

# Evaluation of Nyholt's Procedure for Multiple Testing Correction

Daria Salyakina<sup>1</sup>, Shaun R. Seaman<sup>1</sup>, Brian L. Browning<sup>2</sup>,  
Frank Dudbridge<sup>3</sup>, Bertram Müller-Myhsok<sup>1</sup>

<sup>1</sup> Max-Planck Institute of Psychiatry, Munich, Germany

<sup>2</sup> GlaxoSmithKline, Research Triangle Park, North Carolina

<sup>3</sup> MRC Biostatistics Unit, Cambridge, UK

Corresponding author:

Daria Salyakina

Kraepelinstr. 2–10,

Max-Planck Institute of Psychiatry,

Munich,

Germany

Tel: +49 30622 222

Fax: +49 30622 601

email: [daria@mpipsykl.mpg.de](mailto:daria@mpipsykl.mpg.de)

Running title: Multiple Testing Evaluation

Key words: linkage disequilibrium, case-control study, spectral decomposition

# Abstract

## Objective

A simple method for accounting efficiently for multiple testing of many SNPs in an association study was recently proposed by Nyholt (2004), but its performance was not extensively evaluated. The method involves estimating an ‘effective number’ of independent tests and then adjusting the smallest observed p-value using Šidák’s formula based on this number of tests. We sought to carry out an empirical and theoretical evaluation of Nyholt’s method.

## Methods

Nyholt’s method was applied to a sample of 31 genes typed at a total of 291 SNPs and permutation used to determine the type-I error rate for each gene. **Based on our empirical results, we algebraically investigated the effective number of independent tests for a simple model of haplotype block structure.**

## Results

The nominal 5% type-I error rate varied from under 3% to over 7%, and was dependent on linkage disequilibrium. Theoretical considerations show further that the method can be very conservative in the presence of haplotype block structure.

## Conclusion

Although Nyholt’s approach may be useful as an exploratory tool, it is not an adequate substitute for permutation tests.

# 1 Introduction

Case-control studies are used to test for association between a trait, e.g. disease, and candidate genes or regions. A number of SNPs in the genes or regions are genotyped and then either each SNP tested individually for association or haplotypes derived and then association between trait and haplotypes tested. As many tests are being performed, it is necessary to take this into account when judging their statistical significance. Otherwise, the probability of obtaining at least one false-positive association (the family-wise type-I error rate,  $\alpha$ ) will potentially be much greater than its nominal value  $\alpha_{\text{nom}}$  (usually 0.05). The Bonferroni correction is a popular method for multiple-testing adjustment. Here,  $p_{\text{min}}$ , the minimum p-value from the set of tests is multiplied by  $M$ , the number of tests, and the global null hypothesis that no marker is associated is rejected when  $p_{\text{adj}} = Mp_{\text{min}} < \alpha_{\text{nom}}$ . This guarantees that  $\alpha \leq \alpha_{\text{nom}}$ . However, Bonferroni's method is conservative (i.e.  $\alpha < \alpha_{\text{nom}}$ ) when the tests are positively correlated, as is the case for SNPs in linkage disequilibrium (LD). An alternative formula is Šidák's correction,  $p_{\text{adj}} = 1 - (1 - p_{\text{min}})^M$  (Šidák 1967). This gives  $\alpha = \alpha_{\text{nom}}$  when the tests are independent, but is also conservative for SNPs in LD. Permutation tests (Churchill and Doerge 1994) yield the correct family-wise error rate even for dependent tests, but are less popular because they are more complicated to carry out and can be computationally intensive.

