Letters to the Editor

Am. J. Hum. Genet. 67:1352–1355, 2000

On a Randomization Procedure

To the Editor:

Zhao et al. recently (1999) proposed a novel way of determining the statistical significance of a given test statistic in the context of allele-sharing methods. The procedure promises to be applicable to any pedigree structure and to both qualitative and quantitative traits and is based on a randomization approach. The classical randomization test, as introduced by Fisher (1935), has proved to be a widely applicable and powerful tool for geneticists and scientists in general. It requires fewer assumptions than many other standard tests and has appealing small-sample properties, as the *P* values it produces can be claimed to be "exact." As flexible as the classical randomization procedure is, its validity depends crucially on certain symmetry/exchangeability conditions. For example, in a case/control study, a "case" and a "control" are exchangeable under the null hypothesis. Close inspection reveals that the method of Zhao et al., which generates a distribution of the NPL score (Kruglyak et al. 1996) by randomization of the conditional probabilities of different inheritance vectors in a specific way, does not, in general, satisfy such exchangeability conditions. As a consequence, it cannot be assumed automatically that this procedure will share the appealing properties of the classical randomization test. Indeed, as demonstrated below, not only can *P* values that are computed with a small sample be very misleading, but the results for large samples can be off systematically as well. For example, for affected full-sib pairs with missing data on the parents—a common design for late-onset diseases—the method underestimates the variance of the NPL score by a factor of 2 asymptotically, an effect that corresponds to inflating a Z_{stat} (a test statistic that has, asymptotically, a standard normal distribution) by a factor of $\sqrt{2}$.

We start by reviewing how the randomization procedure works with the family structure of two parents and two affected sibs and the scoring function S_{pairs}. In theory, the inheritance vector has four "bits," one paternal bit and one maternal bit for each of the two sibs. For example, the paternal bit of the first sib indicates whether the allele he or she inherited from the father originates from the paternal grandfather or the paternal grandmother. However, with no data on the grandparents, there can be no information on an individual bit. The information is on whether the paternal bits of the two sibs are the same, which corresponds to whether the two sibs have the paternal allele IBD (identical by descent), and on the same information with the maternal bits. Hence, to understand/describe what the randomization procedure is doing, one can focus on a reduced vector with two bits, one paternal sharing bit (one for IBD sharing, zero for not sharing) and one maternal sharing bit. This sharing inheritance vector has four possible states: (0,0) corresponds to the two sibs sharing zero alleles IBD, (1,1) corresponds to sharing both alleles IBD, (1,0) corresponds to IBD sharing of the paternal allele but not of the maternal allele, and (0,1) corresponds to not sharing the paternal allele but sharing the maternal allele. If the sharing vector can be determined without uncertainty, then (1,1) gives an NPL score of $\sqrt{2}$, (0,0) gives an NPL score of $-\sqrt{2}$, and both (1,0) and (0,1) give an NPL score of 0. In general, with incomplete information, the NPL score for the pair is defined as

$$\begin{aligned} V(1) &= (\sqrt{2})p(1,1) + (-\sqrt{2})p(0,0) \\ &\quad +(0)[p(0,1) + p(1,0)] \\ &= (\sqrt{2})[p(1,1) - p(0,0)] \ , \end{aligned}$$

where $p(\cdot,\cdot)$ are the conditional probabilities of the various configurations of the sharing vector, given the marker data. Apart from the actual NPL score $V(1)$, three other hypothetical NPL scores are generated by the randomization procedure by flipping one or both bits:

$$\begin{aligned} V(2) &= (\sqrt{2})p(0,1) + (-\sqrt{2})p(1,0) \\ &\quad +(0)[p(1,1) + p(0,0)] \\ &= (\sqrt{2})[p(0,1) - p(1,0)] \ , \end{aligned}$$

obtained by flipping the paternal sharing bit;

$$V(3) = (\sqrt{2})p(1,0) + (-\sqrt{2})p(0,1) + (0)[p(0,0) + p(1,1)] = (\sqrt{2})[p(1,0) - p(0,1)] ,$$

obtained by flipping the maternal sharing bit; and

$$V(4) = (\sqrt{2})p(0,0) + (-\sqrt{2})p(1,1) + (0)[p(1,0) + p(0,1)] = (\sqrt{2})[p(0,0) - p(1,1)] ,$$

obtained by flipping both sharing bits. Note that $V(4) = -V(1)$ and $V(3) = -V(2)$. The four values are given equal probabilities by the procedure when it is applied to generate a randomization distribution.

