dures discussed here are known to control the familywise error rate or false-discovery rate in particular situations (e.g., independent covariates), their performance in more general situations needs further investigation.

PETER KRAFT
Departments of Epidemiology and Biostatistics
Harvard School of Public Health
Boston

References

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

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Bickel PJ, Doksum KA (1977) Mathematical statistics: basic ideas and selected topics. Prentice Hall, Englewood Cliffs, New Jersey

Bugawan TL, Mirel DB, Valdes AM, Panelo A, Pozzili P, Erlich HA (2003) Association and interaction of the IL4R, IL4, and IL13 loci with Type 1 diabetes among Filipinos. Am J Hum Genet 72:1505–1514

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York

Culverhouse R, Suarez BK, Lin J, Reich T (2002) A perspective on epistasis: limits of models displaying no main effect. Am J Hum Genet 70:461–471

Devlin B, Roeder K, Wasserman L (2003) Analysis of multilocus models of association. Genet Epidemiol 25:36–47

Sarkar S (1998) Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture. Ann Stat 26:494–504

Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. Biometrika 73:751–754

Address for correspondence and reprints: Dr. Peter Kraft, 665 Huntington Avenue, SBuilding 2, Room 109, Boston, MA 02115. E-mail: pkraft@hsph.harvard.edu

0002-9297/2004/7403-0027$15.00

Am. J. Hum. Genet. 74:584–585, 2004

Reply to Kraft

To the Editor:

Our study (Bugawan et al. 2003) reported a negative association of a specific IL4-524 haplotype with type 1 diabetes (T1D), consistent with a previous report (Mirel et al. 2002), and presented evidence for a genetic interaction between IL4-524 and IL4R SNPs. To test the latter, we computed relevant *P* values by permuting multilocus genotypes separately in case and control groups.

The criticism raised by Kraft (2004 [in this issue]) is not directed at our implementation of permutation testing, per se, but at permutation testing in general. His argument is that permutation testing does not properly account for multiple comparisons, resulting in an increase in false claims of significance, or type I familywise error (FWE). In the place of permutation testing, Kraft advocates the use of the Simes method—an elaboration of the classic Bonferroni procedure. In response, we wish to show that permutation testing can be used to obtain a desired false-positive error rate (as, indeed, can be demonstrated using Kraft's example) and, moreover, that such an approach has the added advantage of providing additional protection against false claims of nonsignificance, or type II error.

It should be noted that permutation methods are well established as a robust approach for obtaining overall significance levels while minimizing type II error (e.g., Good 1994; Doerge and Churchill 1996; Lynch and Walsh 1998), that such methods are extensible to multiple-testing scenarios (Westfall and Young 1993), and that examples of their application to human genetics are not uncommon (e.g., Lewis et al. 2003). However, as with any statistical method, the validity is dependent on correct application. Kraft provides an analysis of the permutation testing by discussing the distribution of two *P* values obtained from hypothetically permuted distributions (i.e., independent and uniformly distributed under the null hypothesis). The joint cumulative distribution function (CDF) for these two *P* values is given as $F(P_{(1)},P_{(2)}) = P_{(1)}(2P_{(2)} - P_{(1)})$, where $P_{(1)}$ and $P_{(2)}$ are, respectively, the first- and second-ordered *P* values. As such, Kraft notes that the $\Pr(P < .05)$ for this joint distribution is ~0.1, indicating that we would expect to see the smaller *P* value, or $P_{(1)} < .05$, about 10% of the time. Kraft's argument, therefore, is that for independent tests, use of a critical value of .05 leads to a type I error rate of 10%.

In fact, the proper approach for permutation testing—adjusted or unadjusted for multiple comparisons—is to find the critical value corresponding to the desired type I error rate. Specifically, if we consider the simulations presented by Kraft as equivalent to the result of a permutation test, we would seek the value of $\boldsymbol{x}$ in the permuted distribution for which $\Pr(P < \boldsymbol{x})$ is actually $\leqslant\alpha$ and would use that value, not the .05 value as Kraft appears to suggest. For $P_{(1)}$, this critical value would be .0253, as can be shown either by simulation or by solving Kraft's joint CDF for $\alpha = 0.05$, given $P_{(2)} = 1$ (in effect, solving the marginal CDF for $P_{(1)}$). It is interesting to note that the first *P* value that Kraft gives (.10) corresponds to the Sidak multiple comparison–adjusted *P* value for observed $\alpha = 0.05$ and $k = 2$ tests, whereas

the value we give corresponds to the Sidak-adjusted threshold ($1 - [1 - \alpha]^{1/k}$). As such, this example nicely illustrates that permutation testing, for two independent tests, yields familiar and contextually appropriate results.

