Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 11814–11818, October 1996
Medical Sciences

Studies on locus content mapping

(radiation hybrids/genetic mapping/map integration/human genome)

J. W. TEAGUE, A. COLLINS, AND N. E. MORTON

Human Genetics, University of Southampton, Level G, Princess Anne Hospital, Southampton SO16 5YA, United Kingdom

Contributed by N. E. Morton, June 24, 1996

ABSTRACT **Locus content maps are derived from monosomic or disomic chromosomes broken by radiation, shearing, or other clastogen, the fragments being distributed among clones by dilution or incorporation into the cells of another species and scored for segregation of markers. Locus content maps provide evidence about radiosensitivity of chromosome regions, support for order, and approximate location. Omission of the most aberrant and least informative clones increases efficiency of localization. Correct analysis must allow for preferential retention of certain sequences, monosomy or polysomy of donor chromosomes, and error filtration. Combination of these refinements extracts substantially more information from fewer clones. Because of unmodeled peculiarities in the data, the best analysis does not recover the physical map but roughly localizes markers that may be monomorphic and therefore unsuitable for linkage mapping. As with linkage for polymorphic loci, distance in the composite map should be confirmed by physical methods.**

Data for a locus content map consist of observations on presence or absence of loci or alleles in chromosome fragments of unspecified length produced by random breakage. The first locus content maps used protein markers for fragments of disomic human chromosomes broken by x-rays and incorporated into mammalian cells (1). Each clone could give information about all chromosomes. After introduction of restriction fragment length polymorphisms, an alternative method based on monosomic human chromosomes from a somatic cell hybrid was developed (2). It is less efficient, since each clone gives information about only one chromosome. Although nearly all locus content mapping uses radiation hybrids, the sensitivity of DNA amplification by the polymerase chain reaction (PCR) allows clones to be established by dilution of fragments produced from disomic (or in principle monosomic) human chromosomes without rescue by a cell of another species (3). Obviously these approaches are not limited to the human, but applications to other organisms are rare.

The object of these studies is to derive a connected map of the loci examined, giving their order and (if breakage is uniform) their physical location. If breakage is not uniform, regions of differential sensitivity to the clastogen may be identified. The value of the data depends on the locus content map, since somatic cell panels established from these experiments contain more allogenic DNA than other vectors and are less stable than cosmid contigs. Locus content mapping is not required for connectivity or physical localization, since it must compete with sequence tagged site content maps, yeast artificial chromosome and cosmid contigs, fluorescence *in situ* hybridization, and projection of the genetic or chiasma map. For a fair comparison, the data must be analyzed under a valid model that spans monosomy, disomy, and synthetic polysomy induced by pooling of clones.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Fragments of unspecified length produced by radiation and shearing have been assumed to be exponentially distributed (4). Both presence and absence of loci are scored. The theory for monosomic fragments has been extended to disomy (5). Each experiment has a dose-dependent scale in arbitrary units, but on the hypothesis of uniform breakage the partial map determined from such data may be scaled to megabases (Mbs). Whether or not breakage is uniform, construction of a linear map from DNA fragments broken by a clastogen constitutes locus content mapping, following the convention of "sequence tagged site content mapping" in contigs (6, 7). Locus content is an attribute of a clone and should not be confused with arbitrary attributes of loci (8).

It has been observed that each chromosome has at least one point, usually near the centromere, that is maximally retained in somatic cell hybrids, the retention frequency declining with distance from this point; telomeres may function the same way (4). Perhaps these sequences stabilize mitotic transmission of the chromosome fragment. Alternatively, rarity of expressed sequences in these regions may reduce selection against human fragments in somatic cell hybrids. Elimination of the mammalian host presumably leads to uniform retention, although this has not been studied. "Pushmi-pullyu" hybrids, selected for retention at one side of an interval and for loss at the other side, should have a smooth, steep gradient. The model for retention must accommodate these possibilities yet be robust to departure from a monotonic decline in retention frequency.