Nyholt (2004) proposed a new method for adjustment of the minimum p-value obtained from a set of individual SNP tests in an association study, based on a procedure described by Cheverud (2001) for multiple-testing adjustment in linkage studies. The advantage of Nyholt's method is its low computational requirement and hence rapidity. The procedure is as follows. The eigenvalues,  $\lambda_1, \dots, \lambda_M$ , of a matrix of observed pairwise LD between the  $M$  SNPs are obtained by principal components analysis or, more generally, spectral decomposition. The variance of

these observed eigenvalues,  $\text{Var}(\lambda)$ , is then used in the formula

$$M_{\text{eff}} = 1 + (M - 1) \left( 1 - \frac{\text{Var}(\lambda)}{M} \right) \quad (1)$$

to calculate an ‘effective’ number of independent tests,  $M_{\text{eff}}$ . This is then used in Šidák’s formula in place of  $M$ , the actual number of tests. Meng et al. (2003) used a similar approach for the selection of haplotype tagging SNPs.

Nyholt showed that his method, when applied to two datasets, yielded minimum p-values approximately equal to those estimated by permutation. However, these datasets were small, one containing 10 SNPs, and the other 23 SNPs, and the latter dataset was of limited use, as the intermarker LD was so low that the estimated  $M_{\text{eff}}$  (22.53) was almost equal to  $M$ , the number of SNPs.

Dudbridge and Koeleman (2004) investigated whether the assumption underlying Nyholt’s method, that there really is an ‘effective’ number of independent tests, is true. When  $b$  independent tests are carried out, the minimum p-value has a Beta(1,  $b$ ) distribution. Using data on chromosomes 18 and 21 from the International Hap Map Consortium, Dudbridge and Koeleman fitted a Beta( $a, b$ ) distribution to minimum p-values simulated from their null distribution by permutation and tested the null hypothesis that  $a = 1$ . By rejecting this hypothesis, they showed that the assumption of an effective number of independent tests is false.

The result of Dudbridge and Koeleman shows that Nyholt’s procedure cannot be exact, but does not indicate how well it performs as an approximate method in practice. Here, we describe a more extensive evaluation than that originally reported by Nyholt, using data on candidate genes from various studies in the Max-Planck Institute for Psychiatry on the genetics and pharmacogenetics of psychiatric disorders. We also present a theoretical investigation of the performance of Nyholt’s procedure when certain patterns of LD are assumed.

## 2 Methods

We considered 31 candidate genes genotyped in 1360 individuals, of whom approximately half were cases and half controls. Choosing only SNPs with minor allele frequency above 10% and in Hardy-Weinberg equilibrium, there were a total of 291 SNPs, i.e. an average of 9.4 SNPs per gene. None of the markers were significantly associated with case-control status, and hence we assumed that the LD patterns in the genes were the same in cases and controls. The program SNPSpD, described by Nyholt, was used to calculate  $M_{\text{eff}}$  for each gene.

Define, for each gene,  $M$  as the number of SNPs (and hence tests), the reduction factor as  $R = M/M_{\text{eff}}$ , and  $\bar{\Delta}$  as the mean absolute pairwise Pearson correlation coefficient between the SNPs in the gene (Devlin and Risch 1995) (estimated using the function LD of the R Genetics package). Thus  $R$  is a measure of how much more powerful Nyholt's method is compared to the usual Šidák correction, and  $\bar{\Delta}$  is a measure of LD within the gene. We examined the relation between  $\bar{\Delta}$  and  $R$  in the 31 genes.

The type-I error rates of Šidák's and Nyholt's procedures were investigated using simulation. Each of 20000 simulations was performed under the complete null hypothesis that none of the SNPs is associated with case-control status. Half of the 1360 individuals were randomly assigned to be cases and half to be controls. The association between case-control status and genotype was tested for each SNP using a chi-square test of the  $2 \times 3$  contingency table and, within each gene, the minimum p-value obtained. These minimum p-values were then adjusted using Šidák's correction based on both  $M$  (Šidák's original method) and  $M_{\text{eff}}$  (Nyholt's method) for that gene. For each gene, the proportion of the 20000 simulations in which the adjusted minimum p-value was less than 0.05 represents the 5% type-I error rate,  $\alpha$ , for the null hypothesis that none of the SNPs in that gene is associated with case-control status.

Next, we explored the effect of haplotype block structure on Nyholt’s procedure, using the method of Gabriel et al. (2002), implemented in the program Haploview (Barrett JC et al. 2004), to detect blocks within each gene.