Here, consider the case in which there are no genotype data on the parents of the affected sibs. In this case, it is obvious that the data cannot distinguish (0,1) from (1,0), hence $p(0,1) = p(1,0)$ and $V(2) = V(3) = 0$, a result that has serious consequences. Suppose the data consist of n affected sib pairs with no genotype data on the parents. For $i = 1, \ldots, n$, let W_i be the NPL score for sib pair i, so that the overall NPL score is

$$W = \frac{\sum_{i=1}^{n} W_i}{\sqrt{n}} .$$

Let w_i be the observed value of W_i, and define $X_i, i = 1, \ldots, n$, as independent random variables with discrete distributions $P(X_i = w_i) = 1/4, P(X_i = -w_i) = 1/4$, and $P(X_i = 0) = 1/2$, and

$$X = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} .$$

The P value determined by the randomization procedure is $P(X \geq w)$, where w is the observed value of W. As a small-sample example, consider $n = 10$ and the only genotype data is a single biallelic marker with alleles A and a. Let p and $q = 1 - p$ be respectively the population frequencies of A and a. Suppose for each of the 10 sib pairs, the two sibs have two alleles identical by state (IBS). In this case, w_i is positive for all 10 pairs, and it is easily seen that the randomization P value is

$$P(X \geq w) = P(X = w) = P(X_i = w_i \forall i) = (1/4)^{10} \approx 9.5 \times 10^{-7} .$$

This value is obviously too small, since it is the right P value when the results are IBD instead of IBS; it is also suspicious that this value does not depend on the allele frequencies p and q. Indeed, the probability that all 10 sib pairs have two alleles IBS within pairs is

$$\left\{\frac{1}{4} + \frac{1}{2}\left(p^2 + q^2\right) + \frac{1}{4}\left[\left(p^2 + q^2\right)^2 + \left(1/2\right)\left(2pq\right)^2\right]\right\}^{10} ,$$

which is equal to .0054, .0067, .0127, .0366, and .1592, respectively, for $p = .5, .6, .7, .8$, and $.9$. Obviously, the values of w_i depend on the allele frequencies. However, in this example, because the values of w_i are all positive, the randomization P value is not sensitive to their absolute values. Hence, this can only be considered as a small-sample example, since, with large-sample examples, some values of w_i will be negative and the allele frequencies will have an effect on the answer. For the large-sample behavior of the randomization procedure, note that X_i has mean 0 and variance $w_i^2/2$, and so X has mean 0 and variance $(\sum_i w_i^2)/(2n)$. It follows that the distribution of $X/\sqrt{\sum_i w_i^2/(2n)}$ can be approximated by a standard normal distribution, and the randomization P value,

$$P(X \geq w) = P\left[\frac{X}{\sqrt{\sum_i w_i^2/(2n)}} \geq \frac{w}{\sqrt{\sum_i w_i^2/(2n)}}\right] ,$$

can be approximated by

$$1 - \Phi\left[\frac{w}{\sqrt{\left(\sum_i w_i^2\right)/\left(2n\right)}}\right] ,$$

where $\Phi(\cdot)$ denotes the cumulative distribution of the standard normal. In other words, asymptotically, the randomization procedure corresponds to a method that treats

$$Z^* = \frac{W}{\sqrt{\left(\sum_i W_i^2\right)/\left(2n\right)}}$$

as a statistic that has a standard normal distribution under the null hypothesis. However, under the null hypothesis, $E(W_i) = 0$ and $\mathrm{Var}(W_i) = E(W_i^2)$. Asymptotically $(n \to \infty)$,

$$\frac{\sum_i W_i^2}{\sum_i \mathrm{Var}(W_i)} = \frac{(1/n)\sum_i W_i^2}{\mathrm{Var}(W)} \to 1 .$$

with probability 1, and

$$Z_{\text{adj}} = \frac{W}{\sqrt{\left(\sum_i W_i^2\right)/n}} = \frac{\sum_i W_i}{\sqrt{\left(\sum_i W_i^2\right)}}$$

has, asymptotically, a standard normal distribution under the null hypothesis. A discussion concerning Z_{adj} and other test statistics that are asymptotically valid can be found in Teng and Siegmund (1998) and Nicolae et al. (1998). The key here, however, is to note that

$$Z^* = \sqrt{2}Z_{\text{adj}} .$$

So when $Z_{\text{adj}} = 2$, which gives a P value of $1 - \Phi(2) = 0.023$, the randomization procedure will give a P value that is approximately $1 - \Phi(2.83) = 0.0023$.