A further complication is that deviations from uniform breakage appear to occur, in part because of factors that may be grouped as "error." They include false positives and negatives due to nonspecific or weak probes, recording errors, and clonal chimerism. There may well be a proportion of double breaks caused by a single ionization or shear, analogous to the linear component in dominant lethals, which increase approximately as the 1.6th power of the mutagenic dose rather than as the quadratic expected for two-hit phenomena. Such "errors" tend to affect single loci even in a dense map, making pairs with other loci uninformative about order. A valid analysis must provide a statistical filter for all these errors by incorporating them into the probability model (9).

In this paper, we develop a general model incorporated in the computer program MAP+ (10) and apply it to monosomic, disomic, and polysomic data under different hypotheses to assess the affect of erroneous approximations. The model includes error filtration assuming a uniform distribution of errors for both retention and loss. This is oversimplified, but locus-specific errors seem intractable. Following recent developments in map integration, the location data base LDB+ is automatically trawled for the best current physical map to begin iteration. MAP+ and LDB+ are available on the World Wide Web at http://cedar.genetics.soton.ac.uk/public_html or anonymous ftp at cedar.genetics.soton.ac.uk in the directory/pub.

Abbreviation: Mb, megabase.

Monosomy

Monosomic probabilities without error filtration have been derived (4). Consider a pair of loci h and i, where h is closer to the point of maximal retention (which for brevity we shall call the centromere, as is usually the case). The four outcomes may be denoted hi, $h.$, $.i$, and .., where a letter signifies retention of the locus and a dot signifies loss (see Appendix). Missing or ambiguous observations are neglected, and clones retaining all markers or no markers may be discarded. In the absence of error, the expected retention frequency for the ith locus would be

$$Q_i = H \exp[-Mw_{ci}] + (1 - \exp[-Mw_{ci}])\, L$$

where w_{ci} is the distance between the centromere and i in Mb, H is the retention probability of the centromere, M is the ratio of map length in R units [rays (the unit of length with a mean of 1 break at the given clastogen dosage)] and Mb, and L is the retention rate after a break has occurred.

The conditional probability that i is lost given loss of h would be

$$S. = \exp[-Mw_{hi}] + (1 - \exp[-Mw_{hi}])\,(1 - L)$$

Our error model replaces Q_i by

$$B_i = (1 - C)\, Q_i + C\,(1 - Q_i),$$

where C is the error frequency for retention and $S.$ by $(1 - E)$ $S. + E\,(1 - S.)$, where E is the error frequency for loss.

Then the four joint probabilities are

$$P_{..} = (1 - B_h)\,[(1 - E)\, S. + E\,(1 - S.)]$$

$$P_{.i} = 1 - B_h - P_{..}$$

$$P_{h.} = 1 - B_i - P_{..}$$

$$P_{hi} = 1 - (1 - B_h) - (1 - B_i) + P_{..}$$

with $\Sigma P = 1$.

Polysomy

This theory is readily extended to polysomy on the assumption that error acts on each donor chromosome independently. Then the joint probabilities for n-somy are

$$P(..) = (P..)^n$$

$$P(.i) = (1 - B_h)^n - (P..)^n$$

$$P(h.) = (1 - B_i)^n - (P..)^n$$

$$P(hi) = 1 - (1 - B_h)^n - (1 - B_i)^n + (P..)^n$$

The marginal retention probability is $P_i = 1 - (1 - B_i)^n$ and the conditional probabilities are

$$P(hi\,|\,h+) = [1 - (1 - B_h)^n - (1 - B_i)^n + (P..)^n]/P_h$$

$$P(..\,|\,h-) = [P../(1 - B_h)]^n = P(..)/(1 - P_h)$$

These equations hold for monosomy ($n = 1$) and disomy ($n = 2$), as well as in the general case.