Finally, as Dudbridge and Koeleman (2004) note, if the assumption that the minimum p-value corresponds to that of  $M_{\text{eff}}$  independent tests were true, it should be distributed  $\text{Beta}(1, M_{\text{eff}})$  when all null hypotheses are true. We fitted a  $\text{Beta}(a, b)$  distribution to the 20000 simulated minimum p-values for each of the 31 genes in turn, to determine how close the maximum likelihood estimates,  $\hat{a}$ , of  $a$  were to 1. We also fitted the beta distribution with  $a$  fixed at 1, and compared, for each gene, the maximum likelihood estimate,  $\hat{b}$ , of  $b$  with its corresponding  $M_{\text{eff}}$  value. A likelihood ratio test of the null hypothesis that  $a = 1$  was also performed for each gene.

### 3 Results

The average intermarker spacing within a gene was 8.6 Kb, with standard deviation (SD) 7.6 and range 0.37–33 Kb. The mean of  $\bar{\Delta}$  in the 31 genes was 0.44 (SD 0.20, range 0.11–0.84). There appeared to be an approximate quadratic relation between  $R$  and  $\bar{\Delta}$  (see Figure 1). Thus, the stronger is the LD within a gene the greater tends to be the increase in power of Nyholt’s method compared to the simple Šidák correction.

Figure 2 shows the type-I error rates for the 31 genes calculated using Šidák’s procedure (open circles and cross) and Nyholt’s procedure (closed circles and plus) plotted against  $\bar{\Delta}$ . As expected, for Šidák’s method there is a negative correlation: the more (positively) dependent the tests, the more conservative is Šidák’s procedure. Using Nyholt’s method all but five of the 31 genes had  $\alpha$  between 4% and 6%. For the 23 genes with  $\bar{\Delta} < 0.6$ ,  $\alpha$  was very close to 5% or slightly less, whereas for genes with  $\bar{\Delta} > 0.6$ , Nyholt’s method was somewhat anti-conservative and became increasingly so as  $\bar{\Delta}$  increased. The highest value of  $\alpha$  was 7.1% for a gene with  $\bar{\Delta} = 0.8$ .

The standard error of the 5% type-I error rate estimated from 20000 simulations is 0.15% when the true rate is 5%. For 19 of the 31 genes, the 95% confidence interval for the true type-I error rate excluded 5%.

Similar results were obtained when the 1% type-I error rate was examined.

Next, we used the program Haploview, to detect blocks in each gene. Only one gene, containing 46 typed SNPs, had more than one block: there were five blocks and four SNPs that did not fit into blocks. This multi-block gene was the gene with the lowest  $\alpha$  (3.7%) and among those with the smallest  $\bar{\Delta}$ . Nyholt's procedure was then applied to the five blocks of this gene separately.  $\bar{\Delta}$  within blocks was greater than  $\bar{\Delta}$  in the gene as a whole, as LD between SNPs in different blocks is weaker than between SNPs in the same block, and it was found that the five  $\alpha$  values for these blocks (closed triangles in Figure 2) were all greater than that for the gene as a whole (cross in Figure 2). For four of the five blocks the method was anti-conservative, and for one block  $\alpha$  was as high as 7.9%. The five blocks contained a total of 42 SNPs. The sum of the five  $M_{\text{eff}}$  values estimated in each block separately was 21.6. Adding the four SNPs not in blocks gave a total  $M_{\text{eff}}$  of 25.6. This compared with  $M_{\text{eff}} = 40.5$  estimated for the 46 SNPs together. Thus the reduction factor,  $R$ , is greater when the blocks are considered separately. Based on  $M_{\text{eff}} = 40.5$  (method 1) and 25.6 (method 2), respectively,  $\alpha$  was calculated as 0.037 and 0.060.

