

Supplementary Methods online). We also genotyped three SNPs, 001Msp (rs577001), 012Taq (rs237012) and 018Hha (rs237018) flanking 163A→G in the families from the UK, US, Finland, Romania, Norway and Northern Ireland and in the British case-control samples and obtained no evidence of association with T1D (Supplementary Tables 2 and 3 online).

Guo *et al.*² did report evidence of association from diverse ethnic groups, but the bulk of their supporting evidence came from European Americans, who would be well represented by the European American and British samples studied here (in fact, of 558 families, ~300 directly overlap). Our results were not influenced by age-at-onset in the cases ($P = 0.09, 0.47, 0.29$ and 0.23 for 163A→G, 001Msp, 012Taq and 018Hha, respectively) or in the affected offspring ($P = 0.87, 0.86, 0.90$ and 0.37 , respectively), as shown by regression analysis.

Our study, with 18,132 individuals (3,007 transmissions and 7,230 cases and controls; Supplementary Table 4 online), is much larger than the two previously published studies combined. Taking into account all the results, we conclude that there is no convincing evidence for the association of the 163A→G SNP with T1D. The previous result of Guo *et al.*² ($P = 2.9 \times 10^{-5}$) may be a false positive. In genetic association studies of common diseases, there is a very low prior probability of detecting a true positive result, given the large numbers of genes and polymorphisms in the genome, such that a P value on the order of 10^{-5} could still be false⁴⁻⁶. The problem of false positive results can be compounded by selection biases in the collection of samples, genotyping errors, population substructure and post-hoc subgroup analyses⁴⁻⁶. There are other explanations for the discrepant results obtained, such as complex gene-gene or gene-environment interactions. In the first instance, however, it is necessary to attempt to exclude the possibility that the initial finding is a false positive. Both Bohren *et al.* and Guo *et al.* reported evidence of functional differences between the 55M and 55V allotypes of SUMO4 (refs. 1,2). Functional studies of candidate genes and variants are essential in the study of common diseases, but they do not solve the problem of false positives.

URLs. Information from the UK GRID is available at <http://www-gene.cimr.cam.ac.uk/ucdr/grid.shtml>. Information from the 1958 British Birth Cohort is available at <http://www.cls.ioe.ac.uk/Cohort/Ncdfs/mainncds.htm>.

Note: Supplementary information is available on the Nature Genetics website.

Deborah J Smyth¹, Joanna M M Howson¹, Christopher E Lowe¹, Neil M Walker¹, Alex C Lam¹, Sarah Nutland¹, Jayne Hutchings¹, Eva Tuomilehto-Wolf², Jaakko Tuomilehto²⁻⁴, Cristian Guja⁵, Constantin Ionescu-Tirgoviste⁵, Dag E Undlien⁶, Kjersti S Ronningen⁷, David Savage⁸, David B Dunger⁹, Rebecca C J Twells¹, Wendy L McArdle¹⁰, David P Strachan¹¹ & John A Todd¹

- Bohren, K.M., Nadkarni, V., Song, J.H., Gabbay, K.H. & Owerbach, D. *J. Biol. Chem.* **279**, 27233–27238 (2004).
- Guo, D. *et al. Nat. Genet.* **36**, 837–841 (2004).
- Owerbach, D., Pina, L. & Gabbay, K.H. *Diabetes* **53**, 1890–1893 (2004).
- Wang, W.J.S., Barratt, B.J., Clayton, D.G. & Todd, J.A. *Nat. Rev. Genet.* **6**, 109–118 (2005).
- Thomas, D.C. & Clayton, D.G. *J. Natl. Cancer Inst.* **96**, 421–423 (2004).
- Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L. & Rothman, N. *J. Natl. Cancer Inst.* **96**, 434–442 (2004).

To the editor:

In the August 2004 issue, Guo *et al.*¹ reported a highly significant association of type 1 diabetes (T1D) with a haplotype encompassing the gene encoding SUMO4, a new member of the family of small ubiquitin-like modifiers with effects on IκBα action. The G allele of the 163A→G SNP (resulting in the amino acid substitution M55V) was overtransmitted to affected children with high statistical significance when all families were examined, but in a small subset of 92 British multiplex families, the A allele was nonsignificantly overtransmitted. In a simultaneously published paper, Bohren *et al.*² reported a highly significant overtransmission of the A allele in a cohort of 478 multiplex families, derived in almost equal parts from the Human Biological Data Interchange (mostly US cases) and the Warren Repository (UK). Guo *et al.* attribute the discrepancy between the European and UK families to genetic heterogeneity. What is notable is not the concept of locus heterogeneity in T1D but its marked subpopulation dependence and the possible genetic mechanisms and population structure that might explain these findings. Allelic heterogeneity in *cis* would imply that the 163A→G SNP is only a marker for some other functional variant and might have been obviously different in haplotype structure or allele frequencies in the British subset compared with the rest of the samples in the multiple SNP data that Guo *et al.* generated. The absence of comment to this effect may be interpreted