Panel Selection

Every set of radiation hybrids for a particular chromosome arm that we have examined has two gross departures from our model (and all published models). On one hand, there are clones with a great excess of negative reactions, the few positive ones suggesting errors that in our model have probability $\lambda = 1 - (1 - C)^n$. These "errors" may either be due to typing or represent inclusion of short human sequences by a process different from incorporation of multiple loci. On the other hand, there are clones with a great excess of positive reactions, the few negative ones suggesting errors that in our model have probability $\in = 1 - (1 - E)^n$. These "errors" may be due either to typing or to excision of short sequences from an otherwise intact human chromosome arm. Since these phenomena are not part of the model, we introduce an approximate likelihood,

$$P(\text{clone } j) = \begin{cases} \lambda^r(1 - \lambda)^m \text{ if the human arm is lost} \\ \in^m(1 - \in)^r \text{ if the human are is retained} \\ \prod P_k \text{ if the human arm is partly retained} \end{cases}$$

where r is the number of positive reactions, m is the number of negative reactions, and the P_k are taken from the polysomic distribution. For the most proximal scored locus h, the value of P_k is P_h if positive and $1 - P_h$ if negative. For successive pairs of scored loci h,i

$$P_k = \begin{cases} P(hi\,|\,h+) \text{ if both are positive} \\ P(..\,|\,h-) \text{ if both are negative} \\ 1 - P(hi\,|\,h+) \text{ if } h + i - \\ 1 - P(..\,|\,h-) \text{ if } h - i + . \end{cases}$$

Clone j is considered to be a gross error if

$$\begin{cases} \lambda^r(1 - \lambda)^m/\prod P_k > 100 \\ \text{or } \in^m(1 - \in)^r/\prod P_k > 100. \end{cases}$$

Exclusion of aberrant clones usually increases agreement with the physical map.

A clone is relatively uninformative if $q = 4\,rm/(r + m)^2 \ll 1$ (11). We found that clones with $q < 0.1$ did not contribute appreciably to agreement with the physical map. Exclusion of aberrant and uninformative clones leaves a subset of selected clones. The mean retention frequency of selected clones is near 0.5. Censoring of the panel substantially reduces the number of tests required to map a new marker and is preferable to raising the retention frequency by random pooling of clones that may be gross errors (11).

Numerical Analysis

Data on the ith locus content panel are kept as file r_j in the location data base (LDB+) chromosome directory (10). The variables are locus, clone, and score, where loci are rows and clones are columns. Because of its similarity to linkage analysis, radiation hybrid mapping is assigned to the MAP+ program (10).

Each locus is scored for presence (1) or absence (0) of loci, with blanks and other codes being treated as unknown. Clones that do not contain at least two scores of 0 and at least two scores of 1 are omitted by the program. Optionally, other clones may be excluded by a delete command. The remaining data are summarized as retention frequencies for loci and conditional concordance frequencies for pairs of loci. Initially, there are up to six parameters that must be estimated: the retention probability of the centromere (H), the retention probability of loci far from the centromere (L), the ratio of map length in R units (rays) and in Mb (M), and the two error frequencies C and E. The first analysis tests hypotheses about these parameters by constraining to the composite (physical) map in the location database LDB+. The physical map is constructed by projection of the genetic linkage map formed by fitting a double logistic curve to each chromosome arm. This projects the sex averaged genetic distance onto the physical

scale through the cytogenetic band midpoint (scaled to megabases) (12, 13). Curve fitting is weighted by the inverse of the cytogenetic band width. A recent study using fluorescence *in situ* hybridization to localize yeast artificial chromosomes shows substantial agreement for most loci (14), although there are fewer loci in the physical map. Loci not present in the physical map are assigned the location with the highest correlation in the 2×2 contingency tables for pairs of loci. Maximum likelihood iteration uses finite differences to approximate derivatives.

Given estimates of the parameters (which in some circumstances might be consistent with $H = L, L = 0, H = 1$, or $C = E = 0$), constraining a locus content map to physical data is one way to obtain map connectivity. This has been the motive for much of the interest in R mapping. Another motive is to detect through discrepancies between locus content and physical maps regions that are sensitive or resistant to the given clastogen. To detect discrepancies, the mapping instructions are first performed under the order constraints of physical data. Then the constraints are removed and the significance of improvement is tested by likelihood rato (LR) χ^2. This cycle may be repeated with a new trial map to verify stability of the estimates.