Eight chromosomes contained 23 of the 31 genes. The procedure used above for blocks within genes was now used for genes within chromosomes. That is, the type-I error rate for each chromosome (for the null hypothesis that none of the genotyped SNPs in the chromosome is associated with case-control status) was estimated based both on the  $M_{\text{eff}}$  estimated for all the genes on the chromosome together (method 1) and on the sum of  $M_{\text{eff}}$  estimated for each gene on the chromosome separately (method 2). Table 1 shows  $\bar{\Delta}$ ,  $R$  and  $\alpha$  for methods 1 and 2 and the eight chromosomes. The value of  $\bar{\Delta}$  calculated using all SNPs in a chromosome was naturally much smaller than the weighted average of the  $\bar{\Delta}$ s calculated within genes on that

chromosome. Likewise, the eight  $R$  and  $\alpha$  values calculated using method 1 were smaller than for those obtained from method 2. Using method 1 the mean values of  $R$  and  $\alpha$  were respectively 1.08 and 0.042, not very much different from a simple Šidák correction based on  $M$  tests, where  $M$  is the number of genotyped SNPs in the chromosome. Method 2 was better: mean values of  $R$  and  $\alpha$  were 1.27 (range 1.08–1.60) and 0.048 (range 0.041–0.055).

A beta( $a, b$ ) distribution was fitted to the simulated minimum p-values for each of the 31 genes. **The maximum likelihood estimates,  $\hat{a}$ , of  $a$  ranged from 0.91 to 1.01.** The mean was 0.97 and the interquartile range (0.95, 0.99). The same was done for the eight chromosomes. The mean of the eight  $\hat{a}$  values was 0.95; the interquartile range was (0.94, 0.97) and the range (0.92, 0.99). For 20 of the 31 genes and seven of the eight chromosomes the null hypothesis that  $a = 1$  was rejected at the 5% level. Thus, the assumption that the minimum p-values have a Beta(1,  $b$ ) null distribution for some  $b$  is not supported by the data, but the maximum likelihood estimates of  $a$  are not very different from 1. Table 2 shows the maximum likelihood estimates,  $\hat{b}$ , of  $b$  for the 31 genes and eight chromosomes when  $a$  was fixed at 1. The  $\hat{b}$  values for the 31 genes were on average 8% lower than the corresponding  $M_{\text{eff}}$  values, ranging from 40% lower (in the case of the multiblock gene) to 21% higher. For the eight chromosomes,  $\hat{b}$  was on average 24% less than  $M_{\text{eff}}$ , ranging from 40% lower to 6% higher. Thus,  $M_{\text{eff}}$  estimated by Nyholt’s procedure does not in general correspond to the best estimate of an effective number of independent tests.

## 4 Theoretical Investigation

The results above show that Nyholt’s method can perform poorly when applied to regions of DNA containing haplotype blocks, e.g. multiblock genes or multiple genes on the same chromosome. In this case, method 1 tends to overestimate  $M_{\text{eff}}$  and the power of the procedure is little different from that of a simple Šidák correction. An algebraic treatment of a very simple model example serves to illustrate why this



happens. Suppose there are  $2N$  SNPs with the first  $N$  in perfect LD ( $\bar{\Delta} = 1$ ) and the last  $N$  in perfect LD. Thus the effective number of independent tests is at most 2. Let  $C$  denote the correlation matrix for these SNPs, and let  $\delta$  be the Pearson correlation coefficient between a SNP in the first block and a SNP in the second block. The entries of the  $2N \times 2N$  correlation matrix  $C$  are

$$C_{i,j} = \begin{cases} 1 & \text{if } i, j \leq N \text{ or } i, j \geq N + 1; \\ \delta & \text{otherwise.} \end{cases}$$

The two largest eigenvalues of  $C$  are  $N(1 \pm \delta)$ . The eigenvalue  $N(1 + \delta)$  corresponds to an eigenvector whose entries are all 1, and the eigenvalue  $N(1 - \delta)$  corresponds to an eigenvector whose first  $N$  entries are 1 and whose last  $N$  entries are  $-1$ . Since the rank of  $C$  is 2, the remaining  $2N - 2$  eigenvalues are 0.

