Letters to the Editor

Am. J. Hum. Genet. 63:1245, 1998

Allele Identical by Descent Sharing at Any Point of a Chromosome of a Sib Pair

To the Editor:

The distribution of identical by descent (IBD) alleles on a chromosome is a key component of multipoint linkage analysis (Goldgar 1990; Kruglyak and Lander 1995; Whittemore 1996). Goldgar (1990) and Guo (1994) considered a proportion of genetic material shared IBD by sibling pairs. Kruglyak and Lander (1995) used "inheritance vectors" (Lander and Green 1987) to calculate the probability that a sib pair shares 0, 1, or 2 alleles IBD. I propose a simple and straightforward procedure, based on the Haseman and Elston (1972) approach.

Suppose a chromosome has m markers, the distances between them being known. Assuming no crossover interference, the Haldane mapping function is used. Family data on marker phenotypes provide the probability $f_{i_k} \equiv P(i_k|M_k)$ that a sib pair has i_k alleles IBD at the kth marker loci, for $k = 1, 2, \ldots, m$ and $i_k = 0$, 1, or 2 (Haseman and Elston 1972, table 2). Denote by z the coordinate, on the chromosome, of the point studied that is between markers c ("closest") and $c + 1$.

The probability $P(i_z|M)$ that a sib pair shares i_z alleles IBD ($i_z = 0$, 1, or 2) at a point z, conditional on all the marker data M, is calculated by use of the formulas of total and conditional ("chain") probabilities:

$$\begin{aligned} P(i_z|M) &= \sum_{i_1,\ldots,i_m} P(i_z, i_1, \ldots, i_m|M) \\ &= \sum_{i_1,\ldots,i_m} P(i_1|i_2, \ldots, i_m, M)P(i_2|i_3, \ldots, i_m, M) \\ &\quad \times \cdots \times P(i_c|i_z, i_{c+1}, \ldots, i_m, M) \\ &\quad \times P(i_z, i_{c+1}, \ldots, i_m|M) \ . \end{aligned}$$

With the important assumption of no crossover interference, the allele sharing at any locus depends only on the marker data and the neighboring locus:

$$\begin{aligned} P(i_k|i_{k+1}, \ldots, i_m, M) &= P(i_k|i_{k+1}, M_k) \\ &= P(i_k, i_{k+1}, M_k)/P(i_{k+1}, M_k) \\ &= P(i_k|i_{k+1})P(i_k|M_k)/P(i_k) \ . \end{aligned}$$

The unconditional probabilities $P(i)$ are $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ for $i = 0$, 1, and 2, respectively. The conditional probability $\Psi'_{kl} \equiv P(k|l)$ that the sib pair has k alleles IBD at one locus if this pair has l alleles IBD at another locus is based on the corresponding joint probability $\Psi''_{kl} \equiv P(k, l)$, derived by Haseman and Elston (1972, table 4). Therefore,

$$P(i_k|i_{k+1}, \ldots, i_m, M) = \Psi'_{i_k i_{k+1}} f_{i_i}/P(i_k) \equiv \Psi_{i_k i_{k+1}} f_{i_k} \ ,$$

where $\Psi_{i_k i_{k+1}} = \Psi'_{i_k i_{k+1}}/P(i_k) = \Psi''_{i_k i_{k+1}}/[P(i_k)P(i_{k+1})]$ or, in the matrix notation,

$$\Psi = \begin{bmatrix} 4\psi^2 & 4\psi(1-\psi) & 4(1-\psi)^2 \\ 4\psi(1-\psi) & 4(\frac{1}{2} - \psi + \psi^2) & 4\psi(1-\psi) \\ 4(1-\psi)^2 & 4\psi(1-\psi) & 4\psi^2 \end{bmatrix},$$

where $\psi = r^2 + (1 - r)^2$, and r is the recombination fraction, calculated from the known distance between the marker loci studied, by use of the Haldane mapping function; the indices i_k and i_{k+1} are omitted for ψ, Ψ, and r.