The supported locus content map is kept as a partial (p) map and integrated into the *rhmb* field of the summary map (12). In our experience, locus content maps are not accurate enough to compete with physical maps, and so we prefer to interpolate loci missing from physical maps into the composite map that integrates genetic and physical data. A standard error conditional on the rest of the map indicates the reliability of this location (10). Since two loci rarely have exactly the same patterns of reactions (due in part to error), "resolution" greatly exceeds reliability.

Results

We analyzed monosomic (2) and disomic (15) hybrids for human chromosome 21 (Table 1). These and other recent studies on radiation hybrids show three peculiarities that had not been noticed before. First, radiosensitivity depends on chromatin composition. G bands stain brightly with Giemsa and trypsin unless Giemsa is followed by heat treatment, which quenches G bands. They are A+T rich and tend to have a low density of expressed genes and microsatellites, a low frequency of recombination, and to be radioresistant (16). R bands stain dully with Giemsa and trypsin but brightly if Giemsa is followed by heat treatment. They are G+C rich and tend to have a high frequency of expressed genes and microsatellites, a high frequency of recombination, and to be radiosensitive. The most G+C rich R bands, conventionally called T bands, have these tendencies to the greatest degree. Within G and R categories, there are consistent differences among bands in staining intensity. Differences in chromatin content largely account for a 9-fold variation in radiosensitivity as kb/cM noted for intervals defined by linkage (17), and a 3-fold variation among chromosomes (18), but the correlation between G+C content and recombination frequency distorts relation to the physical map. Fig. 1 shows a comparison of distances in monosomic and disomic RH maps with the physical map. There are a small number of order discrepancies and evidence of expansion in the radiation hybrid maps for T band 21q22.3 when compared with the physical map, and noise in the rest of the chromosome.

Table 1. Chromosome 21 panels

Panel (ref. no.)	Genomes	Dose, rads	All clones	Segregant clones	Selected clones	Loci
Cox (2)	1	8000	99	69	67	32
Gene bridge 4 (15)	2	3000	93	52	46	56

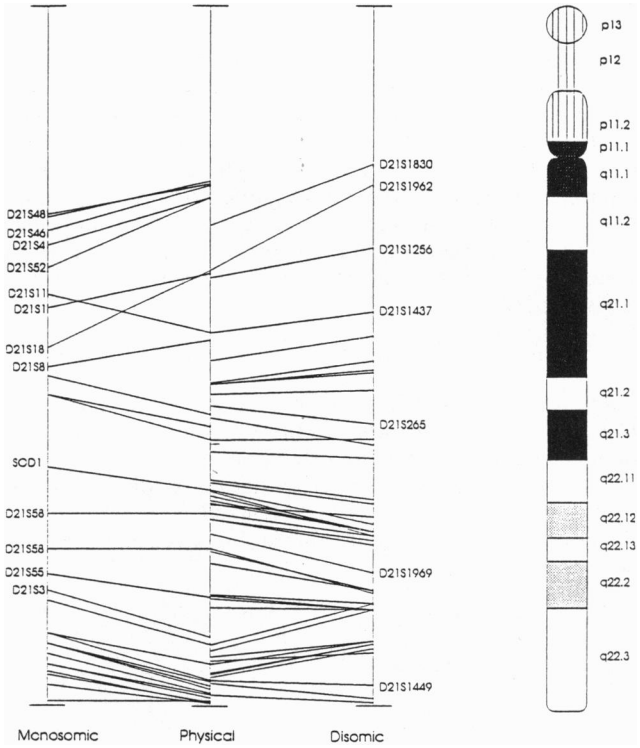

FIG. 1. The effect of chromatin content on radiosensitivity. Radiation hybrid maps (n = 16, see Table 2) are expanded for T band 21q22.3. Selected loci are located.

