true loci}$	0.9153	0.8187	0.8418	0.8559

TABLE V. Expected numbers of true and false loci identified from $L=10,000$ markers, conditional on rejection of global null (10,000 simulations; sample sizes as in Table IV)

No. of true loci	2		5		10		20	
	True	False	True	False	True	False	True	False
RTP, $K=10$	1.93	8.07	3.55	6.45	4.73	5.27	5.45	4.55
TPM, $\tau=0.05$	2.0	537.42	4.85	534.75	9.22	529.15	16.52	522.07
TPM, $\tau=K/2L$	1.94	7.07	3.62	6.23	4.95	5.75	5.91	5.52
BH FDR	1.56	0.14	1.92	0.18	2.10	0.19	2.03	0.21
Bonferroni	1.48	0.06	1.57	0.07	1.58	0.07	1.45	0.08

Hochberg, 1995]. These are not strict comparisons, because the natures of the hypotheses are different; but the general goal is to proceed to follow-up studies with the greatest number of true positives and the least number of false positives.

Results for 10,000 markers are given in Table V. The RTP gave similar results to the TPM with $\tau=K/2L$, and its overall power is higher. The TPM with $\tau=0.05$ identifies more true loci, but with considerably more false loci. The RTP identifies most of the true loci when K is an overestimate, but it must include a number of false loci. When K is an underestimate, about half of the identified loci are true at this level of power; this proportion will increase up to 100% detection of true loci with the appropriate increase in sample size. The BH and Bonferroni procedures identify fewer true loci, and considerably fewer false loci, with reduced power compared to the RTP.

DISCUSSION

The rank truncated product is complementary to threshold truncation in the sense that they are appropriate for different situations. In a meta-analysis, especially in the presence of reporting bias, threshold truncation should be preferred. It is also relevant when we wish to relate nominally significant results to the familywise error. We

propose that rank truncation should be preferred when the aim is to detect a small set of fixed effects among a large number of null effects. Furthermore, it allows us to specify, in advance, the number of hypotheses to be explored in a follow-up study. These features make the RTP particularly suitable for genomewide association scans. It develops the ideas of Hoh et al. [2001], with the advantage that the analytic distribution is available, allowing preliminary simulations to explore different sample sizes and truncation points. Our results indicate that, by concentrating on the combined evidence from a fixed size subset of hypotheses, the RTP often has greater power than other methods to detect association.

The RTP does not declare on individual hypotheses. It can be expressed in a form which permits use of the closed testing procedure [Marcus et al., 1976], but we do not recommend this course because the familywise error is only bounded above by α [Hochberg and Tamhane, 1987], and may be much lower in practice. We prefer to reject the most significant hypotheses in an exploratory sense, giving a candidate set of ordered hypotheses for follow-up. In the situations we considered, the RTP detected more true loci with greater power than the other methods. More false loci were detected than by BH or Bonferroni, but in a genomewide scan this seems acceptable. Note that the loci detected by stepwise procedures contain or are contained in the set detected by RTP, and can be given priority for follow-up as they represent the most likely true loci. BH might be preferred if the priority is to control the *proportion* of false loci, but this is only guaranteed post hoc if the prior power is high [Zaykin et al., 2000]; and in this case, the RTP enjoys similar power to distinguish true from false loci. Furthermore, for a given underlying model, the truncation rank can be chosen to identify, on average, the same number of loci as a BH procedure with an arbitrary false discovery rate, while maintaining control of the type I error.

We caution against finding the truncation point with the strongest significance, both because the cost of optimization results in little net gain in power, and because it is unclear what the optimum actually represents. Stochastic variation in the significance of both true and false loci can result in an unclear optimum and poor correlation with the actual number of true loci. Although, in expectation, the optimum may be accurate, it

seems unwise to depend on a point estimate which is not well-understood. We prefer to fix the number of follow-up hypotheses, which could be compared with parallel studies to build a consensus set of candidate loci.

