Finally, Whitfield implies it is important to determine whether social factors or LD are responsible for the differences in effects of *ADH1B*47His* in eastern Asia and in Europe. Social factors can indeed be very important in modifying risk associated with different genotypes. For example, it will be difficult to determine the actual risk associated with ADH variation in northern Africa and the Middle East, since most populations in those regions are Muslim and consumption of alcohol is proscribed by their religion. However, we note that the relevant "genetic" component is not strictly differences in LD but differences in what haplotypes are present. A site in the ADH cluster but not in LD with the *ADH1B* Arg47His site could have an epistatic effect such that only those chromosomes with particular alleles in coupling account for the protective effect. Moreover, background genotype clearly differs at the population level between eastern Asian populations and European populations (e.g., Calafell et al. 1998) adding yet another level of confounding on the path to understanding the role of the *ADH1B* Arg47His polymorphism in risk of alcoholism.

KENNETH K. KIDD, MICHAEL V. OSIER, ANDREW J. PAKSTIS, AND JUDITH R. KIDD
Department of Genetics
Yale University
New Haven

References

Calafell F, Grigorenko EL, Chikanian AA, Kidd KK (2001) Haplotype evolution and linkage disequilibrium: a simulation study. Hum Hered 51:85–96

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1998) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49

Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi K, Okonofua F, Parnas J, Schulz LO, Bertranpetit J, Bonne-Tamir B, Lu R-B, Kidd JR, Kidd KK (2002) A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. Am J Hum Genet 71:84–99

Whitfield JB (2002) Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. 71:1247–1250 (in this issue)

Address for correspondence and reprints: Dr. Kenneth K. Kidd, Yale University School of Medicine, Department of Genetics, 333 Cedar Street, P.O. Box 208005, New Haven, CT 06520-8005. E-mail: kidd@biomed.med.yale.edu

Detecting Polymorphisms and Mutations in Candidate Genes

To the Editor:

Currently, there is no consensus in the literature as to the number and the nature of controls that should be studied to distinguish between polymorphisms and disease-causing mutations (Bridge 1997). The quandary becomes particularly acute when we are trying to determine if a missense alteration in a candidate gene is important (disease associated). How many control samples should be tested, and what other considerations should go into the selection of controls? It is important to consider and report the status, race or ethnic background, and sex (if appropriate) of controls. Furthermore, how many patients should be studied when one is screening a gene for mutations? We have attempted to address these concerns in this letter.

First, one needs to consider if the control subjects could have the same disorder as the case subjects. Where did the control subjects originate, and how were they selected? The use of convenient control subjects (newborn samples, unused diagnostic samples, etc.) may inadvertently include individuals who are carriers or affected. If one uses control subjects selected for a particular study, they may or may not be appropriate for a different study. When studying psychiatric disorders, one needs to ensure that the controls do not have undiagnosed problems. When studying late-onset diseases, one needs to confirm that the control subjects are past the age of onset.

Marchuk (1998) suggested typing controls from similar racial, ethnic, and geographic backgrounds, since allele frequencies can differ between groups. In the past, ignoring this important tenet has caused some mutations to be misclassified. The peripheral myelin protein 22 Thr118Met substitution was believed to be a mutation in Charcot-Marie-Tooth disease, but was found to be a Swedish polymorphism (Nelis et al. 1997). The fibrillin-1 P1148A substitution was initially considered to be a Marfan syndrome mutation in a mixed population of patients, because it had not been found in white or African American control subjects. However, it was later found to be a polymorphism in Asians (Wang et al. 1997). The homeo box A1 A218G polymorphism was reported to increase susceptibility to autism; however, it was found to be more common in African Americans than in whites (Collins et al., in press). Thus, one could misinterpret a negative result if only a single racial or ethnic group is utilized as a control population.

The sex of the control subjects is of obvious importance in testing for polymorphisms in X-linked genes.

Table 1

Sample Sizes Needed to Detect Polymorphisms

N[a]	Polymorphism Frequency	Alpha	Power
40	.05	.05	.80
65	.05	.05	.95
210	.01	.05	.80
340	.01	.05	.95
2,400	.001	.05	.80
3,910	.001	.05	.95

[a] N signifies the number of chromosomes, and this applies to either X-linked or autosomal diseases.

