4. Gibbons, R.J., Picketts, D.J., Villard, L. & Higgs, D.R. *Cell* **80**, 837–845 (1995).
5. Picketts, D.J. *et al. Hum. Mol. Genet.* **5**, 1899–1907 (1996).
6. Villard, L. *et al. Am. J. Hum. Genet.* **58**, 499–505 (1996).
7. Craddock, C.F. *et al. EMBO J.* **14**, 1718–1726 (1995).
8. Eisen, J.A., Sweder, K.S. & Hanawalt, P.C. *Nucleic Acids Res.* **23**, 2715–2723 (1995).
9. Aasland, R., Gibson, T.J. & Stewart, A.F. *Trends Biochem. Sci.* **20**, 56–59 (1995).
10. Gibbons, R.J. *et al. Am. J. Med. Genet.* **55**, 288–299 (1995).
11. Kurosawa, K., Akatsuka, A., Ochiai, Y., Ikeda, J. & Maekawa, K. *Am. J. Med. Genet.* **63**, 505–506 (1996).
12. Myers, R.M., Larin, Z. & Maniatis, T. *Science* **230**, 1242–1246 (1985).
13. Schwabe, J.W. & Klug, A. *Nature Struct. Biol.* **1**, 345–346 (1994).
14. Semenza, G.L. *Hum. Mutat.* **3**, 180–199 (1994).
15. Castilla, L.H. *et al. Nature Genet.* **8**, 387–391 (1994).
16. Schmeichel, K.L. & Beckerle, M.C. *Cell* **79**, 211–219 (1994).
17. Orlando, V. & Paro, R. *Curr. Opin. Genet. Dev.* **5**, 174–179 (1995).

Diabetes, dependence, asymptotics, selection and significance

Nature Genetics recently published two articles[1,2] identifying susceptibility genes for non–insulin-dependent diabetes mellitus (NIDDM), with subsequent comments by Daly and Lander[3] on statistical analyses. We appreciate the opportunity to comment on statistical issues and interpretation of these reports.

Hanis *et al.*[1] reported a *P* value of 10^{-6} for IBS analysis, which, Daly and Lander noted, was too small because its calculation ignored the dependence of sib pairs in sibships with more than two affected. On the basis of simulations, they suggested a *P* value of 2×10^{-5}.

Two components contribute to underestimation of *P* values in the presence of such dependencies: i) joint dependence in allele sharing among the sib pairs in a large sibship and ii) pairwise dependence. When alleles IBD are summed over all possible pairs in a sibship with more than two affected, the sharing is pairwise but not jointly independent. Hence, the mean and variance of the sharing are the same as if the pairs were from separate families, but the distribution of sharing is skewed instead of symmetric. This skewing affects the performance of the normal approximation and usually leads to underestimation of the *P* value. When alleles IBS are tallied among all pairs in a sibship with more than two affected, the resulting information is not even pairwise independent, and without proper adjustment, the variance of any test statistic on the IBS tallies will be underestimated, leading to further underestimation of the *P* value. The effect of skewing is negligible with a large number of sibships, but the underestimation of the variance remains. Because the data of Hanis *et al.* consist of many (247) independent sibships, skewing contributes little to the deviation of the actual from the reported *P* value, which is due mainly to the underestimation of the variance of the chi-squared statistic on the IBS tallies.

Related statistical issues arise in the study by Mahtani *et al.*[2], who report linkage analyses of six families with the lowest insulin secretion levels selected from an original sample of 26 NIDDM families. NPL scores of 4.1 and 4.2, respectively, are obtained, depending on whether deceased affected individuals are considered 'affected or unknown'. The corresponding exact *P* values calculated by GENEHUNTER[4] are 5×10^{-4} and 2×10^{-4}, both of which are substantially higher than the originally reported *P* value of 2×10^{-5} based on a normal approximation. Such a difference between the exact and asymptotic *P* values would not arise with large sample sizes, but occurs here because of skewing in the distribution of NPL scores due to the small number of families. Assessment of genome-wide significance is not straightforward, because with the small number of families there is no simple rule for equating genome-wide *P* values with single-test *P* values. Recent simulations (M.J.D. & E.S.L.) suggest that the genome-wide significance (uncorrected for selection of the families) is about 0.11, rather than the 0.05 originally reported.

Mahtani *et al.* noted that the selection of the six families requires additional correction of the *P* value. A Bonferroni correction factor of 2 to 4 was initially suggested, to correct for choosing the highest of the scores for the four quartiles (Z(6), Z(12), Z(18) and Z(24)). Based on simulations (M.J.D. & E.S.L.), the appropriate correction factor for this selection appears to be ~2.8. A genome-wide *P* value of 0.11 and a correction factor of 2.8 yields a corrected genome-wide significance of 0.28. A larger correction factor may be appropriate (corresponding to *P* of 0.3–0.4), given that the reported score is the highest for all values Z(n), and that the removal of the first six families with excess sharing results in a corresponding deficit of sharing in the unselected families (*t* statistic=–2.09, $P<0.05$).

The discussion above leads us to conclude that i) the *P* value of the IBS test for NIDDM1 should be increased from 10^{-6} to 2×10^{-5}, which still meets criteria for genome-wide significance[5] ($P=0.05$); and ii) the genome-wide *P* value for NIDDM2 should be increased from 0.05 to 0.11 (before correction for selection). The result does not meet criteria for genome-wide significance for an anonymous locus, but remains very interesting from a biological perspective because of the evidence for linkage in the region of the MODY3 gene, recently identified as *HNF1α* (ref. 6), which causes diabetes associated with low insulin secretion.

Both findings would be strengthened by confirmatory studies in other populations, but the key confirmations will be biological: the identification of mutations that increase the risk of NIDDM. While searching for mutations may be more straightforward for NIDDM2 than for NIDDM1, mutations in HNF-1α (or related sequences) may be found in only a subset of the six NIDDM2 families. If so, there is no reason to expect there to be an additional NIDDM susceptibility locus in this region in the remaining families.

Augustine Kong[1], Mike Frigge[1], Graeme I. Bell[2,3], Eric S. Lander[4], Mark J. Daly[4] & Nancy J. Cox[2]

Departments of [1]Statistics, [2]Medicine and [3]Biochemistry and Molecular Biology and Howard Hughes Medical Institute, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637, USA. [4]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.

1. Hanis, C.L. *et al. Nature Genet.* **13**, 161–166 (1996).
2. Mahtani, M.M. *et al. Nature Genet.* **14**, 90–94 (1996).
3. Daly, M.J. & Lander, E.S. *Nature Genet.* **14**, 131–132 (1996).
4. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
5. Lander, E.S. & Kruglyak, L. *Nature Genet.* **11**, 241–247 (1995).
6. Yamagata, K. *et al. Nature* **384**, 455–458 (1996).