The use of loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families

M D Teare, K Rohde, M F Santibáñez Koref

Abstract

A method is described to investigate the inheritance of disease predisposition in cancer families. It is an extension of classic genetic linkage analysis, which enables information on loss of constitutional heterozygosity (LOCH) to be incorporated into the model. This adapted model treats LOCH data as additional observations on the disease phenotype. One of the major benefits of this approach is that isolated parent-offspring pairs are now potentially informative for linkage analysis. Examples are presented.

(J Med Genet 1994;31:448-452)

Inherited predisposition to a variety of cancers has been shown to be consistent with Knudson’s two hit hypothesis. According to this hypothesis inactivation of both copies of a tumour suppressor gene is necessary for certain tumours to occur. Patients with an inherited predisposition carry a defective allele of a tumour suppressor gene in the germline. In tumours from these patients the second allele is inactivated by a somatic event, such as the deletion of the corresponding chromosome region. In these cases it is not only the cosegregation of the cancer phenotype with certain markers which indicates the location of the defective gene, but also the loss of genetic material from a certain region of the genome in the tumours. This somatic event can be assessed as loss of constitutional heterozygosity (LOCH). Our aim is to examine how LOCH can be used to ascertain the location of genes predisposing to cancer.

Assuming the validity of the tumour suppressor model, loss of heterozygosity in the tumour material may represent an important source of additional information. It enables the use of small families, such as parent-offspring pairs which are uninformative using classical linkage analysis. These are not only easier to obtain than larger families but are also associated with defects which impair reproductive ability. This is the case, for example, when tumours at an early age are common among affected persons, as in the Li-Fraumeni syndrome.

Methods

Let us follow the nomenclature from Ott and consider a family pedigree of the size $m$. Let $x_i$ denote the phenotype and $g_i$ the genotype of the $i$-th pedigree member. Assuming that the phenotypes are, by given genotype, conditionally independent, the likelihood can be written as:

$$L = \sum_{x_1} \ldots \sum_{x_m} \prod_{i} P(x_i | g_i)$$

where $P(x_i | g_i)$ denotes probability of the $i$-th pedigree member to display a certain phenotype $x_i$ given a specific genotype $g_i$ and is usually designated as penetrance, and $P(g_1, \ldots, g_m)$ denotes the probability to observe the array of genotypes $(g_1, \ldots, g_m)$ in the pedigree.

We will consider a particular penetrance function which takes into account not only the development of a malignancy $d_i$ but also the presence of genetic alterations $l_i$ in the malignant tissue.

Let us consider in more detail the case of an inherited disease predisposition which follows Knudson’s two hit hypothesis and assume that a single marker is used to assess loss of constitutional heterozygosity.

At the disease locus, $d$ denotes the wild type allele and $D$ the disease predisposing allele:

$$(d, D)$$

$$g_i = \begin{cases} 
  d & \text{homozygous for the wild type allele} \\
  dD & \text{paternal allele disease predisposing} \\
  DD & \text{maternal allele disease predisposing} \\
  DD & \text{homozygous for the disease predisposing allele}
\end{cases}$$

At the marker locus, $l$ denotes that the allele is retained in the tumour and $0$ denotes that loss of the allele has been observed:

$$l = \begin{cases} 
  11 & \text{no LOCH observed} \\
  01 & \text{maternal allele lost} \\
  10 & \text{paternal allele lost} \\
  00 & \text{loss of both alleles detected in the tumour}
\end{cases}$$

Let us consider in more detail the case of an inherited disease predisposition which follows Knudson’s two hit hypothesis and assume that a single marker is used to assess loss of constitutional heterozygosity.
$P(d_i | g_i)$ denotes the probability to observe a malignancy $d_i$ given the genotype $g_i$. This function may also be dependent on factors such as age and sex.

$P(d_i | g_i, s)$ denotes the probability to observe LOCH at the marker locus in the malignant tissue, given a genotype $g_i$ and disease manifestation $d_i$. It will depend on the probability to detect the loss of one or both alleles at the marker locus and, in the presence of an inherited predisposition, on the probability $p$ that the allele lost at the marker locus and the disease predisposing allele are on the same chromatin strand in the germline. The latter can occur when the loss of one allele at the marker locus is the result of some other mechanism, like the presence of another tumour suppressor gene in the same chromosome, or through somatic recombination between the marker and the disease locus. Therefore $p$ will depend on the distance between disease and marker locus and their location in the chromosome as does the meiotic recombination fraction $\theta$. The rate of recombination between homologous chromosomes in somatic cells will determine the value of $p$. This rate is assumed to be very low since pairing homologous chromosomes outside meiosis is rare. If the tumour suppressor and marker are not linked then $p = 0.5$ and $\theta = 0.5$. Since somatic recombination between homologous chromosomes is less frequent than meiotic recombination, it is possible that in some cases $\theta = 0.5$ while $p < 0.5$. In those cases the location of the disease predisposing gene can be inferred in spite of a high meiotic recombination frequency ($\theta = 0.5$) between marker and disease locus.