Often, the sex of the control subjects used is not mentioned in the literature. If one looks at chromosomes from normal females in X-linked mental retardation (XLMR), the significance of finding an alteration is unclear, because females are not likely to be affected by a change that could be pathogenic in a male. Therefore, it is imperative, in studying XLMR, to examine chromosomes solely from males of normal intelligence in a reference population and to cite this in any publication.

How many normal controls should be analyzed to detect a 5%, 1%, or 0.1% polymorphism? We used power calculations performed by the Power and Precision program (Biostat) to determine the number of chromosomes required to detect a significant difference between the polymorphism frequency in the reference population and the expected frequency. The polymorphism proportion in the hypothetical control group was set to 0.001% (as close to 0% as possible), since 0% of the controls would be expected to carry a disease-causing mutation. The alpha, or significance level, was set to 5%. The power, or percent of studies expected to yield a significant effect, was set to both 80% and 95%. Power is commonly set at 80%; however, at that level, a polymorphism would be missed 20% of the time. If a power of 95% were used, there would be only a 5% possibility of missing a polymorphism.

Table 1 displays the number of chromosomes that should be examined to significantly determine if the polymorphism frequency in the reference population differs from the expected frequency. The examination of a minimum of 65 chromosomes is necessary to detect a 5% polymorphism with 95% power. Therefore, 95% of the time, the polymorphism will be detected if it is in the population. For a 1% polymorphism, a minimum of 340 chromosomes should be examined. This number is close to Marchuk's (1998) proposal of typing 300–400 or more chromosomes to detect a 1% polymorphism. Finally, 3,910 chromosomes would be required to detect a 0.1% polymorphism with 95% power.

These numbers of chromosomes can also be applied to the search for mutations in disease genes. If each dis-

ease gene has been found to cause a certain percentage of a disease, one can utilize that information to determine how many affected individuals should be screened for mutations. For example, it would appear that each known XLMR gene accounts for ~1% of XLMR (Chelly and Mandel 2001). Therefore, a minimum of 340 unrelated males with XLMR should be tested to detect a single alteration in any candidate gene with 95% power.

In summary, when a potential mutation is detected, the status, race, and number of control subjects used for polymorphism detection need to be carefully considered. For X-linked conditions, the sex of controls used should also be taken into consideration. These characteristics are crucial to formulating accurate results. The sample sizes in table 1 can also be applied to the identification of candidate gene mutations in affected individuals.

Acknowledgments

JULIANNE S. COLLINS AND CHARLES E. SCHWARTZ
JC Self Research Institute
Greenwood Genetic Center
Greenwood, SC

References

Bridge PJ (1997) The calculation of genetic risks: worked examples in DNA diagnostics. (2nd ed) Johns Hopkins University Press, Baltimore, p 137

Chelly J, Mandel J-L (2001) Monogenic causes of X-linked mental retardation. Nat Rev Genet 2:669–680

Collins JS, Schroer RJ, Bird J, Michaelis RC. The HOXA1 A218G polymorphism and autism: lack of association in white and black patients from the South Carolina Autism Project. J Aut Devel Disord (in press)

Marchuk DA (1998) Laboratory approaches toward gene identification. In: Haines JL, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley-Liss, New York, pp 371–372

Nelis E, Holmberg B, Adolfsson R, Holmgren G, Van Broeckhoven CV (1997) PMP22 Thr(118)Met: recessive CMT1 mutation or polymorphism? Nat Genet 15:13–14

Wang M, Mathews KR, Imaizumi K, Beiraghi S, Blumberg B, Scheuner M, Graham JM, Godfrey M (1997) P1148A in fibrillin-1 is not a mutation anymore. Nat Genet 15:12

Address for correspondence and reprints: Dr. Julianne S. Collins, JC Self Research Institute, Greenwood Genetic Center, 1 Gregor Mendel Circle, Greenwood, SC 29646. E-mail: julianne@ggc.org