Am. J. Hum. Genet. 61:418–422, 1997

Analysis of Genetic Linkage and Somatic Loss of Heterozygosity in Affected Pairs of First-Degree Relatives

K. Rohde,[1] M. D. Teare,[2] and M. Santibáñez Koref[3]

[1]Department of Bioinformatics, Max Delbrück Centre for Molecular Medicine, Berlin; [2]CRC Genetic Epidemiology Unit, Institute of Public Health, University of Cambridge, Cambridge; and [3]MRC Gene and Genome Evolution Group, Hammersmith Hospital, London

Summary

Recently, data on loss of constitutional heterozygosity (LOH) have been used by several groups to increase the power to detect linkage in pedigrees with an inherited cancer predisposition. This approach assumes that the predisposition is due to the inheritance of the defective copy of a tumor suppressor. In order to assess the gain of power expected from the inclusion of LOH data, we simulated segregation and somatic loss of alleles in pedigrees consisting of an affected pair of first-degree relatives. We explored the effects of pedigree structure, frequency of loss, penetrance, and recombination rate on the expected LOD score. The results indicate that, for establishment of genetic linkage, isolated parent-offspring pairs can be as informative as sib pairs and that they could represent an additional source of information in linkage studies.

Recently several groups have used data on loss of constitutional heterozygosity (LOH) to increase the power to detect genetic linkage (Rebbeck et al. 1994; Steichen-Gersdorf et al. 1994; Teare et al. 1994; Lustbader et al. 1995; Rohde et al. 1995). This approach is applicable when the cancer predisposition is caused by the inheritance of the defective copy of a tumor-suppressor gene. In this situation, inactivation of the wild-type allele in the tumor by deletion will be accompanied by loss of alleles at marker loci close to the tumor suppressor. The alleles retained in the tumor will tend to originate from the same parent as the defective copy of the tumor-suppressor gene.

In this communication we explore the consequences that incorporation of LOH has on both the expected LOD scores of isolated pairs of affected first-degree relatives and the influence of various parameters. Sib pairs are widely used to locate genes predisposing to disease. Isolated parent-offspring pairs could represent a valuable additional source of information but are uninformative without additional data. They can become informative when LOH is included. Observations from the tumor of only one of the affected pedigree members can already provide indications regarding the location of the marker with respect to the disease locus. Two such cases are depicted in figure 1. In figure 1*a*, marker allele 2 is retained in the parental tumor and is transmitted to the affected offspring. This is consistent both with marker allele 2 being located on the same chromosome as the defective tumor-suppressor allele and with the marker and tumor-suppressor loci being close enough for both to be deleted by the same event and for no recombination to have happened between them. The pedigree and typing depicted in figure 1*b*, where the allele retained in the tumor is not the one transmitted to the affected child, would not support these assumptions.

In a previous paper we presented a framework for the formal incorporation of LOH into linkage analysis (Teare et al. 1994). This approach can be outlined as follows: For an individual with the genotype g, we incorporate LOH as a component of the phenotype x: $x = (d,l)$, where d describes the disease status of the individual and l describes the genetic changes observed at the marker locus in the tumor of this individual. We can therefore partition the penetrance function $P(x|g)$: $P(x|g) = P(d|g)P(l|g,d)$. To describe $P(l|g,d)$—that is, the conditional probability of showing the LOH status l, given the genotype g and the disease status d—we introduced the parameters λ_{ij}, ε_{ij}, and ρ. λ_{ij} and ε_{ij} are the probabilities of observing a loss of one or both alleles at the marker locus, given the alleles i and j at the disease locus ($i,j \in \{d,D\}$, where d designates wild-type allele and D denotes the allele associated with cancer predisposition). The parameter ρ applies to the case when an individual heterozygous at the disease locus shows in the tumor the loss of one allele at the marker locus. ρ describes the probability that the marker allele retained in the tumor and the disease-predisposing allele do not originate from the same parent. It is equivalent to the parameter λ in the report by Rebbeck et al. (1994) and to ϕ in the report by Steichen-Gersdorf et al. (1994).

For example, if an individual is heterozygous at the marker locus and has inherited a wild-type (d) and a

Received January 27, 1997; accepted for publication May 23, 1997.