Let $\lambda$ and $\epsilon$ designate the probability to observe loss of only one or both alleles respectively in the marker at the presence of an inherited disease predisposition, and $\lambda$ and $\epsilon$, the corresponding probabilities for a sporadic case.

Therefore if no malignancy has developed then

$$P(s_i = 1 | g_i, d_i = 0) = 1$$

In the case that the proband carries one cancer predisposing allele and has developed a tumour.

| $P(s_i = 0 | g_i, d_i = 1)$ | both alleles lost at the marker locus in the tumour; |
|-----------------------------|----------------------------------------------------|
| $P(s_i = 0 | g_i, d_i = 1)$ | $\epsilon_0$ one allele is lost at the marker locus and it is in phase with the disease predisposing allele in the germline; |
| $P(s_i = 1 | g_i, d_i = 1)$ | $\lambda_0 p$ if the detected loss is in the opposite phase to the disease predisposing locus in the germline; |
| $P(s_i = 1 | g_i, d_i = 1)$ | $1 - \lambda_0 - \epsilon_0$ no allele loss in the tumour is detected; |

and if it is a sporadic case:

| $P(s_i = 0 | g_i, d_i = 1)$ | both alleles lost; |
|-----------------------------|----------------------------------------------------|
| $P(s_i = 1 | g_i, d_i = 1)$ | $\lambda_0 2$ one allele is lost; |
| $P(s_i = 1 | g_i, d_i = 1)$ | $1 - \lambda_0 - \epsilon_0$ no allele loss is detectable. |

We have further to consider the case that the proband is homozygous for the predisposing allele

| $P(s_i = 0 | g_i, d_i = 1)$ | both alleles are lost; |
|-----------------------------|----------------------------------------------------|
| $P(s_i = 1 | g_i, d_i = 1)$ | $\epsilon_1$ one allele is lost; |
| $P(s_i = 1 | g_i, d_i = 1)$ | $1 - \lambda_1 - \epsilon_1$ no allele loss is detectable. |

**Examples**

Let us consider the family represented in fig 1A and use the following additional designations:

$f_i$: frequency of allele $a_i$ and the following abbreviations for the function $P(d_i | g_i)$:

$P_i$: the cumulative probability for a person with genotype $DD$ to develop cancer by age $a_i$, where $a_i$ is the age of the $i$-th person.
$p_i(1)$: the probability for a person with genotype Dd to develop cancer at age $a_i$, where $a_i$ is the age at onset of disease for the $i$-th person.

$P^2(i)$: as in $P_i(1)$ but for a $dd$ genotype, $p_i(1)$: as in $p_i(1)$ but for a $dd$ genotype.

The disease status is censored for unaffected persons, and the likelihood must be calculated according to survival analysis methods.\(^3\) We will assume that de novo mutations at the disease locus are infrequent enough to be neglected and that $f_0$ is small enough to disregard terms of the order of $f_0^2$ or higher.

The likelihood for this pedigree assuming the following phase

\[
\begin{array}{c|c|c}
\text{MOTHER} & \text{FATHER} \\
\hline
D & d \\
1 & 2 \\
3 & x \\
d & d
\end{array}
\]

is

\[ p_d(1) (1 - P_c(2)) \]

\[ \left(1 - \theta p_s(3) \lambda_d(1 - p) + \theta p_s(3) \frac{\lambda_d}{2}\right)^{1/2} \]

\[ (p_s(4) + p_d(4)) \]

The alternative phases can be calculated correspondingly.

We obtain for the likelihood the following expression

\[
L(0, \theta) = f_d f_s (p_s(4) + p_d(4))/p_d(1) \]

\[ (1 - P_c(2)) \left( p_d(3) \lambda_d(1 - p) + p_s(3) \frac{\lambda_d}{2}\right) \]

\[ p_d(1) (1 - P_c(2)) (p_s(3) \frac{\lambda_d}{2} + p_s(3) \lambda_d p) \]

\[ + f_s f_p (p_s(1) (1 - P_c(2)) p_s(3) \frac{\lambda_d}{2}) \]

\[ + f_s f_p p_s(1) (1 - P_c(2)) p_s(3) \frac{\lambda_d}{4} \]

\[ + f_s f_p p_s(1) (1 - P_c(2)) p_s(3) \frac{\lambda_d}{4} \]

This likelihood does not depend explicitly on $\theta$ and shows that this pedigree would be uninformative if only the constitutional genotypes are considered.