A second peculiarity of radiation hybrid maps is that metacentric chromosomes have extreme expansion around the centromere when both arms are mapped simultaneously (15, 18). Fig. 2 shows the frequency of joint retention for adjacent (ordered) markers against their midpoint physical location on chromosome 2. The low joint retention of markers close to the centromere is striking and there is an abrupt transition distinguishing p and q arms. This cannot be explained by the low density of microsatellites in pericentric regions, associated with low radiosensitivity, but is a logical consequence of mitotic selection against dicentric and acentric chromosomes. If the rodent centromere is retained, as may usually be the case, an incorporated human fragment will be strongly selected against unless it has lost its centromere and therefore the proximal

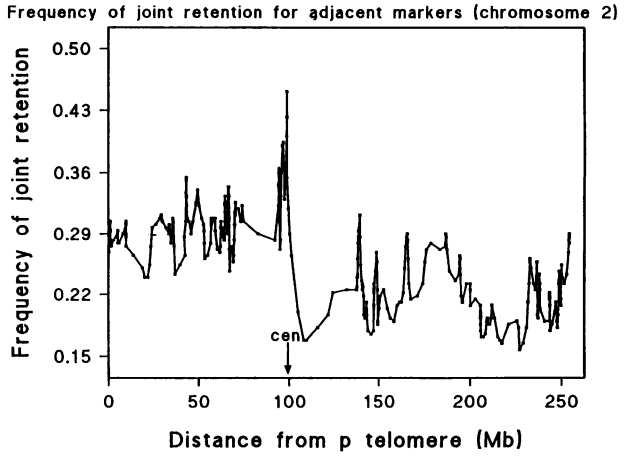

FIG. 2. Frequency of joint retention for adjacent markers (gene bridge 4 panel, chromosome 2).

Table 2. Effects of assumed n-somy on χ^2 for selected clones

Panel (ref. no.)	somy n	χ^2	H	L	M	E	C
Cox (2)	1	1106.3	1.000	0.317	0.177	0.025	0.000
	2	1085.0	0.826	0.172	0.177	0.011	0.000
	4	1075.0	0.504	0.090	0.176	0.005	0.000
	16	1067.7	0.144	0.023	0.175	0.001	0.000
Gene bridge 4 (15)	1	2530.0	1.000	0.378	0.128	0.043	0.000
	2	2493.1	0.501	0.208	0.105	0.024	0.000
	4	2478.2	0.296	0.109	0.103	0.012	0.000
	16	2469.1	0.084	0.028	0.101	0.003	0.000

part of the other chromosome arm. If the rodent centromere is lost, incorporation of a centric human fragment will usually entail loss of a chromosome arm, since retention decreases distally. Centromeric expansion is minimized by mapping each arm separately as MAP+ does. Somewhat surprisingly, Fig. 2 also indicates a lower mean retention frequency for q arm markers than for those on the p arm. The reason for this is not known.

A third and somewhat surprising feature of these maps is that goodness of fit increases with the assumed level of polysomy (Table 2). Since clones were not pooled to induce polysomy, this must reflect departure from Poisson breakage. There is a small effect on map distances and on inferred order of less securely placed markers, but this is negligible compared with other sources of error (Table 3). Conditional standard errors are greater than 160 kb. Since this does not include errors in the model and in order, greater precision requires a different method.

The effect of these three peculiarities is that proportionality between locus content and physical maps is only a rough approximation. Within this constraint, the information in a locus content map is maximized by error filtration (by estimating C and E), allowance for preferential retention ($H \neq L$), and selection of clones. Table 3 shows that omission of these refinements, which are present in MAP+ but absent in other programs (11, 15), loses much of the information on order and distance measured as efficiency by dividing χ^2 for the full model by χ^2 for the simplified model. Estimates of error frequencies and other parameters seem reasonable (Table 2), but the hypothesis that response is proportional to radiation dose is not supported. This discrepancy may reflect problems in dosimetry, variation in breakage repair, or complexity of radiosensitivity.

Discussion

Locus content mapping has been almost entirely restricted to radiation hybrids, and it is not known whether other clastogens or recovery protocols show the same or other peculiarities. If radiosensitivity of G+C rich chromatin is due to the larger target of transcriptionally active loci, shearing and some other clastogens might act in the same way, while protocols that do not depend on incorporation of fragments into an intact chromosome impose different constraints.