Address for correspondence and reprints: Dr. K. Rohde, Department of Bioinformatics, Max Delbrück Centre for Molecular Medicine, Robert Rössle Straße 10, Berlin Buch D-13122, Germany. E-mail: rohde@mdc-berlin.de

0002-9297/97/6102-0021$02.00

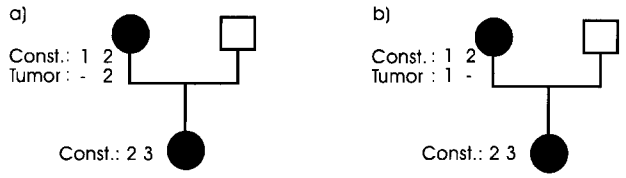

Figure 1 Examples of affected parent-offspring pairs that become informative when LOH data are considered. We assume that the mother's tumor was typed. The numbers represent alleles at the marker locus. Const. = constitutional genotype; and Tumor = genotype of the tumor.

defective (D) allele at the disease locus, the probability of observing a loss at the marker locus, given that the retained marker allele and the cancer-predisposing allele D were inherited on the same chromosome, is $P(l|g,d) = \lambda_{dD}(1 - \rho)$. A more detailed discussion of this model, as well as examples of its application, can be found in the work of Teare et al. (1994) and Rohde et al. (1995).

In order to ascertain the gain of power expected from the inclusion of LOH data, we adapted the C version of the SLINK program (Ott 1989; Weeks et al. 1990; Cottingham et al. 1993) to simulate the segregation and somatic loss of alleles. We implemented the modification of the penetrance function in SLINK, as described for MLINK elsewhere (Rohde et al. 1995). The modified program can simulate LOH at autosomal and X-linked loci in a variety of pedigrees. Its source code and some examples are available via the anonymous ftp server at ftp.mdc-berlin.de (login: anonymous; password: E-mail address; file: lslink.tar.Z in the directory pub/linkage).

We used familial breast cancer linked to BRCA1 as a model in our simulations, since LOH at loci surrounding BRCA1 has been extensively studied in families with disease predisposition linked and unlinked to this locus, as well as in sporadic cases (Cornelis et al. 1995, and references therein). Therefore, affected sib pairs were represented by sister-sister (SS) pairs and parent-offspring pairs by mother-daughter (MD) pairs. The data from Easton et al. (1993) were used both for the liability classes and for the frequency of the predisposing allele (f_D). We assumed throughout the calculations that normal and tumor material were typed for each affected individual, that affected individuals had only one tumor, that no phenotypic or genotypic information was available for unaffected individuals, and that homozygous losses are rare enough to be neglected ($\varepsilon_{dd} = \varepsilon_{dD} = \varepsilon_{DD} = .00$). For each set of parameters and pedigree, 1,000 replicates were generated and analyzed by use of the modified versions of MLINK or ILINK (Rohde et al. 1995).

It is intuitively clear that, the more frequently LOH is observed, the more power we expect to gain. The effects of the frequency of loss at the marker locus are explored in figure 2. As in standard linkage analysis, the more polymorphic the marker, the more information can be extracted. Figure 2 shows that the expected LOD scores for a MD pair and a SS pair become more similar when the frequency of loss moves toward one. In the case where no loss at the marker locus is observable, the MD pair becomes uninformative. For a SS pair the expected LOD score more than doubles for a frequency of loss of .4. It is perhaps surprising that the expected LOD score obtainable for a MD pair can be similar to that achieved for a SS pair, although, if a tumor is not observed, the former is always uninformative. However, we should bear in mind that, in the absence of linkage, there is a .5 probability that a particular marker allele present in one sister—that is, the allele retained in her tumor—is shared, by descent, with the other sister. This is also the probability that a particular allele will be shared between a mother and her daughter. For (meiotic) recombination fraction $\theta = .00$, these probabilities are similar, and hence so are the LOD scores.

The frequency of loss at the marker locus may depend not only on its location—in particular, with respect to the disease locus—but also on the genotype at the disease locus. Cornelis et al. (1995) reported LOH for markers surrounding BRCA1 in 50 of 58 tumors from members of families with disease predisposition linked to BRCA1. The retained alleles did always originate from the same parent as did the disease predisposition.

