Volume 17 Number 17 1989 Nucleic Acids Research

Happy mapping: a proposal for linkage mapping the human genome

P.H.Dear and P.R.Cook*

Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, UK

Received July 18, 1989; Accepted August 3, 1989

ABSTRACT

A theoretical approach for linkage mapping the genome of any higher eukaryote is described. It uses the polymerase chain reaction, oligonucleotides of random sequence and single haploid cells. Markers are defined and then the DNA of a single sperm is broken at random (eg by γ-rays) and physically split into 3 aliquots. Each aliquot is screened for the presence of each marker. Closely-linked markers are more likely to be found in the same aliquot than unlinked markers. The entire process is repeated with further sperm and the frequency that any two markers co-segregate determined. Closely-linked markers co-segregate from most cells; unlinked markers do so rarely. A map can then be constructed from these co-segregation frequencies. A specific application for determining the order and distance between sets of closely-linked and previously-defined markers is also described.

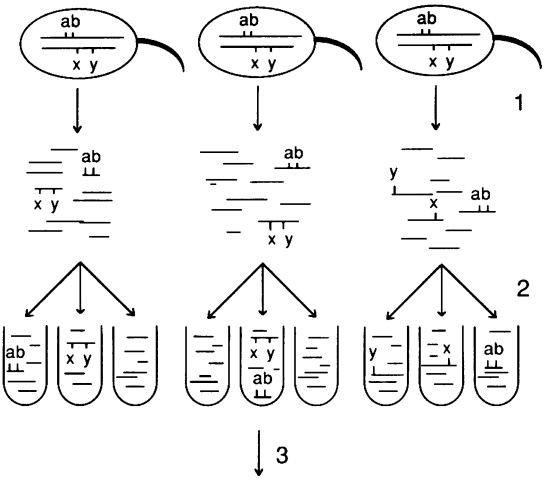
INTRODUCTION

Various strategies are being applied to map the $\sim 3 \times 10^9$ bp of the human genome (1–6). All suffer a number of drawbacks. Random cloning and chromosome 'walking' are impeded by repeated sequences and our inability to clone some loci (7). Difficulties associated with cutting DNA into large fragments (8–10) and then resolving them (11,12) prevent efficient restriction mapping. The **in**frequency of meiotic recombination also limits resolution of classical linkage mapping to, at best, one marker per 10^6 base pairs (6). All these add to the problem of closing the final few gaps in the map (13). Moreover, all strategies are inefficient as maps are generated piecemeal. We suggest an approach that sidesteps these problems; it uses **hap**loid cells and the **p**olymerase chain reaction (14,15), hence 'happy mapping'.

We first exemplify the general approach, which is applicable to genomes of any size, using a specific example in which ~5000 genomic markers are spaced ~600kbp apart. This resolution is approximately twofold better than the limit of conventional linkage analysis, and a twenty-fold improvement on the current map (5). Moreover, 600kbp is comparable to the size of inserts cloned in yeast artificial chromosomes (1,2) and so would form a useful basis for more precise mapping of particular regions. This general approach contains an untried step that has yet to be shown to work in practice. We then discuss a specific application of the general approach—the local mapping of known markers—in which the untried step is eliminated; all steps in this specific application have been shown to be feasible.

THE GENERAL APPROACH

Fig. 1 outlines the approach. First, ~5000 markers scattered throughout the genome are characterised. Then DNA of a **single** sperm is isolated and broken **at random** (eg by γ-



	a	b	x	y
a	-	3	1	1
b	3	-	1	1
x	1	1	-	2
y	1	1	2	-

Fig. 1. The strategy, illustrated using 3 sperm, each containing 2 chromosomes and 4 markers (a,b,x,y). (1) DNA from each sperm is purified, broken randomly and (2) the fragments physically segregated into 3 aliquots. Each aliquot is screened for the presence of each marker and (3) a table constructed showing the number of times each marker is found with each other marker in an aliquot. The values in this table reflect marker proximity. Thus, closely-linked markers a and b are found together three times, loosely-linked markers x and y twice and unlinked markers (eg a and x) once.

