Am. J. Hum. Genet. 41:350–355, 1987

DSLINK: A Computer Program for Gene-Centromere Linkage Analysis in Families with a Trisomic Offspring

SUSAN L. HALLORAN AND ARAVINDA CHAKRAVARTI

Human Genetics Program, Department of Biostatistics, University of Pittsburgh, Pittsburgh

SUMMARY

Trisomic individuals provide information for gene-centromere mapping, since two of the four chromatids in a meiotic tetrad can be recovered. When centromeric markers are available, linkage analysis between the centromere and any marker locus can be performed in nuclear families having one or more trisomic offspring. Since conventional linkage programs consider only disomic individuals, we have written a FORTRAN computer program, DSLINK, that performs gene-centromere linkage analysis on the basis of information on trisomic and disomic offspring. This program makes it possible to study the relationship between recombination and chromosome segregation.

INTRODUCTION

Gene-centromere distances can be calculated from the centromeric and marker genotypes in trisomic individuals, since two of the four members of a meiotic tetrad can be recovered from the disomic gamete transmitted by the parent in whom nondisjunction has occurred. Although the principle is analogous to tetrad mapping in fungi (Strickberger 1985), the present method is referred to as unordered half-tetrad mapping, since only two of the four chromatids in a meiotic tetrad can be studied. The half-tetrad arising by nondisjunction can result from either a meiosis I or meiosis II nondisjunctional event.

By studying polymorphic centromeric markers in both of the parents and in the trisomic child, one can determine which parent contributed the nondisjoined gamete and whether nondisjunction occurred during meiosis I or meiosis

Received December 4, 1986; revision received March 18, 1987.

Address for correspondence and reprints: Dr. Aravinda Chakravarti, Room A310 Crabtree Hall, GSPH, Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261.

 0002-9297/87/4103-0002$2.00

TABLE 1

PROBABILITY OF A HETEROZYGOUS DISOMIC GAMETE RESULTING FROM MEIOTIC NONDISJUNCTION

Nondisjunction at Meiosis	Probability of Heterozygous Disomic Gamete
I	$1 - \frac{y}{2}$
II	y

II. Consider a set of marker loci on the particular chromosome of interest, markers that are heterozygous in the parent in whom nondisjunction has occurred. Clearly, in the absence of recombination, the genotypes of all marker loci in the disomic gamete transmitted to the trisomic offspring will be heterozygous under a meiosis I error and homozygous under a meiosis II error. Any departure from these zygosities is, then, an indication of recombination between the particular marker locus and the centromere, the magnitude of the departure being an indication of the amount of recombination. This information can be used to estimate gene-centromere distances, as outlined by Chakravarti and Slaugenhaupt (in press), in terms of a new linkage parameter, y. Specifically, for a particular marker locus the probability of a heterozygous disomic gamete for each type of meiotic error can be expressed in terms of y, as shown in table 1.

The linkage parameter y can be related to the recombination value θ between the centromere and the marker locus by using a function appropriate for the chromosome under study. If w is the map distance between the centromere and the marker locus, then $y = 0$ when $\theta = 0$ ($w = 0$). Additionally, when $\theta = \frac{1}{2}$ and there are numerous crossovers, $y = \frac{2}{3}$ (Mather 1938; Morton 1982). Also, when θ is small, $w \approx \theta$ and $y \approx 2\theta$ (Ott et al. 1976). The relationship between y and θ will also vary with the degree of interference assumed, as shown in table 2. When there are at most two chiasmata, a map function relating w and θ is required; for humans, one may use the generalized map function of Rao et al. (1977) with interference parameter $p = 0.398$, as suggested by Morton et al. (1985). The formula $y = \frac{2}{3}\,\mathrm{Sin}(3w)$ is based on empirical data on nonreduction from yeast and *Drosophila* (Ott et al. 1976). For the smaller human chromo-

TABLE 2

FUNCTIONS RELATING y AND w FOR VARIOUS NUMBERS OF CHIASMATA

No. of Chiasmata/Interference	Function	Reference
No interference	$y = \frac{2}{3}(1 - e^{-3w})$	Mather 1938; Ott et al. 1976
At most one chiasma	$y = 2w$	Ott et al. 1976
At most two chiasma	$y = 3\theta - w$	Morton and MacLean 1984
Empirical	$y = \frac{2}{3}\,\mathrm{Sin}(3w)$	Ott et al. 1976