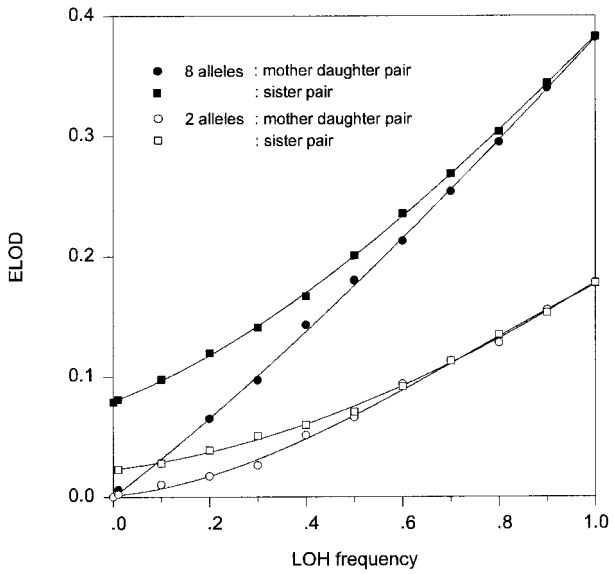

Figure 2 Expected LOD scores for different frequencies of loss and either eight (*blackened symbols*) or two (*unblackened symbols*) alleles at the marker locus, for affected pedigree members diagnosed with breast cancer before the age of 31 years (liabilities are .00002, .00167, and .00167 for the genotypes *dd, dD,* and *DD,* respectively); $f_D = .003$, $\theta = .00$, and $\rho = .00$. Simulation and analysis were done by use of the same parameters. MD pairs are represented by circles, and SS pairs are represented by squares.

Table 1

Effects That Frequencies of LOH Have on Expected LOD Scores, Depending on Genotype at Disease Locus, for an SS Pair and an MD Pair

Age at Onset (Penetrance for Age Interval, by Genotype [*dd, dD, DD*])	LOH Frequencies Used in: Simulation λ_{dd}	λ_{dD}	λ_{DD}	Analysis: All λ = .3, MD Pair	All λ = .3, SS Pair	λ as in Simulation, MD Pair	λ as in Simulation, SS Pair
< 31 years (.00002, .00167, .00167)	.8	.8	.8	.30	.31	.30	.31
	.4	.8	.2	.31	.31	.32	.32
	.2	.8	.2	.32	.32	.32	.33
51–60 years (.00137, .01711, .01711)	.8	.8	.8	.03	.03	.03	.03
	.4	.8	.2	.04	.04	.06	.06
	.2	.8	.2	.04	.04	.07	.08

Note.—The calculations were performed for eight alleles at the marker locus; f_D = .003, θ = .00, and ρ = .00. Unless specified otherwise, the same parameters were used for simulation and analysis.

For families that were unlinked or for which linkage to BRCA1 was uncertain, LOH was detected in 9 of 19 tumors. Similar observations have been reported for LOH at 13q in families with and without disease linked to BRCA2 (Collins et al. 1995).

The effects of three different sets of LOH frequencies on the expected LOD scores are shown in table 1. The values for λ_{dd} and λ_{dD} in the second set are close to the frequencies described by Cornelis et al. (1995). The choice of λ_{DD} = .2 is arbitrary. It corresponds to the average LOH frequency determined by Sato et al. (1990) for a series of markers across the genome. It has little influence on the calculations, given the low frequency assumed for allele *D*. The third set of values (λ_{dd} = .2, λ_{dD} = .8, and λ_{DD} = .2) should help in the investigation of the effects of a larger difference between λ_{dd} and λ_{dD}. Table 1 also shows that the expected LOD scores increase the smaller λ_{dd} is with respect to λ_{dD}. The relative increase is more marked the more similar the liability values for a carrier and a noncarrier are. Misspecification of the frequencies of loss, as presented in table 1, results in a drop in the expected LOD scores. In the first row for each liability class, either the first and the third entries or the second and the fourth entries are identical. In general, when the analysis is performed under the assumption that the frequency of loss is independent of the genotype at the disease locus (i.e., $\lambda_{dd} = \lambda_{dD} = \lambda_{DD} = \lambda > 0$), the obtained LOD score is independent of the specific value chosen for λ, provided that data on the normal material are also available for every tumor typed. This is equivalent to multiplying each value of a liability class—that is, the penetrance for each possible genotype—by a constant factor. Therefore, $\lambda_{dd} = \lambda_{dD} = \lambda_{DD} = \lambda > 0$, $\varepsilon_{dd} = \varepsilon_{dD} = \varepsilon_{DD} = \varepsilon \geqslant 0$, and $\varepsilon + \lambda < 1$ ($\varepsilon > 0$, if homozygous losses have been observed) seem a reasonable choice in the absence of estimates for the frequencies of loss, but, as the results from table 1 suggest, it may result in loss of power.