It should also be noted that multiple-testing methods that rely on raw Bonferroni-type inequalities fail to incorporate correlation structures between tests. Therefore, although such methods (e.g., Simes 1986; Hochberg 1988; Rom 1990) provide control of FWE, they nevertheless are expected to be less powerful than methods that account for such dependencies. Indeed, these methods may be made more precise through resampling-based approaches (Westfall and Young 1993). In particular, the data from which the tests in table 7 (Bugawan et al. 2003) were derived are strongly correlated, and, therefore, tests that assume independence are not expected to be the most powerful. Moreover, Kraft fails to take into account the nonindependence of genotype distributions between chromosome 5 and chromosome 16 SNPs presented in table 6 (Bugawan et al. 2003). Applying the Simes correction suggested by the author for 10 comparisons (two sets: patients and controls, and five SNPs), the independence between IL4-524 and IL4R patient genotypes would be rejected with $P < .01$, supporting our conclusion of an interaction between chromosome 5 and chromosome 16 in T1D susceptibility.

In conclusion, what is needed, from a methodological perspective, are statistical procedures that adequately protect against false claims of significance while simultaneously addressing the correlated nature of multiple testing. The various methods discussed by Kraft address the former but do not address the latter. Having said this, whatever the statistical approach, the strongest test of the significance of any reported genetic interaction lies neither in initial-discovery *P* values nor in biologic plausibility—which we believe is high in this case—but in the ability to reproduce observations in independent cohorts.

Ana Maria Valdes, Brian Rhees, and Henry Erlich

Roche Molecular Systems
Alameda, CA

References

Bugawan TL, Mirel DB, Valdes AM, Panelo A, Pozzilli P, Erlich HA (2003) Association and interaction of the IL4R, IL4, and IL13 loci with type 1 diabetes among Filipinos. Am J Hum Genet 72:1505–1514

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142: 285–294

Good P (1994) Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer-Verlag, New York

Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75:800–802

Kraft P (2004) Multiple comparisons in studies of gene × gene, gene × environment interaction. Am J Hum Genet 74: 582–584 (in this issue)

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, et al (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am J Hum Genet 73:34–48

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associated, Sunderland, MA, pp 441–442

Mirel DB, Valdes AM, Lazzeroni LC, Reynolds RL, Erlich HA, Noble JA (2002) Association of IL4R haplotypes with type 1 diabetes. Diabetes 51:3336–3341

Rom DM (1990) A sequentially rejective test procedure based on a modified Bonferroni inequality. Biometrika 77:663–665

Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. Biometrika 73:751–754

Westfall PH, Young SS (1993) Resampling-based multiple testing: examples and methods for *P*-value adjustment. John Wiley & Sons, New York

Address for correspondence and reprints: Dr. Brian K. Rhees, Roche Molecular Systems, Department of Human Genetics, 1145 Atlantic Avenue, Alameda, CA 94501. E-mail: brian.rhees@roche.com

0002-9297/2004/7403-0028$15.00

Am. J. Hum. Genet. 74:585–588, 2004

Revisiting the Clinical Validity of Multiplex Genetic Testing in Complex Diseases

To the Editor:

The usefulness of genetic testing to identify high-risk patients for common multifactorial diseases is subject to debate. Optimism about the public health opportunities is counterbalanced with skepticism, since genetic factors appear to play a role in only a minority of patients with complex diseases, the number of genes involved is large, and their penetrance is incomplete (Holtzman and Marteau 2000; Vineis et al. 2001).

In last March's issue of the *Journal,* Yang and colleagues addressed the question of whether prediction of disease is improved by multiplex genetic testing (Yang et al. 2003). At first sight, their results seem promising. In a simulation study, they considered five genetic tests (g_1–g_5), which each could have a positive ($g_i = 1$) or negative result ($g_i = 0$). Yang et al. used the likelihood ratio to indicate the magnitude of change in disease probability before and after genetic testing. Positive test