The lod score is

\[ \log_{10} \left( \frac{L(0, \theta)}{L(1/2, 1/2)} \right) \]

We can reduce the number of parameters by setting $\lambda_c = \lambda_i$ and $\epsilon_c = \epsilon_i$. This simplification seems reasonable since estimates of $\lambda_i$ and $\epsilon_i$ may be difficult to obtain when dealing with small families and when there is no possibility to distinguish between cancers resulting from an inherited predisposition and sporadic cases. It is apparent from equation (2) that for $\lambda_i = \lambda_i = \lambda$ equation (3) becomes independent from $\lambda$. An estimate of $\lambda$ is in that case not required for the computation of the lod score. On the other hand, this reduction in the number of parameters will lead to loss of information.

The pedigrees in the example represents a Li-Fraumeni family. About half of these families have germline p53 mutations in the exons or at the splice junctions.\(^3\) In some of the remaining families it is unlikely that p53 is the primary lesion since affected relatives do not share p53 alleles.\(^3\) In other families, aberrations elsewhere in the gene cannot be excluded. It is therefore important to estimate the proportion of families which are linked to p53 in order to assess the relevance of other loci for this syndrome. The use of LOCH data (additional information) will improve the estimate of the proportion of families with p53 defects, since families which would be otherwise uninformative can be used for the analysis.

Let us assume that the typing results shown in fig 1 were obtained with a marker at the TP53 locus. Estimates of the penetrance and allele frequencies for such families are given by Lustbader et al.\(^4\) Using these estimates, that is, $p_1(1) = 0.0009$, $p_1(1) = 0.02407$, $P_c(2) = 0.0255$, $P_c(2) = 0.06697$, $p_3(3) = 0.00013$, $p_3(3) = 0.02016$, $p_4(4) = 0.00013$, $p_4(4) = 0.02016$, and $f_0 = 0.00005$, and setting $\theta = 0$, $\theta = 0$ (marker at the disease locus) and $\lambda_s = \lambda_i$, we obtain a lod score of 0.282.

Fig 1B shows the influence of $\lambda_s$ and $\lambda_i$ on the lod score. For $\lambda_s = 0.2$ and $\lambda_i = 1$ the lod score is 0.238 while for $\lambda_s = 1$ and $\lambda_i = 0$ it is 0.295. These results show that the lod score is little influenced by a wide range of $\lambda_s$ and $\lambda_i$ values.

Fig 2A represents a breast/ovarian cancer mother-daughter pair. The likelihood is described by the equation.

\[
L(0, \theta, \rho) = f_d f_s (p_s(1) (1 - P_c(2)) \lambda_i (p_s(3) \lambda_i (1 - p) \]

\[ + (1 - \theta) (1 - p) + \theta p_s(3) \frac{\lambda_i}{2} (1 - p) + (1 - \theta) p) \]

\[ + f_s f_p (p_s(1) (1 - P_c(2)) \lambda_i (p_s(3) \lambda_i + p_s(3) \frac{\lambda_i}{2}) \]

\[ + f_s f_p p_s(1) (1 - P_c(2)) p_s(3) \frac{\lambda_i}{4} \]

\[ + f_s f_p p_s(1) (1 - P_c(2)) p_s(3) \frac{\lambda_i}{4} \]

Equation (4) depends on $\theta$ and $\rho$. Therefore a mother–daughter pair, which would not be informative without LOCH data, can be used to estimate $\theta$ and $\rho$. As shown in fig 2B, a lod score close to log$_{10}$ 4 can be obtained for $\theta = \rho = 0.02$. It also shows that for $\lambda_s \geq \lambda$, the lod score depends little on $\lambda_s$ and $\lambda_i$.

**Discussion**

The method we propose will enable linkage analysis to be undertaken on series of cancer family clusters that previously lacked sufficient statistical power to detect linkage. It also
The use of loss of constitutional heterozygosity data to ascertain the location of predisposing genes in cancer families

Figure 2. (A) Mother-daughter breast ovarian cancer pair. The mother (1) was diagnosed with ovarian cancer at the age of 40 and the daughter (3) had a breast cancer diagnosed before the age of 30. The father is aged 60 and unaffected. The mother and daughter are typed constitutionally and for tumour. (B) Dependency of the lod score on $\lambda_0$ and $\rho$. Estimates for allele frequencies and penetrance are from Easton et al. and assuming $\theta = 0.02, \rho = 0.02$.