The increase in power expected from the inclusion of LOH depends on the certainty with which we can infer the phase between the cancer-predisposing allele and the marker allele retained in the tumor. The parameter ρ describes the probability that these two alleles do not originate from the same parent. This can happen either when loss at the marker locus is unrelated to inactivation at the disease locus or when there has been a recombination event between both loci after conception and prior to the occurrence of the deletion. Steichen-Gersdorf et al. (1994) assumed ρ = 0 for markers close to BRCA1. This assumption seems justified in light of the results reported by Cornelis et al. (1995). However, deletion of the allele mutated in the germ line has been observed in tumors from members of Li-Fraumeni families with a germ-line p53 mutation (Varley et al. 1997). It is possible that, in these cases, the retained allele has been inactivated during tumor development. This may be a phenomenon particular to p53, but it illustrates that, even in cases in which marker and disease locus are very close, to assume that the mutant allele is on the same homologue as the marker alleles retained in the tumor may result in a misspecification of ρ. Figure 3 shows that the expected LOD score decreases with increasing ρ, in particular for a MD pair as compared with a SS pair. Nevertheless, for ρ = .2—that is, retention of "the wrong allele" in one of five tumors—the expected LOD score for a MD pair is still as high as that expected for a SS pair without LOH.

As shown in figure 4, when ρ is misspecified the maximum expected LOD score is achieved for a θ different from the one used to generate the data—that is, misspec-

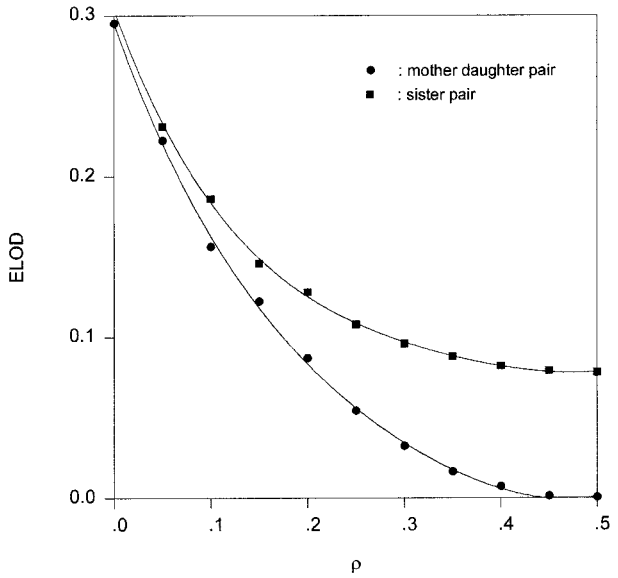

Figure 3 Expected LOD score for different values of ρ, for a SS pair (*squares*) and a MD pair (*circles*), for affected pedigree members diagnosed with breast cancer before the age of 31 years, with eight alleles at the marker locus, frequency of loss .8 irrespective of the genotype at the disease locus, and θ = .00. Simulation and analysis were done by use of the same ρ values.

ification of the "postzygotic" recombination fraction not only decreases the obtainable LOD score to the extent that any indication of linkage may vanish but also results in a wrong estimate of θ. This effect is more

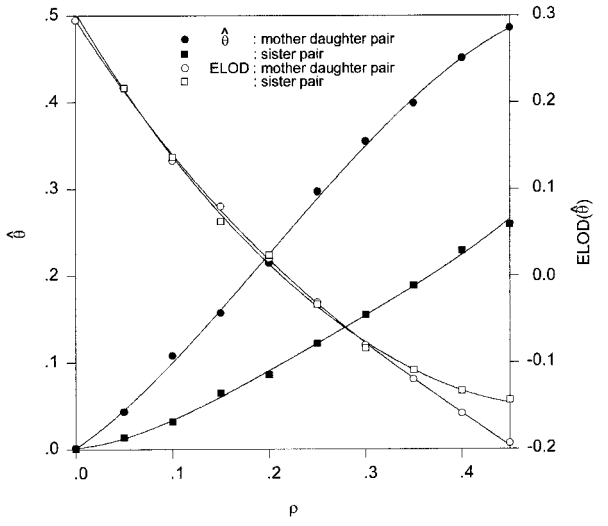

Figure 4 Expected LOD score and estimate of θ, by misspecification of ρ for a SS pair (*squares*) and a MD pair (*circles*); parameters are the same as those used in figure 3. Under the assumption of ρ = .00 in the analysis, maximum-likelihood estimates of θ (*unblackened symbols*) were determined by use of a modified version of ILINK (Rohde et al. 1995). The blackened symbols denote the corresponding expected LOD scores.