Recall that the large-sample behavior of the randomization procedure presented above is based on the case of affected sib pairs with no data on the parents. In general, the large-sample behavior of the procedure depends on both the family structure and the missing data patterns. For example, it can be shown that, for affected half sibs, the randomization procedure is calibrated for large samples and is asymptotically similar to using Z_{adj}. Although, as demonstrated, the procedure is anticonservative for sib pairs with no data on parents, it can be shown that—at least for the single-marker case—it is asymptotically slightly conservative for sib-pair data with genotypes on both parents. Real data sets tend to have a mixture of family structures and missing data patterns, and, hence, there is no simple way to make adjustments. Zhao et al. (1999) found that their randomization procedure gives smaller P values than the likelihood methods of Kong and Cox (1997) in most of the examples they looked at. Since the likelihood methods are asymptotically efficient, given a specific model, and are asymptotically equivalent to other methods that are efficient (Cox and Hinkley 1974), this suggests that the randomization procedure might be anticonservative in many of the examples.

To gain some understanding of why this randomization procedure does not, in general, give exact P values, it may help to consider the special situation where it does. Suppose we have sib-pair data and are always able to determine the sharing vector with no uncertainty. This means that for a pair, given the data, one of $p(1,1)$, $p(0,1)$, $p(1,0)$ and $p(0,0)$ is equal to 1. One can see that the four values $V(1)$, $V(2)$, $V(3)$, and $V(4)$ will always be some permutation of $\sqrt{2}$, 0, 0, and $-\sqrt{2}$. Hence, the four values of V always correspond to the four possible values of the NPL score. In addition to the values, the randomized distribution generated is exactly the same as the distribution of the NPL score under the null hypothesis. In general, with complete descent information, the randomization procedure gives valid exact P values that are the same as those obtained by direct simulation and the "exact P values" of GENEHUNTER (Kruglyak et al. 1996). This might have been the scenario which stimulated the development of the procedure. For comparison, consider the classical randomization procedure in a matched-pairs study. Within a pair, the procedure permutes the responses of the case and the control. The idea is that if we are given the two response values of the case and the control, but not the correspondence between the subjects and the responses, then the two permutations have the same probability under the null hypothesis. Hence, the classical randomization test can be considered as a conditional test that conditions on the observed response values without the correspondences. The randomization distribution of Zhao et al., in general, cannot be interpreted as any conditional or unconditional distributions of the outcome. Indeed, consider sib pairs with no data on the parents. If the two sibs have 0 alleles IBS, then $p(0,0) = 1$, $V(1) = -\sqrt{2}$, and the hypothetical value $V(4) = \sqrt{2}$. But $\sqrt{2}$ is not even a possible outcome, since, with no data on the parents, the NPL score generally will be positive but smaller than $\sqrt{2}$, even if the two sibs have two alleles IBS. So, one way to understand why the randomization procedure here does not give exact P values is that, although the different configurations of the inheritance vector have some obvious exchangeability properties for complete information, the same symmetry does not hold for every missing data pattern. It is unfortunate that this lack of symmetry affects not only the small-sample properties, but also the large-sample behavior.

AUGUSTINE KONG[1,3] AND DAN L. NICOLAE[2]
Departments of [1]Human Genetics and [2]Statistics, The University of Chicago, Chicago; and [3]deCODE genetics, Reykjavík

References

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman & Hall, London

Fisher RA (1935) The design of experiments. 1st ed. Oliver and Boyd, Edinburgh

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Nicolae DL, Frigge ML, Cox NJ, Kong A (1998) Discussion of Teng and Siegmund (1998). Biometrics 54:1271–1274

Teng J, Siegmund DO (1998) Multipoint linkage analysis using

affected relative pairs and partially informative markers. Biometrics 54:1247–1265

Zhao H, Merikangas KR, Kidd KK (1999) On a randomization procedure in linkage analysis. Am J Hum Genet 65: 1449–1456

Address for correspondence and reprints: Dr. Dan Nicolae, Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637. E-mail: nicolae@galton.uchicago.edu

0002-9297/2000/6705-0035$02.00

Am. J. Hum. Genet. 67:1355–1356, 2000

Reply to Kong and Nicolae

To the Editor:

We thank Kong and Nicolae (2000) for their insightful discussion of our proposed randomization procedure for linkage analysis (Zhao et al. 1999). In light of the example in their discussion, we agree that our proposed method is anticonservative for nuclear families when both parents are missing. For families of other structures, Kong and Nicolae stated that "it can be shown that, at least for the single marker case, it is asymptotically slightly conservative for sib-pair data with genotypes on both parents." They further stated that, "In general, with complete descent information, the randomization procedure gives valid exact *P* values that are the same as those obtained by direct simulation and the 'exact *P* values' of GENEHUNTER (Kruglyak et al. 1996)." We agree that, with complete descent information, the randomization procedure gives valid statistical inference. However, we do not think that the results from the randomization procedure are the same as those obtained by direct simulation and the "exact *P* values" of GENEHUNTER. In this letter, we use examples to illustrate the differences between our proposed randomization procedure and the two alternative methods—direct simulation and the perfect data approximation in GENEHUNTER—in the determination of statistical significance for genetic linkage.

For any direct simulation method, a crossover process model must be specified to describe the distribution of the recombination events along the chromosomes during meiosis. Because crossover interference has been shown to exist in humans (e.g., see Broman and Weber 2000), direct simulations should be based on a model that can incorporate crossover interference—for example, the χ^2 model (Zhao et al. 1995)—instead of the more commonly used Poisson model, which assumes the absence of crossover interference. However, the appropriateness of such crossover process models needs to be tested using extensive empirical data. Moreover, the effect of model misspecifications cannot be determined. On the other hand, the randomization procedure proposed in our article depends only on the observed recombination patterns rather than on a particular crossover model. Consider a family with one child, his or her two parents, and all four grandparents. For each marker, the inheritance vector for the child has two components (*f,m*), where $f = 0$ or 1 if the grandpaternal or grandmaternal allele was transmitted to the child from his/her father and $m = 0$ or 1 if the grandpaternal or grandmaternal allele was transmitted to the child from his/her mother. Assume the most ideal case, in which we can identify the grandparental origin for the two chromosomes in the child for all genetic markers being studied—that is, we can uniquely determine the inheritance vector of the child for all markers; therefore, (*f,m*) is known without ambiguity. In this case, we can pull the *f* component in the inheritance vectors for all the markers into a vector to summarize the transmissions from the father to the child across all the markers and the *m* component for all the markers in a separate vector to represent the transmission from the mother to the child across all the markers. For example, consider 10 markers and the following two vectors representing transmissions from the father and the mother, respectively, to the child: (1,1,0,0,0,0,0,1,1,1) and (0,0,0,0,1,1,1,1,1,1). For this example, the child inherited the grandmaternal alleles from the father at markers 1, 2, 8, 9, and 10 and the grandpaternal allele from the father at markers 3–7. Similarly, the child inherited the grandmaternal alleles from the mother at markers 5–10 and the grandpaternal allele from the mother at markers 1–4. Under the randomization procedure proposed in our article, for the 10 markers for this child, it is equally likely that each randomization would generate the following four inheritance vector pairs: (a) (1,1,0,0,0,0,0,1,1,1) and (0,0,0,0,1,1,1,1,1,1); (b) (0,0,1,1,1,1,1,0,0,0) and (0,0,0,0,1,1,1,1,1,1); (c) (1,1,0,0,0,0,0,1,1,1) and (1,1,1,1,0,0,0,0,0,0); and (d) (0,0,1,1,1,1,1,0,0,0) and (1,1,1,1,0,0,0,0,0,0). Therefore, the number of recombination events and the distribution of the recombinations are preserved in each randomized sample, and no specific crossover process models are used in the simulations. In contrast, for direct simulation methods, the number and positions of recombination events will differ across simulations.