The sum for the first marker is

$$\sum_{i_1} P(i_1|i_2, M_1) = \sum_{i_1} \Psi_{i_1 i_2} f_{i_1} \equiv f^{(1)}_{i_2} \ .$$

The right notation emphasizes that the probabilities indexed for the second marker "picked up" information from the first one. For the second marker,

$$\sum_{i_1, i_2} P(i_1|i_2, M_1)P(i_2|i_3, M_2) = \sum_{i_2} \Psi_{i_2 i_3} f_{i_2} f^{(1)}_{i_2} \equiv f^{(1,2)}_{i_3}$$

and so on, up to the closest marker to the left of the trait locus (included):

$$f^{(1,\ldots,c)}_{i_z} = \sum_{i_c} \Psi_{i_c i_z} f_{i_c} f^{(1,\ldots,c-1)}_{i_c} \ . \qquad (1)$$

Remember that $\Psi_{i_c i_z}$ depends on z and that the Haldane mapping function is used. In the part of the chromosome to the right of point z, we proceed in the opposite direction:

$$P(i_z, i_{c+1}, \ldots, i_m | M) = P(i_m | i_{m-1}, M)P(i_{m-1} | i_{m-2}, M) \times \cdots \times P(i_{c+1} | i_z, M)P(i_z) \ .$$

Again, by virtue of the assumption of no crossover interference,

$$P(i_k | i_{k-1}, \ldots, i_1, M) = P(i_k | i_{k-1}, M_k) = \Psi'_{i_{i-1} i_i} f_{i_i} / P(i_{k-1}) \equiv \Psi_{i_{i-1} i_i} f_{i_i} \ .$$

The sum for the last marker is

$$\sum_{i_m} P(i_m | i_{m-1}, M_m) = \sum_{i_m} \Psi_{i_m i_{m-1}} f_{i_m} \equiv f^{(m)}_{i_{m-1}} \ ;$$

then,

$$\sum_{i_{m-1}, i_m} P(i_{m-1} | i_{m-2}, M_{m-1}) = \sum_{i_{m-1}} \Psi_{i_{m-1} i_{m-2}} f_{i_{m-1}} f^{(m)}_{i_{m-1}} \equiv f^{(m-1,m)}_{i_{m-2}}$$

and so on, up to the closest marker to the right of the trait locus (included):

$$f^{(c+1,\ldots,m)}_{i_z} = \sum_{i_{c+1}} \Psi_{i_{c+1} i_z} f_{i_{c+1}} f^{(c+2,\ldots,m)}_{i_{c=1}} \ . \qquad (2)$$

Finally, the probability at point z is the joint probability from the left (formula [1]) and right (formula [2]) parts of the chromosome:

$$P(i_z | M) = f^{(1,\ldots,c)}_{i_z} P(i_z) f^{(c+1,\ldots,m)}_{i_z} \equiv f^{(1,\ldots,m)}_{n_z} \ .$$

So, the prior probability $P(i_z)$ at point z is "corrected" by the marker data from both sides. When z is to the left of the first marker, $c = 0$ and the left factor disappears; when z is to the right of the last marker, $c = m$ and the right factor disappears; when z is at the position of the kth marker, $P(i_z | M_k)$ replaces $P(i_z)$, meaning that a noninformative marker receives information from its neighbors. If a marker is fully informative, only one of f_0, f_1, or f_2 is equal to 1; others are equal to 0, thus cutting the "probability chain."

The number of calculations is proportional to the number of marker loci in this multipoint method. If intermediate results are stored, this method leads to a fast algorithm for the calculation of allele IBD sharing at any point of a chromosome, for every sib pair. With this distribution, linkage tests for quantitative and qualitative traits may be derived, by use of likelihood, regression, scores, or other methods, which will be the subject of a separate communication.

Acknowledgments

This work was supported by research grant F05 TW05285 from the Fogarty International Center and resource grant P41 RR03655 from the National Center for Research Resources, National Institutes of Health.

EUGENE DRIGALENKO
Department of Epidemiology and Biostatistics
Rammelkamp Center for Education and Research
MetroHealth Campus
Case Western Reserve University
Cleveland

References

Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47:957–967

Guo S-W (1994) Computation of identity-by-descent proportions shared by two siblings. Am J Hum Genet 54: 1104–1109

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Lander ES, Green P (1987) Construction of multipoint genetic linkage maps in humans. Proc Natl Acad Sci USA 84: 2363–2367

Whittemore AS (1996) Genome scanning for linkage: an overview. Am J Hum Genet 59:704–716

Address for correspondence and reprints: Dr. Eugene Drigalenko, Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, 2500 MetroHealth Drive, Room R258, Cleveland, OH 44109-1998. E-mail: dei@darwin.mhmc.cwru.edu

0002-9297/98/6304-0047$02.00