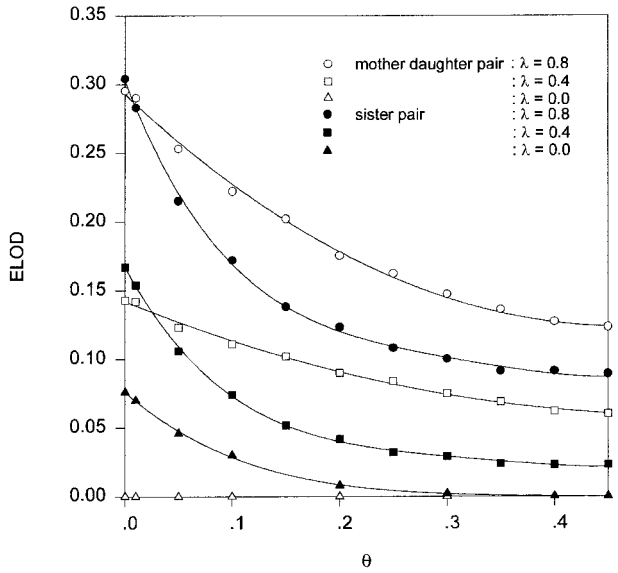

Figure 5 Expected LOD scores for different θ's and frequencies of loss, for a sib pair (*blackened symbols*) and a MD pair (*unblackened symbols*), for frequency of loss .8 (*circles*), .4 (*squares*), or .0 (*triangles*).

marked in a MD pair than in a SS pair. These observations suggest that ρ and θ should be jointly estimated.

Under some circumstances, MD pairs can be expected to be more informative than SS pairs. For a MD pair, only one meiosis takes place between the affected relatives, whereas, even if both sisters have inherited the same predisposing allele from the same parent, this allele is involved in one meiosis for each sister. Therefore, for constant ρ and increasing θ, the expected LOD score should decrease faster for a sib pair as compared with a parent-offspring pair. Figure 5 shows, for example, that, for an LOH frequency of .4 and a θ = .05, the expected LOD score for a MD pair is higher than that for a SS pair. These calculations were done under the assumptions that deletions extending to the marker locus affect the wild-type allele at the disease locus and that postzygotic recombination events between these two loci are rare enough to be neglected. These assumptions are likely to be correct for small values of θ.

In this communication we have investigated the gain of power expected from inclusion of LOH information in the analysis of genetic linkage of isolated pairs of affected first-degree relatives. We modified the SLINK program to simulate segregation and somatic loss of alleles in pedigrees. We have previously shown that isolated parent-offspring pairs can be used for establishment of genetic linkage when LOH information is included. We report here that, for loci with a high LOH frequency and consistent loss of the wild-type allele in the tumor, isolated parent-offspring pairs can be expected to be as informative as sib pairs. Therefore, iso-

lated parent-offspring pairs should be considered a valuable source of information for localization of loci that contribute to an inherited cancer predisposition.

References

Collins N, McManus R, Wooster R, Mangion J, Seal S, Lakhani SR, Ormiston O, et al (1995) Consistent loss of the wildtype allele in breast cancers from a family linked to the BRCA2 gene on chr 13q12-13. Oncogene 10:1673–1675

Cornelis RS, Neuhausen SL, Johanson O, Kelsell D, Ponder BAJ, Tonin P, Hamann U, et al (1995) High allele loss rates at 17q12-q21 in breast and ovarian tumors from BRCA1-linked families. Genes Chromosom Cancer 13:203–210

Cottingham RW Jr, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53:252–263

Easton DF, Bishop DT, Ford D, Crockford GP, The Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Lustbader ED, Rebbeck TR, Buetow KH (1995) Using loss of heterozygosity data in affected pedigree member linkage tests. Genet Epidemiol 12:339–350

Ott J (1989) Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci USA 86:4175–4178

Rebbeck TR, Lustbader ED, Buetow KH (1994) Somatic allele loss in genetic linkage analysis of cancer. Genet Epidemiol 11:419–429

Rohde K, Teare MD, Scherneck S, Santibáñez Koref M (1995) A program using loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families. Hum Hered 45:337–345

Sato T, Tanigami A, Yamakawa K, Akiyama F, Kasumi F, Sakamoto G, Nakamura Y (1990) Allelotype of breast cancer: cumulative allele losses promote tumour progression in breast cancer. Cancer Res 50:7184–7189

Steichen-Gersdorf E, Gallion HH, Ford D, Girodet C, Easton DF, DiCioccio RA, Evans G, et al (1994) Familial site-specific ovarian cancer is linked to *BRCA1* on 17q12-21. Am J Hum Genet 55:870–875

Teare MD, Rohde K, Santibáñez Koref M (1994) The use of loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families. J Med Genet 31:449–452

Varley JM, Thorncroft M, McGown A, Appelby J, Kelsey AM, Tricker KJ, Evans DGR, et al (1997) A detailed study of heterozygosity on chromosome 17 in tumours from Li-Fraumeni patients carrying a mutation to the TP53 gene. Oncogene 14:865–871

Weeks D, Ott J, Lathrop GM (1990) SLINK: a general simulation program for linkage analysis. Am J Hum Genet Suppl 47:A204