Consider a family with two parents and two affected children. Using the notation by Kong and Nicolae (2000), we distinguish four states, for this pedigree, among the two affected children: (0,0) corresponds to the two sibs sharing zero alleles identical by descent (IBD); (1,1) corresponds to sharing both alleles IBD; (1,0) corresponds to IBD sharing of the paternal allele but not the maternal allele; and (0,1) corresponds to not

sharing the paternal allele but sharing the maternal allele. Assume that all four individuals in the pedigree have been genotyped at a single genetic marker and that the father has genotype (*A,A*), the mother has genotype (*B,C*), the first affected child has genotype (*A,B*), and the second affected child has genotype (*A,B*). Because the father is homozygous at this marker, we cannot uniquely determine the number of alleles IBD between the two affected children. With the notation defined by Kong and Nicolae (2000), for this pedigree $p(1,1) = p(0,1) = 1/2$ and the nonparametric linkage analysis score is $1/2 \times 1 + 1/2 \times 2 = 1.5$. The randomization procedure would generate the following four sets of probabilities with equal chance: (a) $\{p(0,0) = 0, p(0,1) = 1/2, p(1,0) = 0, p(1,1) = 1/2\}$; (b) $\{p(0,0) = 0, p(0,1) = 1/2, p(1,0) = 1/2, p(1,1) = 0\}$; (c) $\{p(0,0) = 1/2, p(0,1) = 0, p(1,0) = 1/2, p(1,1) = 0\}$; and (d) $\{p(0,0) = 1/2, p(0,1) = 0, p(1,0) = 1/2, p(1,1) = 0\}$. Therefore, in the randomized sample, the test statistic NPL = .5 and 1.5 with equal probability, whereas the "exact *P* value" in GENEHUNTER is calculated by means of a different reference distribution, in which the NPL = 0, 1, and 2 with probability 1/4, 1/2, and 1/4, respectively. Therefore, the procedure in GENEHUNTER-PLUS overestimates the variance for the NPL statistic for this particular family. In fact, this conservative approach of the statistical significance level evaluation in GENEHUNTER was the motivation of a likelihood-based approach in GENEHUNTER-PLUS by Kong and Cox (1997).

As a final note, the families analyzed in the insulin-dependent diabetes mellitus data set in our article have both parents available. Therefore, for this particular data set, the differences between the results from GENEHUNTER-PLUS and the randomization procedure are not likely to be due to the bias caused by incomplete parental information in the data set.

HONGYU ZHAO,[1,2] KATHLEEN R. MERIKANGAS,[1] AND KENNETH K. KIDD[2]
Departments of [1]Epidemiology and Public Health and [2]Genetics
Yale University School of Medicine
New Haven

References

Broman KW, Weber JL (2000) Characterization of human crossover interference. Am J Hum Genet 66:1911–1926

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kong A, Nicolae DL (2000) On a randomization procedure. Am J Hum Genet 67:000–000

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Zhao H, Merikangas KR, Kidd KK (1999) On a randomization procedure in linkage analysis. Am J Hum Genet 65: 1449–1456

Zhao H, Speed TP, McPeek MS (1995) Statistical analysis of crossover interference using the chi-square model. Genetics 139:1045–1056

Address for correspondence and reprints: Dr. Hongyu Zhao, Department of Epidemiology and Public Health, 60 College Street, Yale University School of Medicine, New Haven, CT 06520-8034. E-mail: hongyu.zhao@yale.edu

0002-9297/2000/6705-0036$02.00

Am. J. Hum. Genet. 67:1356–1359, 2000

The Promise and Pitfalls of Telomere Region–Specific Probes

To the Editor:

A complete set of telomere region–specific FISH probes designed to hybridize to the unique subtelomeric regions of every human chromosome was initially described in 1996 (National Institutes of Health et al. 1996), and an update was recently reported in the *Journal* (Knight et al. 2000). It was anticipated that these probes would be extremely valuable in the identification of submicroscopic telomeric aberrations that were thought to account for a substantial yet previously underrecognized proportion of cases of mental retardation in the population. Recently, a version of these probes was made commercially available as part of a diagnostic device that allows for simultaneous analysis of the telomeric regions of every human chromosome, except the p arms of the acrocentric chromosomes, on a single microscope slide (Cytocell) (Knight et al. 1997). The utility of these probes is evident in that numerous reports now exist describing cryptic telomere rearrangements or submicroscopic telomeric deletions that were undetectable by standard cytogenetic banding techniques but that were revealed by these FISH probes (Horsley et al. 1998; Ballif et al. 2000; reviewed in Knight and Flint 2000). Furthermore, several recent studies that have used these

Table 1

Clinically Significant Telomeric Aberrations Detected Using Telomere Region–Specific FISH Probes

Telomeric Aberration	No. Observed	Probe(s) Used
ish del(1)(qter)	1	PAC 160H23
ish der(2)t(2q;17q)pat	1[a]	P1 210E14, cosmid B37c1
ish der(18)t(7p;18q)mat	1	PAC 164D18, PAC 964M9
ish der(22)t(14q;22p)mat	1	PAC 820M16, D14Z1/D22Z1[b]

[a] Source: Bacino et al. (2000).