The sum of the eigenvalues of  $C$  equals  $2N$ , so the mean of the observed eigenvalues is 1. The variance of the observed eigenvalues is

$$\begin{aligned} \text{Var}(\lambda_{\text{obs}}) &= \frac{1}{2N - 1} [(N(1 + \delta) - 1)^2 + (N(1 - \delta) - 1)^2 + (2N - 2)(0 - 1)^2] \\ &= \frac{2N}{2N - 1} [N(1 + \delta^2) - 1], \end{aligned}$$

and the estimate,  $M_{\text{eff}}$ , of the effective number of independent tests is, from equation (1),

$$\begin{aligned} M_{\text{eff}} &= 2N - \left( \frac{2N - 1}{2N} \right) \text{Var}(\lambda_{\text{obs}}) \\ &= 2N - (N(1 + \delta^2) - 1) \\ &= (1 - \delta^2)N + 1. \end{aligned}$$

However, the true effective number of tests is at most 2, independent of  $N$ .

The preceding calculation can be generalized to  $K$  blocks of SNPs when any two markers from different blocks are in linkage equilibrium. Suppose the  $j$ th block contains  $2N_j$  SNPs with the first  $N_j$  in perfect LD ( $\bar{\Delta} = 1$ ) and the last  $N_j$  in perfect LD, and that the correlation between one of the first  $N_j$  SNPs and one of the last  $N_j$  SNPs is  $\delta_j$ . Let  $N = \sum N_j$ . In this case

$$M_{\text{eff}} = 2N + 1 - \sum_{j=1}^K \frac{N_j^2}{N} (1 + \delta_j^2). \quad (2)$$

We can now model a gene with 5 blocks and 8 SNPs per block by setting  $K = 5$ ,  $N_j = 4$ , and  $\delta_j^2 = 0.75$  (say) for  $j = 1, 2, \dots, 5$ . Now  $M_{\text{eff}} = 34$ , but the effective number of independent tests is less than 10. Thus, the estimated effective number of independent tests,  $M_{\text{eff}}$ , is at least a factor of 3.4 too large.

## 5 Discussion

In this article we have reported an empirical and a theoretical investigation of the performance of Nyholt's method. Nyholt's method is less conservative than Šidák's correction (although when applied to a longer length of DNA containing several sets of SNP markers in different haplotype blocks, it may be only slightly less conservative than Šidák), but it may also be anti-conservative when there is strong LD between SNPs. In view of this observation, we recommend that any significant result obtained using Nyholt's procedure which is not already significant using Šidák's original correction should be confirmed by a permutation test.

Nyholt's derivation of the formula for  $M_{\text{eff}}$  (equation (1)) is based on an interpolation of two extreme patterns of LD. When all SNPs are in linkage equilibrium, the eigenvalues,  $\lambda_1, \dots, \lambda_M$ , of the correlation matrix  $C$  are all equal to 1, so that  $\text{Var}(\lambda) = 0$  and  $M_{\text{eff}} = M$ . When all SNPs are in perfect LD, the eigenvalues are  $M, 0, \dots, 0$ , so that  $\text{Var}(\lambda) = M$  and  $M_{\text{eff}} = 1$ . Our theoretical investigation, however, shows that the correction is wrong when perfectly correlated and independent markers are mixed. In this situation,  $M_{\text{eff}}$  appears to be overestimated, so that the method is conservative.

In addition, as Dudbridge and Koeleman (2004) have shown for chromosomes 18 and 21 and we have shown for a sample of 31 candidate genes, the distribution of the minimum p-value from the permutation test is, in general, not equal to that of  $L$  independent tests for any integer  $L$ . A further sign that the theory underpinning Nyholt's correction is unreliable is that it yields different answers depending on whether it is applied to a set of markers  $S$  or to each disjoint subset of markers (e.g.

in a haplotype block or gene) whose union is  $S$ .