means that isolated parent-offspring pairs, which are always uninformative for classic linkage analysis, are now potentially informative if tumour material is available. Although this new approach can be applied to any genetic analysis in which a tumour suppressor model is appropriate, its real benefit is in the analysis of those rare syndromes where the age at onset of the tumour directly interferes with reproductivity or the onset of the disease itself is associated with infertility. Large families often do not exist for these syndromes. Collecting sufficient sib pairs is very difficult in these situations when it is often the isolated parent-offspring pairs which are more common. This is the case in small families with clusters of malignancies, especially childhood tumours, loosely fitting the criteria of the Li-Fraumeni syndrome. Although germline p53 mutations have been shown in such families, a significant proportion of even classic Li-Fraumeni families do not show any alterations in the exons or the intron-exon boundaries of that gene. The demonstration of “non-random” loss of p53 alleles may help to decide if further efforts should be devoted to investigating other regions of the gene in particular families and help to estimate the proportion of these families which is the result of germline p53 defects. Similarly, somatic events may be used to determine the proportion of families with an inherited predisposition at a certain locus, although the gene itself may not be known. This is the situation for the BRCA1 locus on 17q in breast cancer. A two step procedure could be envisaged in such cases. Large families known to be linked could be used to estimate the necessary parameters and then using these parameters larger series including parent-offspring pairs could be analysed.

This kind of approach is of course sensitive to effects like imprinting, especially if a large proportion of the studied material comes from mother-daughter pairs. However, the absence of loss from paternal alleles in series which should contain a significant number of sporadic cases would be noticeable.

Another feature of syndromes associated with tumour suppressor genes is the case of multiple primary tumours. In breast ovarian cancer families it is sometimes the case that a woman develops separate primaries. If we assume that the multiple cancers are independently formed tumours, a further extension of the model described could incorporate this additional information.

It is conceivable that patients with a predisposition show a higher frequency of loss in the tumour than those without. The probability of losing one allele in genetic cases should be higher, since only one event would be necessary to achieve growth advantage while in sporadic cases two events affecting that locus are required and alternative pathways of tumour development may be followed. In genetic cases the loss of the wild type allele should also be easier to detect since as an early event it should be present in all the malignant cells. If this argument is correct the detection of a deletion would in itself suggest that there is a predisposing gene in the deletion area. For example, LOCH for markers on 17q seems to be more frequent in ovarian tumours from patients belonging to families linked to the BRCA1 locus than in persons without a known predisposition mapping to that locus. Inclusion of the information given by the difference in frequency of loss in linked v unlinked cases requires estimates not only for $\theta$ and $\rho$ but also for $\lambda_{0,\mu}$, $\lambda_\mu$, and $\lambda_{\mu}$. According to Knudson’s two hit hypothesis, deletion of one allele should not confer any advantage to cells in which both copies of a tumour suppressor are inactive because they carry a mutation. We would expect therefore loss of one allele to be rare in tumours of patients homogygous for a cancer predisposing allele. It is difficult to estimate $\lambda_{0,\mu}$ empirically because of the low frequency of disease associated alleles in the population.

Reported LOCH frequencies can vary widely even if the same markers and comparable tumour types are used. In breast tumours for example, Bierch et al. detected 41% (49) of 121 LOCH using the probe merH on 7q, while Devilee et al. did not find any loss in 32 samples for the same locus (see also Corneille). These discrepancies may be because of the amount of non-malignant material in the samples analysed and the criteria used to assess LOCH. The use of a single parameter
\( \lambda(\lambda = \lambda_H = \lambda_s = \lambda_c) \) will lead to some loss of information but should not bias the results.

The correlation between the frequencies of deletion of one or both alleles at a marker locus will depend on the location of the marker itself. Loss of both alleles will be rare if the marker is close to a gene for which the presence of one copy is essential in a viable cell.

Another factor contributing to the scarcity of reports of homozygous deletions is the increasing use of PCR formatted markers for which it is easier to assess allelic imbalance than the absolute levels of each allele. Contamination with normal tissue will mask homozygous deletions in such cases.

The method outlined assumes a simplest case tumour suppressor gene model. This is a natural starting point and this approach can be further modified. The more general question is the use of information from somatic events to assess the likelihood of the involvement of a certain locus in situations in which somatic events point to the location of the predisposing locus. This is the case in inherited cancer predisposition when the classical two hit hypothesis is valid. A bias in the loss of alleles should also be observed when the predisposing allele acts in a dominant fashion at the cellular level but lies in proximity to a tumour suppressor gene, since alleles carrying the activated oncogene would be preferentially retained.

The described adaptations are currently being incorporated into a linkage analysis program in order to be applied to real data.

This work was supported by the Cancer Research Campaign. We would like to thank Professors D.G. Hammond and S. Scherneck, and Drs D Scott and W Schmidt for encouraging discussion.