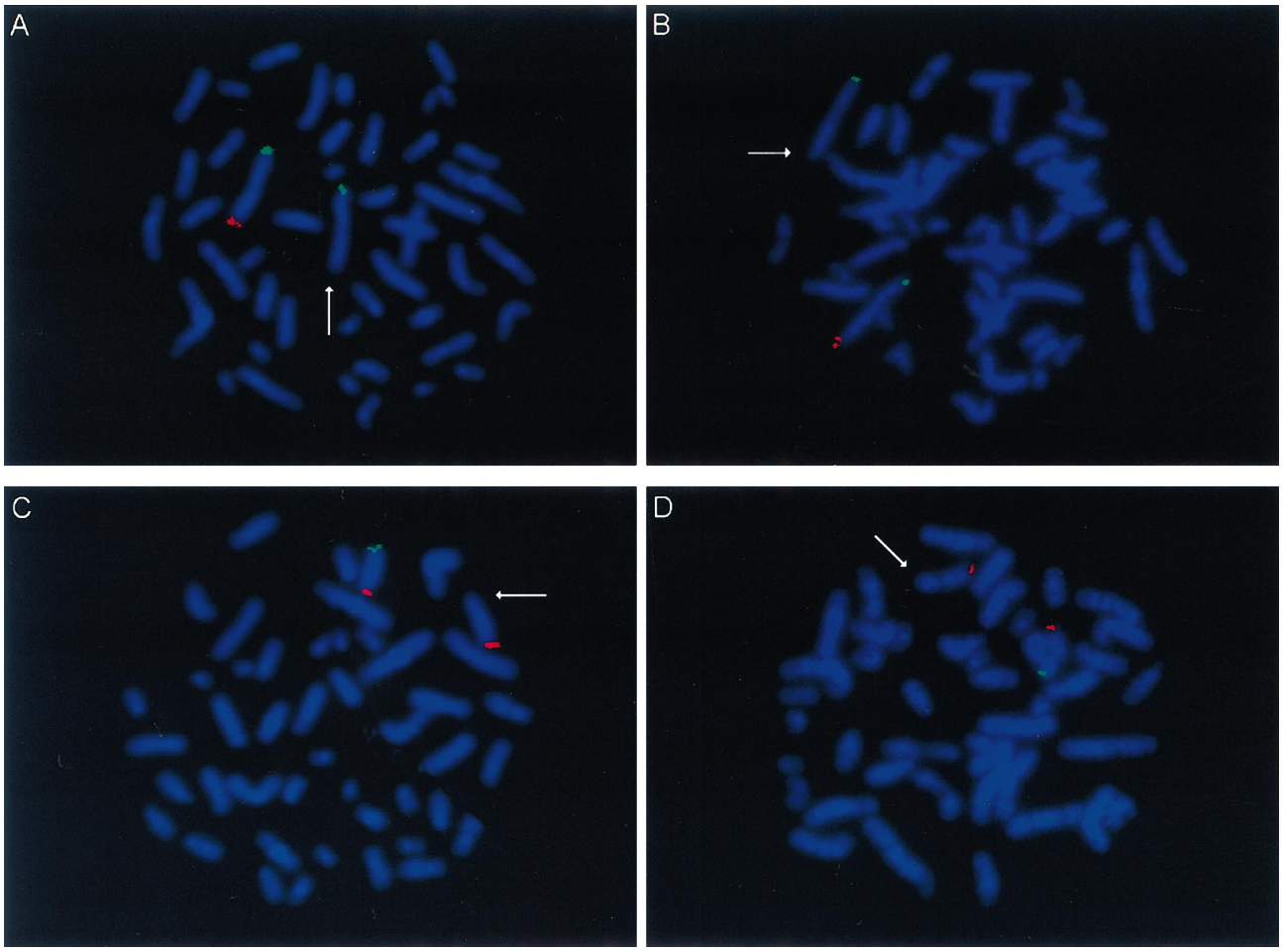

Figure 1 Subtelomeric polymorphisms detected by telomere region–specific FISH probes. Probes specific for p arms fluoresce green, and probes specific for q arms fluoresce red. *A,* Metaphase from a phenotypically normal parent, showing a deletion of the 2q telomere region–specific FISH probe (*arrow*). Note a normal hybridization pattern for the 2p telomere region–specific FISH probe. *B,* Metaphase from the child of the parent shown in *A,* indicating an inherited deletion of the 2q FISH probe (*arrow*). *C,* Metaphase from a patient with a polymorphic deletion of the Xp subtelomeric probe (*arrow*) that was paternally inherited (parental data not shown). *D,* Metaphase from a patient showing a polymorphic deletion, of the 9p FISH probe (*arrow*), that was paternally inherited (parental data not shown).