as indication that such difference was not found. Our data (Supplementary Fig. 1 online), generated by genotyping 163A→G in the 30 families from the Centre d'Étude du Polymorphisme Humain used in the HapMap project³, confirm that 163A→G is part of a solid linkage disequilibrium block with little internal structure. Locus heterogeneity in *trans*, involving allele frequencies at epistatically interacting loci, would be much more interesting biologically, much more difficult to map and equally puzzling in the context of European subpopulations.

As a first step towards generating data that will better define the question, we genotyped the 163A→G polymorphism a set of 1,188 Canadian families with at least one affected child. We found no transmission distortion from heterozygous parents ($P = 0.48$; Supplementary Table 1 online), as tested by the Family Based Association Test software⁴. Transmission was very close to the expected 50%: A:G = 387:376. Our power to detect a difference of the magnitude reported by Guo *et al.* at $\alpha = 0.05$ was >99%. This may mean that 163A→G had no effect on T1D risk in our population or that opposing effects due to allelic heterogeneity cancel each other out. Almost all of our subjects are of mixed European descent; a few are from the Middle East and Asia. The largest single homogenous group is a set of 538 pedigrees (1,995 individuals) from the province of Newfoundland, an isolated population descended almost exclusively from English and Irish settlers⁵. Given that both Guo *et al.* and Bohren *et al.* found an overtransmission of the 163A allele in their British samples, we tested the hypothesis that the 163A allele was overtransmitted in this set but the 163G allele was overtransmitted in the rest of our subjects. This was not the case: transmission in this subset was again equal (A:G = 125:117; $P = 0.31$) and not different from that in the rest of our set ($\chi^2 = 0.12$; $P = 0.73$). Although, because of missing parents, the number of informative transmissions in this subset was relatively low, we still had a power of 89% to detect a difference between subsets comparable to that found by Guo *et al.* at $\alpha = 0.05$. In any event, a genetic particularity in subjects of British descent does not seem to have masked association in our set.

Regardless of our results, the enigma of the opposing findings in the European American subjects of the two studies remains. Even in the absence of differences in ethnic background, population samples

may stratify because of other factors. For example, two samples of the same European American population could differ in age of the subjects at disease onset as a result of differing recruitment settings and strategies that differentially determine age at ascertainment (e.g., pediatric versus adult clinic). To test the hypothesis that age at onset of disease might explain differential effects of *SUMO4* alleles on T1D risk, we carried out transmission disequilibrium testing separately in two sets of subjects with onset before or after 8 years of age. Again, the two populations had similar transmission ratios and were no different from each other (A:G = 146:135 versus 136:124; $\chi^2 = 0.007$; $P = 0.935$).

Age at onset is only one possible difference between the two European American groups in which opposing effects were reported. The two population samples could still differ by many other factors that depend on setting and context of recruitment, such as socioeconomic status, environment, nutrition and exposure to pathogens. If they are not due to some kind of technical or statistical artifact, the contradictory observations of Guo *et al.* and Bohren *et al.* might offer an opportunity to study the interactions of a complex disease locus with other loci or environmental or demographic factors of a magnitude sufficient to cause the effect of an allele to switch from predisposing to protective.

Finally, the possibility remains that these contradictory findings do not represent real associations and were the result of chance alone. As hundreds of laboratories are testing thousands of candidate variants, statistical adjustment for multiple hypothesis testing within each individual research group may not be sufficient protection against publication of spurious results. One solution to this problem may be the centralization of sample resources for each disease into a repository whose size will provide the required power to all researchers in the field, such as the Type 1 Diabetes Genetics Consortium, which aims to collect data from thousands of families with T1D for genetic studies.

URLs. Family Based Association Test software is available at <http://www.biostat.harvard.edu/~fbat/fbat.htm>. The Type 1 Diabetes Genetics Consortium is available online at <http://www.t1dgc.org>.

Note: Supplementary information is available on the Nature Genetics website.