TABLE 3

LIKELIHOOD OF THE TRISOMIC OFFSPRING'S GENOTYPE CONDITIONAL ON PARENTAL GENOTYPES

CHROMOSOME FROM NORMAL SEGREGATION	NONDISJOINED CHROMOSOMES: Meiotic Error	NONDISJOINED CHROMOSOMES: Genotype of Disomic Gamete	LIKELIHOOD OF THE GENOTYPES OF A TRISOMIC CHILD
Recombinant	I	Homozygous	$\theta\left(\frac{y}{2}\right)$
		Heterozygous	$\theta\left(1 - \frac{y}{2}\right)$
	II	Homozygous	$\theta(1 - y)$
		Heterozygous	θy
Nonrecombinant	I	Homozygous	$(1 - \theta)\left(\frac{y}{2}\right)$
		Heterozygous	$(1 - \theta)\left(1 - \frac{y}{2}\right)$
	II	Homozygous	$(1 - \theta)(1 - y)$
		Heterozygous	$(1 - \theta)(y)$

somes, such as chromosome 21, the assumption of complete interference (i.e., the existence of at most one chiasma) is probably valid.

METHODS

The likelihood of the marker and centromere genotypes of a trisomic offspring conditional on parental genotypes can be written in terms of y and θ, y being the gene-centromere "distance" for the chromosomes involved in nondisjunction and θ being the gene-centromere recombination value for chromosomes undergoing normal segregation. Specifically, when the origin of nondisjunction and parental linkage phases are known, the likelihood of the genotype of a trisomic child is the product of one of the probabilities from table 1 (for the disomic gamete) and of the probability for the gamete transmitted by the parent in whom normal chromosome segregation has occurred. This second probability is θ if recombination has occurred; in the absence of recombination, it is $1 - \theta$. These likelihoods are given in table 3. However, when the origin of nondisjunction is ambiguous, the likelihood of the jth trisomic offspring is the weighted average

$$L_{jl}(\theta, y) = \sum_{i=1}^{4} u_i L_{jil}(\theta, y) \quad , \tag{1}$$

where u_1, u_2, u_3, and u_4 ($\Sigma_{i=1}^{4} u_i = 1$) are the probabilities of maternal I, maternal II, paternal I, and paternal II nondisjunction, respectively, and where L_{jil} is the likelihood of the jth trisomic offspring conditional on the ith meiotic error and the lth linkage phase in parents. Clearly, the likelihood of any unaffected offspring in the same family is a function of θ only. Therefore, the likelihood of a family with both unaffected and trisomic offspring is a function of θ and y. For nuclear-family data, parental linkage phase will be unknown, so the likelihood must be averaged over all four possible combinations of parental linkage phase. The likelihood, then, of a family with n unaffected children and t trisomic children is

$$L = \tfrac{1}{4} \sum_{l=1}^{4} \prod_{k=1}^{n} L_{kl}(\theta) \prod_{j=1}^{t} L_{jl}(\theta, y) \ , \tag{2}$$

where $L_{kl}(\theta)$ is the likelihood of the kth unaffected child conditional on the parental linkage phase and $L_{jl}(\theta, y)$ is the likelihood in equation (1). The likelihood in equation (2) may then be converted into a lod score as follows:

$$Z(\theta, y) = \log_{10} \left\{ \frac{L(\theta, y)}{L(\theta = \tfrac{1}{2}, y = \tfrac{2}{3})} \right\} \ . \tag{3}$$

Therefore, for particular marker loci, linkage information may be conveniently summed over all families. By means of standard interpolation methods (Hodge et al. 1983), maximum-likelihood estimates of $\theta(\hat{\theta})$ and $y(\hat{y})$ and the maximum lod score $\hat{Z}(\theta,y)$ may be obtained from the bivariate lod table so constructed. Since all markers to be tested will reside on the chromosome of interest, the null hypothesis of no linkage ($H_0 : \theta = \tfrac{1}{2}; y = \tfrac{2}{3}$) can be tested using the statistic $2\ln(10)\hat{Z}(\theta,y)$, which is distributed as a χ^2 variable with 2 df. Of more interest, however, is the test $H_0 : \theta_t = \theta_c$ versus $H_1 : \theta_t < \theta_c$, where θ_t is the recombination value for the chromosomes undergoing nondisjunction (estimated from y by means of an appropriate mapping function) and θ_c is the recombination value either as estimated for the normal segregants or from conventional linkage studies in control families. This hypothesis may be tested by means of the following statistic:

$$G = 2\ln(10) \{\hat{Z}(\theta_c, \theta_t) - \hat{Z}(\theta_c, \theta_t = \theta_c)\} \ , \tag{4}$$

where $\hat{Z}$ is the maximum lod score under the appropriate hypothesis and G is distributed as a χ^2 variable with 1 df. Since this is a one-sided test, the significance may be tested with the P value $1 - \Phi(\sqrt{G})$, where Φ is the cumulative normal density function and G is obtained from equation (4). Alternatively, one may wish to assume that the recombination values are equal ($\theta_t = \theta_c$) and estimate the mapping function for chromosomes undergoing nondisjunction. Specifically, for a given y value, one chooses that mapping function that estimates a θ_t value equal to θ_c.