probes to investigate the telomeric regions of patients who have idiopathic mental retardation with apparently normal karyotypes indicate that ⩽23% of such cases have cryptic telomeric aberrations (Knight et al. 1999; reviewed in Knight and Flint 2000). This suggests that telomeric anomalies may be second only to Down syndrome as the most common cause of mental retardation (Knight and Flint 2000).

In our clinical cytogenetics laboratory, we have used telomere region–specific probes to examine the telomeric regions of 154 unrelated patients with apparently normal karyotypes, submitted for a variety of clinical indications. The recent report, in the *Journal,* by Knight et al. (2000) prompted us to examine the results, to date, of our telomeric FISH assay. This is not a controlled study of a selected population but, rather, a glance at the telomeric anomalies identified since the inception of the telomeric assay in the laboratory. Metaphase chromosomes obtained from peripheral blood specimens sent by the referring physician were analyzed in all cases. Of these patients, 15/154 (9.7%) had either submicroscopic telomeric deletions or cryptic telomeric rearrangements identified. However, only 4/15 (27%) telomeric abnormalities were shown to potentially contribute to the phenotype, since 11/15 (73%) patients inherited apparently benign telomeric variants from a phenotypically normal parent who carried the same cytogenetic "anomaly" (fig. 1). This reduces the percentage of clinically significant subtelomeric aberrations to 4/154 (2.6%) in our study population.

The four clinically significant telomeric abnormalities are listed in table 1. Patients included in this study un-

Table 2

Telomeric Polymorphisms Detected Using Telomere Region–Specific FISH Probes

Telomeric Polymorphism	No. Observed	Probe(s) Used
ish add(1)(qter)(13qtel+)pat	1[a]	PAC 163C9
ish del(2)(qter)mat or pat	8	PAC 1011O17, P1 210E14[b]
ish del(9)(pter)pat	1	PAC 43N6
ish del(X)(pter)pat	1	Cosmid CY29

[a] Source: Shaffer et al. (1999).

[b] PAC 1011O17 was deleted in all patients, and P1 210E14 was not deleted in all patients.

derwent diagnostic study because of a variety of clinical indications, including developmental delay, mental retardation, dysmorphic features, and/or multiple congenital anomalies. However, the precise details of the clinical diagnoses were not available to the diagnostic laboratory, which limited further extrapolation of these telomeric abnormalities being associated with a particular phenotype or subset of patients. All 11 observed telomeric polymorphisms are listed in table 2. Our data indicate that telomeric polymorphisms may be quite common (occurring in ~7% of patients studied), with a deletion in the 2q subtelomeric region occurring in 8/154 patients (~5% of the population). By means of a cosmid (2112b2), this 2q polymorphism has been detected elsewhere (Knight and Flint 2000; Knight et al. 2000). However, it was noted that the 2q probe present on the commercial telomere device had been recently updated, by the manufacturer, to a PAC probe (Genome Systems PAC 1011O17). Although the updated probe is larger than the first-generation cosmid used and is located <240 kb from the true telomere, it still detects the polymorphism (fig. 1*A* and *B*) (Knight and Flint 2000; Knight et al. 2000). For those patients who show a 2q deletion with PAC 1011O17, FISH using another version of the 2q probe (P1 210E14) (National Institutes of Health et al. 1996; Knight et al. 1997) demonstrated signals on both chromosomes, indicating nondeletion of this locus (data not shown). Although the parents of three patients with 2q deletions were unavailable for study, these patients showed the presence of the previously reported 2q subtelomeric probe (P1 210E14) on both homologues, making it highly likely that the anomalies seen in these patients also represent 2q polymorphisms. In addition, the XpYp subtelomeric cosmid probe (CY29), designed to hybridize to the pseudoautosomal regions of both sex chromosomes, has been shown to detect polymorphic sequences (Knight and Flint 2000; Knight et al. 2000), as found in one of our cases (fig. 1*C*). Detection of a 9pter polymorphism by telomere region–specific probes has not been previously reported (fig. 1*D*). It is expected that, as the limits of the technology are pushed farther toward the ends of the chromosome, more polymorphisms are likely to be identified.