Whatever the test system, the locus content map is likely to depart systematically from the physical map. Moreover, dose response has not been demonstrated to follow any particular law and probably includes both 1 hit and 2 hit breakage. Breakage is more frequent in interphase than during mitosis, and in practice a mixture of interphase and different mitotic stages is radiated. The initial response is modified by rejoining of fragments to an extent that depends on temperature and time until cell fusion (19). The final response as rays per megabase (R/Mb) does not have a simple relation to dose.

The subsequent fate of a human locus depends to an unknown extent on selection in a rodent cell, which increases with the number of cell divisions. The mechanisms by which a human fragment is incorporated into a rodent cell are unknown. Do they require sequence homology at one or both ends, at what stage in the cell cycle does incorporation take place, and are human centromeres and telomeres frequently incorporated? Analytic methods and correspondence with the physical map must remain primitive so long as locus content mapping is a black box.

Now that the assumption of uniform breakage must be abandoned, the investigator is free to exclude aberrant and uninformative clones and to augment the panel with clones (not necessarily derived from a clastogen) that increase resolution in regions of interest, without fear that this will perturb a proportionality to the physical map that does not in fact exist. Locus content mapping remains one of the most efficient ways to assign an approximate location to loci that may be monomorphic, competing well with fluorescense *in situ* hybridization and linkage, including meiotic mapping panels. Loci assigned in this way may be integrated into a locus content map by algorithms implemented in LDB+ (10) that scale with respect to two nearby (and preferably flanking) markers common to the partial and summary maps. Loci without physical assignment to a yeast artificial chromosome or other contig may be integrated into a composite map by the same algorithm, but physical maps should continue to have higher priority than locus content maps.

Appendix: Maximum Likelihood Analysis

The observed retention frequency is $N_{1i}/(N_{1i} + N_{0i})$, where the N_i are the observed numbers for locus i and 0, 1 represent retention or loss, respectively. For any parameter θ, the maximum likelihood score is

$$u_{i\theta} = \frac{\partial \ln L_i}{\partial \theta} = \left(\frac{\partial \ln L_i}{\partial P_i}\right)\left(\frac{\partial P_i}{\partial \theta}\right)$$

where the log likelihood is

Table 3. Effects of simplifying assumptions on χ^2 for selected clones

Panel (ref. no.)	Model, $n = 16$	χ^2	H	L	M	Goodness of fit χ^2	df	Efficiency
Cox (2)	$C{=}E{=}O$	1076.1	0.173	0.024	0.202	8.4	1	0.99
	$C{=}E{=}O$							
	$H{=}L{=}\hat{P}$	1275.0	0.025	0.025	0.196	207.3	2	0.84
	$C{=}E{=}O$							
	$H{=}L{=}\tilde{P}$	1765.7	0.033	0.033	0.193	698.0	3	0.60
Gene bridge 4 (15)	$C{=}E{=}O$	2728.1	0.134	0.029	0.141	258.9	1	0.91
	$C{=}E{=}O$							
	$H{=}L{=}\hat{P}$	2799.1	0.030	0.030	0.142	329.9	2	0.88
	$C{=}E{=}O$							
	$H{=}L{=}\tilde{P}$	3273.8	0.036	0.036	0.138	804.6	3	0.75

$\hat{P}$, estimated by maximum likelihood; $\tilde{P}$, estimated as $\tilde{P} = 1 - \sqrt[n]{B/N}$ where there are B nonretentions amongst a total of N.

$$ln\ L_i = N_{1i}\ ln\ P_i + N_{0i}\ ln\ (1 - P_i)$$

and $\partial Pi/\partial\theta$ is estimated by finite differences. The same principle applies to pairwise frequencies, giving for conditional retention

$$u_{hi\theta} = \frac{\partial\ ln\ L_{hi}}{\partial\theta} = \left(\frac{\partial\ ln\ L_{hi}}{\partial P(hi\,|\,h+)}\right)\left(\frac{\partial\ P(hi\,|\,h+)}{\partial\theta}\right)$$

and for conditional loss

$$u_{hi\theta} = \left(\frac{\partial\ ln\ L_{hi}}{\partial\ P(..\,|\,h-)}\right)\left(\frac{\partial\ P(..\,|\,h-)}{\partial\theta}\right).$$