Huiqi Qu¹, Bhupinder Bharaj², Xiao-Qing Liu², Joseph A Curtis^{3,4}, Leigh Anne Newhook^{3,4},

Andrew D Paterson^{2,5}, Thomas J Hudson⁶ & Constantin Polychronakos¹

¹The McGill University Health Center (Montreal Children's Hospital), Montréal, Quebec, Canada. ²Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto, Ontario, Canada. ³Health Care Corporation of St. John's, Outreach, St. John's, Newfoundland, Canada. ⁴Discipline of Pediatrics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada. ⁵Departments of Public Health Sciences, and Psychiatry, and Institute of Medical Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. ⁶McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada. Correspondence should be addressed to C.P. (constantin.polychronakos@mcgill.ca).

1. Guo, D. *et al.* *Nat. Genet.* **36**, 837–841 (2004).
2. Bohren, K.M., Nadkarni, V., Song, J.H., Gabbay, K.H. & Owerbach, D. *J. Biol. Chem.* **279**, 27233–27238 (2004).
3. The International HapMap Consortium. *Nature* **426**, 789–796 (2003).
4. Horvath, S., Xu, X. & Laird, N.M. *Eur. J. Hum. Genet.* **9**, 301–306 (2001).
5. Newhook, L.A. *et al.* *Diabetes Care* **27**, 885–888 (2004).

To the editor:

In the August 2004 issue, Guo *et al.*¹ reported highly significant evidence for association between type 1 diabetes (T1D) and multiple polymorphisms in and near the gene *SUMO4*, encoding a new small ubiquitin-like modifier. The authors also showed that the 163A→G SNP (rs237025; resulting in the amino acid substitution M55V) influences immune responses by modulating NFκB activity. Both genetic and functional studies suggested that the 163G allele was associated with increased risk for T1D in a collection of European American families. But this conclusion was contradicted by the positive association with the 163A allele observed in the British data set studied by Guo *et al.*¹ and in a separate report² consisting primarily of a British data set. These inconsistent results raised the possibility that the reported association might be a false positive. Therefore, we tested this hypothesis in a Korean case-control cohort consisting of 386 individuals with T1D and 553 normal controls. The individuals with T1D were selected from the Korean Type 1 Diabetes Genetic Consortium^{3,4} using the criteria of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus⁵. All affected individuals were on insulin therapy upon hospital discharge. Their mean age was 13.4 years

(range 0.3–23.0 years). The nondiabetic control subjects had no family history of diabetes and were selected from the same geographical area. Their mean age was 39.2 years (range 18.1–80.7 years). The GG and AG genotypes had a higher frequency in affected individuals (62.0%) than in controls (52.1%), with a relative risk of 1.5 ($P < 0.003$; **Supplementary Table 1** online). Our results, consistent with the report by Guo *et al.*¹, provide additional support for an association between the *SUMO4* 163A→G SNP and T1D. Given the association differences across populations, it will be important to investigate the *SUMO4*-T1D association in other populations to understand the mechanisms responsible for these population differences, including gene-gene and gene-environment interactions.

Note: Supplementary information is available on the Nature Genetics website.

Yongsoo Park¹, Sukyung Park¹, Jungu Kang¹, Seiwon Yang² & Dukhee Kim³

¹Department of Internal Medicine, Hanyang University, College of Medicine, Seoul, Korea. ²Department of Pediatrics, Seoul National University, School of Medicine, Seoul, Korea. ³Department of Pediatrics, Yonsei University, School of Medicine, Seoul, Korea. Correspondence should be addressed to Y.P. (parkys@hanyang.ac.kr).

1. Guo, D. *et al.* *Nat. Genet.* **36**, 837–841 (2004).
2. Bohren, K.M. *et al.* *J. Biol. Chem.* **279**, 27233–27238 (2004).
3. Park, Y. & Eisenbarth, G.S. *Diabetes Metab. Res. Rev.* **17**, 2–11 (2001).
4. Park, Y. *Diabetes Res. Clin. Pract.* **66**, S19–S25 (2004).
5. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **23**, S4–S19 (2000).

In reply:

We reported highly significant evidence for association between type 1 diabetes (T1D) and the gene *SUMO4* (ref. 1). Two studies presented in this issue (Smyth *et al.* and Qu *et al.*) do not support our initial observation, but a third study in the Korean population provides confirmatory evidence (Park *et al.*). Such discrepant associations have been reported for many complex diseases, including T1D. The discrepancies may be caused by genotyping errors, random variation due to small sample size, spurious association, genetic heterogeneity or population differences in gene-gene and gene-environment interactions.

We ruled out genotyping errors by re-genotyping a subset of the samples.