Nyholt (2004) mentions that his approach may be conservative in the presence of very strong LD. Our empirical results, on the other hand, show that it tends to become anti-conservative. The website of the SNPSpD program contains a note describing a way of overcoming this. The suggestion is to exclude all SNPs but one from any set of SNPs that are in perfect LD. This is, however, somewhat contrary to the theoretical justification for the method, given that it is based on the formula for  $M_{\text{eff}}$  being correct when all SNPs are in perfect LD. Moreover, it does not work when SNPs are in high but not perfect LD. For example, when there are  $K$  pairs of SNPs with the two SNPs in any pair having correlation  $\delta = 0.8$  and any two SNPs belonging to different pairs being in linkage equilibrium, equation (2) shows that  $M_{\text{eff}} = 2K - 0.8$ , which is unreasonably high when  $K$  is not small.

In the 31 genes we found only three genes containing SNPs in perfect LD. In each case, there was one pair of SNPs in perfect LD. When one SNP from each pair was removed, the type-I error rate improved slightly for two genes (from 3.7% and 4.5% to 3.8% and 4.8% respectively), but worsened for the other (from 7.1% to 7.3%).

Nyholt's method is similar to that of Chevrud (2001). The latter was developed for linkage studies, but could equally be used for association studies. Using Chevrud's method on our data, we obtained very similar results to those reported here. For most genes, the 5% type-I error rate was the same to an accuracy of one decimal place and it never differed by more than 0.3 (this was for one gene where the type-I error rate was 7.1% using Nyholt's method and 6.8% using Chevrud's method).

Finally, although the principle of an effective number of independent tests was proposed for the Šidák correction, there is no formal connection between the two. There seems no immediate reason to believe it should work better for a Šidák correction than for other methods that account for multiple testing. We investigated whether it could be used with Fisher's product p-value method. For  $M$  independent tests,  $T = -2 \log \left( \prod_{i=1}^M p_i \right)$  has asymptotically a  $\chi_{2M}^2$  distribution when all  $M$  null hy-

potheses are true (Fisher 1932). Analogous logic to that which led to equation (1) suggests that  $T \times M_{\text{eff}}/M$  may have an approximate  $\chi^2_{2M_{\text{eff}}}$  null distribution. However we found that, for the 31 genes, this assumption yielded 5% type-I error rates of between 5.4 and 12.1%, with a mean of 7.9%. Thus, the concept of an effective number of independent tests seems to be not at all applicable to Fisher's product method.

In conclusion, although Nyholt's method may be found useful as an exploratory tool, it is not a replacement for using a permutation test. It is worth noting that although permutation can be computationally intensive, methods are available for reducing the computational effort (Dudbridge and Koeleman 2004; Seaman and Müller-Myhsok 2004).

## **Acknowledgements**

The first two authors are supported by fellowships from the Max-Planck Institute for Psychiatry. This study was supported in part by the German Ministry of Education and Research (BMBF) within the National Genom Research Network (NGFN) and the Bavarian Ministry of Commerce.

## References

- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, in press
- Cheverud JM (2001) A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 87: 52-58
- Churchill GA, Doerge RW (1994) Empirical threshold values For quantitative trait mapping. *Genetics* 138: 963-971
- Devlin B, Risch N (1995) A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. *Genomics* 29: 311-322
- Dudbridge F, Koeleman BPC (2004) Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. *Am J Hum Genet* 75: 424–435
- Fisher RA (1932) *Statistical methods for research workers*. Oliver and Boyd, London
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225-2229
- Meng ZL, Zaykin DV, Xu CF, Wagner M, Ehm MG (2003) Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am J Hum Genet* 73: 115-130
- Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 74: 765–769
- Seaman SR, Müller-Myhsok (2005) Rapid simulation of p-values for product methods and multiple-testing adjustment in association studies. *Am J Hum Genet* (in press)
- Šidák Z (1967) Rectangular confidence regions for the means of multivariate normal

distributions. JASA 62: 626-633

## Table Legends

Table 1: **Type-I error rates ( $\alpha$ ), reduction factor ( $R$ ), and mean absolute LD ( $\bar{\Delta}$ ) for eight chromosomes estimated using Šidák method based on  $M$  tests ( $M = \text{no. SNPs}$ ), Nyholt’s method applied to genes on same chromosome together (Method 1), and Nyholt’s method applied to genes on same chromosome separately (Method 2). <sup>a</sup>‘Chr no.’ is chromosome number; <sup>b</sup>‘No. genes’ is number of genes on that chromosome; <sup>c</sup>‘No. SNPs’ is total number of SNPs genotyped in the genes on that chromosome.**