The total score is

$$U_\theta = \sum_i u_{i\theta} + \sum_{h<i}\sum u_{hi\theta}.$$

For a pair of parameters θ, θ', the information in the j^{th} informative datum with expected frequency ρ is

$$k_{j\theta\theta'} = \left[E\left(-\frac{\partial^2\ ln\ L_j}{\partial\ \rho_j^2}\right)\right]\left(\frac{\partial\ \rho_j}{\partial\theta}\right)\left(\frac{\partial\ \rho_j}{\partial\theta'}\right)$$

where

$$E\left(-\frac{\partial^2\ ln\ L_j}{\partial\ \rho_j^2}\right) = \frac{N_{1j} + N_{0j}}{\rho_j(1 - \rho_j)}$$

and the total information is

$$K_{\theta\theta'} = \sum_j k_{j\theta\theta'}$$

Newton–Raphson iteration on the vector of estimated parameters is $\theta \rightarrow \theta + UK^{-1}$, where K^{-1} is the inverse of the K matrix. At convergence, the standard error is SE $(\theta) = \sqrt{K_{\theta\theta}^{-1}}$. Goodness of fit is tested by the likelihood ratio $\chi^2 = 2\ \Sigma_k\ N_k\ ln\ (P_k\hat{P}_k)$ under the constraint $0\ ln\ 0 = 0$ and with degrees of freedom n^2-m-p, where n is the number of loci, m is the number of parameters estimated, and p is the number of values of χ^2 for which $\Sigma N_h = 0$.

The rest of the analysis follows (10) with map length constrained by the locus farthest from the centromere. In locus-oriented mapping $\theta = S_j$ for successive values of l and

$$S_i = Sh + \delta_{hi}w_{hi} = S_h + \delta_{hi}\,(w_{ci} - w_{ch})$$

$$\delta_{hi} = 1 \text{ if } S_i \geq S_h, \text{ } -1 \text{ else}$$

When intervals w_k are estimated with order fixed, w_{hi} and w_{ci} are incremented for finite differences only if k is included in hi or ci, respectively.

1. Goss, S. J. & Harris, H. (1975) *Nature (London)* **255,** 1445–1458.
2. Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990) *Science* **250,** 245–250.
3. Dear, P. H. & Cook, P. R. (1993) *Nucleic Acids Res.* **21,** 13–20.
4. Lawrence, S., Morton, N. E. & Cox, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7477–7480.
5. Morton, N. E. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1421–1422.
6. Green, E. D. & Olson, M. V. (1990) *Science* **250,** 94–98.
7. Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., *et al.* (1992) *Nature (London)* **359,** 380–387.
8. Matise, T. C., Perlin, M. & Chakravarti, A. (1994) *Nat. Genet.* **6,** 384–390.
9. Shields, D. C., Collins, A., Buetow, K. H. & Morton, N. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6501–6505.
10. Collins, A., Teague, J. & Morton, N. E. (1996) *Genomics*, in press.
11. Lunetta, K. L., Boehnke, M., Large, K. & Cox, D. R. (1995) *Genome Res.* **5,** 151–163.
12. Morton, N. E., Collins, A., Lawrence, S. & Shields, D. C. (1992) *Ann. Hum. Genet.* **56,** 223–232.
13. Francke, U. (1994) *Cytogenet. Cell Genet.* **65,** 206–219.
14. Bray-Ward, P., Menninger, J., Lieman, J., Desai, T., Mokady, N., Banks, A. & Ward, D. C. (1996) *Genomics* **32,** 1–14.
15. Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J. L., Castle, A. B., *et al.* (1995) *Science* **270,** 1945–1954.
16. Holmquist, G. P. (1992) *Am. J. Hum. Genet.* **51,** 17–37.
17. Antonarakis, S. E. (1992) *Genomics* **14,** 1126–1132.
18. Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prad' Homme, J.-F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J. & Goodfellow, P. N. (1995) *Hum. Mol. Genet.* **5,** 339–346.
19. Leach, R. J. & O'Connell, P. (1995) *Adv. Genet.* **33,** 63–99.