DSLINK

DSLINK is a FORTRAN computer program designed for use on a VAX 8600. The likelihood of the offspring conditional on the parental genotypes at the centromere and a particular marker locus, given θ and y, is calculated in the following manner: For each of the possible parental linkage phases, the likelihood of each trisomic child for each meiotic error is calculated and weighted by the probability of the meiotic error; these likelihoods are summed and then multiplied over all trisomic children. The likelihood of each unaffected child is calculated and multiplied over all unaffected offspring. The likelihood for the family is, then, the product of the likelihoods of the two classes of offspring averaged over the four possible parental linkage phases. The θ and y values for which the likelihood is calculated are provided by the user. By requiring that the user supply these values, the generality of the program is maintained, so that data from any chromosome involved in nondisjunction can be examined. The y values should be obtained from θ values by means of the function appropriate for the chromosome under study (table 2).

As has been shown by Hassold and Jacobs (1984), the probabilities of the different meiotic errors are not equal, with maternal meiosis I errors being the most common and paternal meiosis II errors being the least frequent. Additionally, for some families the parental origin and meiotic stage of nondisjunction will be known from the centromeric and marker loci. To allow for the incorporation of this information, DSLINK requires that the user provide the probabilities for each meiotic error for each family. For a family in which the origin of nondisjunction is ambiguous, these probabilities would be the population-based probabilities (Hassold and Jacobs 1984). Thus, for a family in which the error is known—e.g., maternal meiosis I—$u_1 = 1$ and $u_2 = u_3 = u_4 = 0$.

DSLINK requires that there be a centromeric marker and that all loci must be codominant. Gene frequencies are not needed, since the likelihood of the offspring conditional on parental genotypes is calculated; parental genotypes are assumed to be known. The families must consist of two parents and one or more offspring; there may be multiple trisomic offspring. The output consists of both a table of likelihoods for each family for each noncentromeric marker and bivariate lod tables for each noncentromeric marker summed over all families.

DISCUSSION

Linkage data on centromeric and noncentromeric markers from families with a trisomic child are useful in investigating the etiology of nondisjunction. If all nondisjunctional events result from asynapsis, then $\theta_r < \theta_c$, a hypothesis that can be tested easily by using the lod scores, as described above. If, however, the different meiotic errors have different etiologies, the matter becomes more complex; then families must be grouped according to origin of nondisjunction and θ must be estimated for each type of error. In this way, meiosis I nondisjunction may be compared to meiosis II nondisjunction and paternal nondisjunction to maternal nondisjunction. DSLINK has recently been used in the analysis of trisomy 21 (Warren et al., in press). Individuals interested in obtaining a

copy of DSLINK and the documentation should send a 5-inch × 5-inch diskette to the authors.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grant GM33771.

REFERENCES

Chakravarti, A., and S. A. Slaugenhaupt. Methods for studying recombination on chromosomes which undergo nondisjunction. Proc. Natl. Acad. Sci. USA (in press).

Hassold, T. J., and P. A. Jacobs. 1984. Trisomy in man. Annu. Rev. Genet. **18**:69–97.

Hodge, S. E., C. E. Anderson, K. Neiswanger, R. S. Sparkes, and D. L. Rimoin. 1983. The search for heterogeneity in insulin-dependent diabetes mellitus (IDDM): linkage studies, two-locus models, and genetic heterogeneity. Am. J. Hum. Genet. **35**:1139–1155.

Mather, K. 1938. Crossing over. Biol. Rev. **13**:252–292.

Morton, N. E. 1982. Outline of genetic epidemiology. Karger, New York.

Morton, N. E., and C. J. MacLean. 1984. Multilocus recombination frequencies. Genet. Res. **44**:99–108.

Morton, N. E., C. J. MacLean, and R. Lew. 1985. Tests of hypotheses on recombination frequencies. Genet. Res. **45**:279–286.

Ott, J., D. Linder, B. K. McCaw, E. W. Lovrien, and F. Hecht. 1976. Parthenogenic origin of benign ovarian teratomas. Ann. Hum. Genet. **40**:191–196.

Rao, D. C., N. E. Morton, J. Lindsten, M. Hulten, and S. Yee. 1977. A mapping function for man. Hum. Hered. **27**:99–104.

Strickberger, M. W. 1985. Genetics. Macmillan, New York.

Warren, A. C., A. Chakravarti, C. Wong, S. A. Slaugenhaupt, S. L. Halloran, P. C. Watkins, C. Metaxotou, and S. E. Antonarakis. Evidence for reduced recombination on the nondisjoined chromosomes 21 in Down syndrome. Science (in press).