Table 2: For each gene and chromosome, numbers of SNPs ( $M$ ), effective number of independent tests estimated by Nyholt’s method ( $M_{\text{eff}}$ ), and maximum likelihood estimates of  $b$  when Beta( $1, b$ ) distribution fitted to distribution of minimum p-values ( $\hat{b}$ ).

## Figure Legends

Figure 1: Relation between mean LD,  $\bar{\Delta}$ , within a gene and reduction factor,  $R$ , for that gene. Shown are 30 single-block genes (circles) and one multi-block gene (closed circle) containing 5 blocks (closed triangles).

Figure 2: Relation between mean LD,  $\bar{\Delta}$ , within a gene and type-I error rate,  $\alpha$ , for Šidák’s and Nyholt’s methods. Open circles represent 30 single-block genes and Šidák’s correction; closed circles represent the corresponding Nyholt corrections. Cross, star and plus represent the multi-block gene, corrected using Šidák’s, Nyholt method 2 and Nyholt method 1 respectively. Triangles represent the five blocks of the multi-block gene: open for Šidák; closed for Nyholt.



Table 1

Chr no. <sup>a</sup>	No. genes <sup>b</sup>	No. SNPs <sup>c</sup>	Šidák	Method 1			Method 2		
			$\alpha$	$\alpha$	$R$	$\bar{\Delta}$	$\alpha$	$R$	$\bar{\Delta}$
1	2	11	0.045	0.049	1.11	0.19	0.055	1.24	0.41
4	2	16	0.044	0.045	1.03	0.10	0.048	1.08	0.19
5	5	48	0.039	0.041	1.05	0.10	0.047	1.22	0.35
7	2	20	0.042	0.045	1.07	0.15	0.048	1.14	0.28
9	3	57	0.035	0.037	1.09	0.16	0.041	1.21	0.30
10	2	17	0.038	0.042	1.10	0.17	0.046	1.21	0.34
11	4	19	0.034	0.038	1.11	0.15	0.054	1.60	0.63
17	3	34	0.033	0.038	1.13	0.19	0.047	1.43	0.50

Table 2

Gene	M	Meff	$\hat{b}$	Gene	M	Meff	$\hat{b}$
1	13	11.5	9.6	21	5	2.6	2.9
2	6	5.1	4.8	22	14	12.4	10.2
3	5	4.5	4.5	23	7	4.5	4.1
4	3	1.6	1.9	24	4	2.3	2.3
5	4	2.6	2.6	25	8	7.3	6.7
6	6	4.2	3.3	26	6	5.3	5.1
7	4	2.6	2.8	27	17	13.2	9.2
8	3	2.8	2.8	28	46	40.5	24.4
9	9	6.5	5.1	29	9	6.7	5.4
10	7	6.1	5.9	30	12	4.5	5.0
11	10	8.3	7.1	31	18	12.8	9.8
12	5	2.9	3.1	Chr			
13	15	11.3	8.5	1	11	10.0	8.8
14	5	4.6	4.4	4	16	15.5	14.6
15	5	3.6	3.7	5	48	45.8	34.0
16	8	5.1	4.3	7	20	18.7	15.7
17	11	9.1	7.2	9	57	52.1	31.3
18	5	3.9	3.7	10	17	15.5	12.2
19	10	9.8	9.7	11	19	17.2	11.6
20	11	10.1	9.3	17	34	30.1	19.6

Figure 1

Figure 2