The American College of Medical Genetics, in conjunction with the College of American Pathologists, has set forth guidelines for validation of FISH probes (Watson 1999). These guidelines suggest hybridizing five normal specimens with each new FISH probe. This approach will not uncover the frequency of these subtelomeric polymorphisms, and large numbers of normal individuals need to be tested to gather the frequencies of these polymorphic variants in the population. Although identifying these polymorphisms and the frequency with which they occur may help in the understanding of telomere structure and function, as well as in the understanding of the mechanisms that underlie the formation of terminal deletions and subtelomeric rearrangements, polymorphic subtelomeric probes are tenuous for diagnostic purposes. Whenever possible, when abnormalities are observed, parental samples should be tested with the same telomere region–specific probes, prior to the interpretation of the results from the child, to exclude the possibility of a benign familial polymorphism segregating in the family (Shaffer et al. 1999). This approach will improve the usefulness of these probes in the identification of telomeric alterations with true clinical significance.

Acknowledgments

We thank the following clinicians and counselors for submission of cases for clinical study: P. Benke (University of Miami School of Medicine, Miami); L. Celle, K. Russell, and E. Zackai (Children's Hospital of Philadelphia, Philadelphia); J. Graham (Cedar-Sinai Medical Center, Los Angeles); L. Sadler (Children's Hospital of Buffalo, Buffalo); and C. Bacino, A. Beaudet, B. Bejjani, W. Craigen, S. Fernbach, F. Scaglia, R. Sutton, and J. Towbin (Baylor College of Medicine, Houston).

BLAKE C. BALLIF, CATHERINE D. KASHORK, AND LISA G. SHAFFER
Department of Molecular and Human Genetics
Baylor College of Medicine
Houston

References

Bacino CA, Kashork CD, Davino NA, Shaffer LG (2000) Detection of a cryptic translocation in a family with mental retardation using FISH and telomere region-specific probes. Am J Med Genet 92:250–255

Ballif BC, Kashork CD, Shaffer LG (2000) FISHing for mechanisms of cytogenetically defined terminal deletions using chromosome-specific subtelomeric probes. Eur J Hum Genet 8:764–770

Horsley SW, Knight SJL, Nixon J, Huson S, Fitchett M, Boone RA, Hilton-Jones D, Flint J, Kearney L (1998) Del(18p) shown to be a cryptic translocation using a multiprobe FISH assay for subtelomeric chromosome rearrangements. J Med Genet 35:722–726

Knight SJL, Flint J (2000) Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. J Med Genet 37:401–409

Knight SJL, Horsley SW, Regan R, Lawrie NM, Maher EJ, Cardy DLN, Flint J, Kearney L (1997) Development and clinical application of an innovative fluorescence in situ hybridization technique which detects submicroscopic rearrangements involving telomeres. Eur J Hum Genet 5:1–8

Knight SJL, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Lawrie NM, Cardy DLN, Nguyen H, Hudson TJ, Riethman HC, Ledbetter DH, Flint J (2000) An optimized set of human telomere clones for studying telomere integrity and architecture. Am J Hum Genet 67: 320–332

Knight SJL, Regan R, Nicod A, Horsley SW, Kearney L, Homfray T, Winter RM, Bolton P, Flint J (1999) Subtle chromosomal rearrangements in children with unexplained mental retardation. Lancet 354:1676–1681

National Institutes of Health, Institute of Molecular Medicine Collaboration, Ning Y, Roschke A, Smith AC, Macha M, Precht K, Riethman H, Ledbetter DH, Flint J, Horsley S, Regan R, Kearney L, Knight S, Kvaloy K, Brown WRA (1996) A complete set of human telomeric probes and their clinical application. Nat Genet 14:86–89

Shaffer LG, Kashork CD, Bacino CA, Benke PJ (1999) Caution: telomere crossing. Am J Med Genet 87:278–280

Watson MS (ed) (1999) Standards and guidelines for clinical genetics laboratories, 2d ed. American College of Medical Genetics, Bethesda, MD

Address for correspondence and reprints: Dr. Lisa G. Shaffer, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 15E, Houston, TX 77030. E-mail: lshaffer@bcm.tmc.edu

0002-9297/2000/6705-0037$02.00