Our simulations assumed equal odds ratios for the disease loci. This is motivated by the assumption that for polygenic disease there are many low risk factors with very similar odds ratios. In the case of some loci with significantly greater odds ratios, the power will be dominated by the stronger effects, so our results apply just to that subset of loci. We anticipate that strong effects will be detected through traditional methods, and these loci should therefore be excluded so that the RTP is used to detect the combined weak effects of the minor loci.

We only considered two-tailed tests, since these are more likely to be used in genomewide scans, where the direction of association is unknown and varies across loci. In applications such as a meta-analysis, one-tailed tests may be more appropriate, since the evidence combined should be in the same direction. We have not explored the power of the RTP for further applications, and this remains an important question.

The analytic distributions assume that all P -values are uniform on $[0,1]$. Nonuniform P -values can occur in a number of situations. The random variables may be discrete, as in tests of small samples. The P -values may be estimated by simulation, e.g., when some cell counts are small [Sham and Curtis, 1995], resulting in discrete P -values. Also, some statistics have a positive probability of taking a boundary value, as occurs, for example, in tests of variance components which are constrained to be positive [Fulker and Cherny, 1997]. Indeed, the threshold truncated product itself has this property, because there is a positive probability that no P -values exceed the threshold, in which case the product is defined to be 1. In all of these cases, an analytic distribution can be obtained as before, provided that the joint distribution of P -values is known.

APPENDIX

DISTRIBUTION OF TRUNCATED PRODUCT

We give the distribution of the rank-and-threshold truncated product W_{RT} , which reduces to W_R and W_T in special cases. First note that, for $x \geq 0$

and positive integer k ,

$$\begin{aligned} \Pr(\chi_{2k}^2 \geq 2x) &= \int_{2x}^{\infty} \frac{u^{k-1} e^{-\frac{u}{2}}}{2^k \Gamma(k)} du \\ &= \frac{1}{(k-1)!} \int_x^{\infty} t^{k-1} e^{-t} dt \end{aligned}$$

where $t = \frac{u}{2}$

$$\begin{aligned} &= \frac{1}{(k-1)!} \left[x^{k-1} e^{-x} + (k-1) \int_x^{\infty} t^{k-2} e^{-t} dt \right] \\ &= e^{-x} \sum_{s=0}^{k-1} \frac{x^s}{s!} \end{aligned}$$

Suppose W_{RT} is the product of k independent uniform random variables on $[0,t]$. Then $\frac{W_{RT}}{t^k}$ is a Fisher product of uniform random variables on $[0,1]$, and

$$\begin{aligned} \Pr(W_{RT} \leq w|k, t) &= \Pr\left(\chi_{2k}^2 \geq -2 \ln \frac{w}{t^k}\right) \\ &= \frac{w}{t^k} \sum_{s=0}^{k-1} \frac{(k \ln t - \ln w)^s}{s!} \end{aligned}$$

if $w \leq t^k$, or 1 otherwise.

Now let truncation rank K and threshold τ be given. The truncation is by threshold when there are K or fewer P -values less than τ , which occurs according to a binomial distribution with probability τ . The truncation is by rank when there are at least $K + 1$ P -values less than τ , or equivalently, the $(K + 1)$ -th order statistic of the P -values is less than τ . When P_i are independent uniform random variables on $[0,1]$, the order statistic has the $\beta(K+1, L-K)$ distribution

$$\Pr(P_{(K+1)} = t) = \binom{L}{K+1} (K+1) t^K (1-t)^{L-K-1}$$

and the P -values less than $P_{(K+1)}$ are uniform on $[0, P_{(K+1)}]$. The distribution of W_{RT} is therefore

$$\begin{aligned} \Pr(W_{RT} \leq w) &= \sum_{i=1}^K \binom{L}{i} (1-\tau)^{L-i} A(w, \tau, i) \\ &\quad + I(K < L) \cdot \binom{L}{K+1} (K+1) \\ &\quad \times \int_0^{\tau} (1-t)^{L-K-1} A(w, t, K) dt \end{aligned}$$

where

$$A(w, t, k) = \begin{cases} w \sum_{s=0}^{k-1} \frac{(k \ln t - \ln w)^s}{s!} & \text{when } w \leq t^k \\ t^k & \text{otherwise} \end{cases}$$

The integral can be evaluated numerically. If $\ln(t)$ is approximated by a second-order Taylor series, then the integrand is of order at most t^{L+K-3} . We used the extended trapezoidal rule with $L+K$ abscissa. C code is available from the authors.

This general form reduces to W_R when $\tau=1$, W_T when $K=L$, the Fisher product when $\tau=1$ and $K=L$, and Šidák's correction $1 - (1-w)^L$ [Šidák, 1967] when $\tau=1$ and $K=1$.

When L is large, the binomial terms can be replaced by normal approximations, giving

$$\begin{aligned} \Pr(W_{RT} \leq w) &= \sum_{i=1}^K f\left(\frac{i - L\tau}{\sqrt{L\tau(1-\tau)}}\right) \tau^{-i} A(w, \tau, i) \\ &\quad + I(K < L) \cdot (K+1) \\ &\quad \times \int_0^{\tau} f\left(\frac{K+1-Lt}{\sqrt{Lt(1-t)}}\right) t^{-(K+1)} A(w, t, K) dt \end{aligned}$$

where $f(\cdot)$ is the standard normal density. By convention, the approximation is accurate when $L\tau > 5$ and K is large enough that $\Pr(LP_{(K+1)} < 5)$ is tolerably small. When K or i is large (say, > 50), we can use a central limit approximation to the χ^2 probability

$$\begin{aligned} t^{-k} A(w, t, k) &= \Pr\left(\chi_{2k}^2 \geq -2 \ln \frac{w}{t^k}\right) \\ &= 1 - \Phi\left(\frac{k \ln t - \ln w - k}{\sqrt{k}}\right). \end{aligned}$$

REFERENCES

- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc [B]* 57:289–300.
- Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. 2001. High-resolution haplotype structure in the human genome. *Nat Genet* 29:229–32.
- Fan R, Knapp M. 2003. Genome association studies of complex diseases by case-control designs. *Am J Hum Genet* 72:850–68.
- Fisher RA. 1932. *Statistical methods for research workers*. London: Oliver and Boyd.
- Fulker DW, Cherny SS. 1997. An improved multipoint sib-pair analysis of quantitative traits. *Behav Genet* 26:527–32.
- Hochberg Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75:800–2.
- Hochberg Y, Tamhane AC. 1987. *Multiple comparison procedures*. New York: John Wiley & Sons.
- Hoh J, Wille A, Ott J. 2001. Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res* 11:2115–9.
- Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG,

- Todd JA. 2001. Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233–7.
- Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, Ennis S, Ke X, Morton NE. 2002. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci USA* 99:2228–33.
- Marcus R, Peritz E, Gabriel KR. 1976. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63:655–60.
- Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, Ankener WM, Alfisi SV, Kuo FS, Camisa AL, Pazorov V, Scott KE, Carey BJ, Faith J, Katari G, Bhatti HA, Cyr JM, Derohannessian V, Elosua C, Forman AM, Grecco NM, Hock CR, Kuebler JM, Lathrop JA, Mockler MA, Nachtman EP, Restine SL, Varde SA, Hozza MJ, Gelfand CA, Broxholme J, Abecasis GR, Boyce-Jacino MT, Cardon LR. 2003. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 33:382–7.
- Sham PC, Curtis D. 1995. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 59:97–105.
- Šidák Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 78:626–33.
- Simes RJ. 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751–4.
- Xiong MM, Zhao J, Boerwinkle E. 2002. Generalized T^2 test for genome association studies. *Am J Hum Genet* 70:1257–68.
- Zaykin DV, Young SS, Westfall PH. 2000. Using the false discovery rate approach in the genetic dissection of complex traits: a response to Weller et al. *Genetics* 154:1917–8.
- Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. 2002. Truncated product method for combining P -values. *Genet Epidemiol* 22:170–85.