irradiation) into fragments with a mean size of 3.5Mbp. These are physically split into 3 aliquots, each containing many fragments. Each aliquot is now screened for the presence of each marker. If two markers are closely linked, they will probably lie on the same DNA fragment and so will both be in the same aliquot (ie, they 'co-segregate'). Unlinked markers will lie on different DNA fragments and therefore have only a 1 in 3 chance of ending up in the same aliquot. We then repeat the entire process with further sperm and determine the number of times that any two markers co-segregate. Closely-linked markers will co-segregate from most of the cells; unlinked markers will do so rarely. A map can then be constructed from these co-segregation frequencies.

The approach is analogous to classical linkage mapping; each sperm is comparable to a complete family, with markers segregating from the 'parent' (a sperm) amongst the 'children' (the aliquots). But our approach is, in theory, far more efficient; we study 5000 markers per 'family' and break the chromosome much more frequently than meiosis so markers segregate on a finer scale.

MARKERS

We first characterize ~5000 unique markers using **bulk** human DNA. This involves

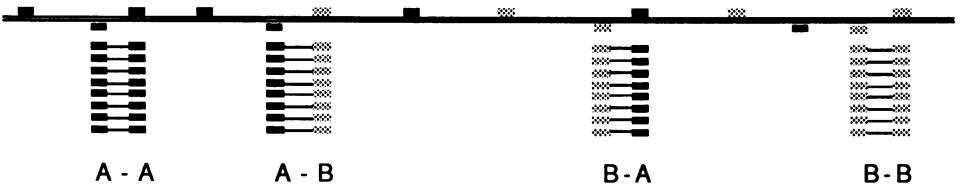


Fig. 2. Chain priming by oligos of random sequence.Two oligos (A, ■ and B, ▒) prime chain synthesis (short lines, flanked by rectangles) on the template (▭) only when their binding sites are sufficiently close together. Half the chains have both A and B at their ends, the other half have either A or B at both ends.

considerable work, but need be done only once. Two oligonucleotides (oligos) of random sequence hybridise at sites scattered randomly throughout the genome; some sites will be <1kbp apart so the intervening sequence can be amplified using the polymerase chain reaction (Fig. 2). [1kbp is a conservative estimate of the maximum length that can be efficiently amplified using the chain reaction.] The result is a number of copies of a number of different sequences. These amplified chains are sized by gel electrophoresis and serve as copies of markers, which are identifiable by length and the particular oligo pair used to produce them.

In most published experiments 'spurious' chains are seen due to non-specific hybridisation so the stringency is increased or longer oligos are used to eliminate them, leaving only the specific chain of interest (14, 15). To us, however, these 'spurious' chains are the chains of interest. We adjust conditions to **generate** as many such chains as can be conveniently resolved on a gel (ie by using shorter oligos or lower stringencies). We assume, conservatively, that 50 chains are resolvable. This requires oligos $9-10$ bases long; unfortunately these prime inefficiently in the chain reaction so longer oligos will probably be used at reduced stringency.

In this way, one oligo pair defines ~ 50 markers and marker number is increased by using more pairs. 20 different oligos (A,B,C...T) are synthesized and used in all possible different pairwise combinations in turn (ie A+B, A+C,...R+T,S+T; ie $[20^2-20]/2) =$ 190 in all) with bulk human DNA. This gives 190 different oligo pairs and so 190 sets of ~ 50 chains, or $\sim 10{,}000$ in all. However, only half of the fifty chains produced per pair have different oligos at their ends (eg A-B and B-A; Fig. 2); half have one or other of the oligos at both ends (eg A-A and B-B). A-A copies appear whenever oligo A is used (ie with A+C, A+D etc), and B-B copies whenever oligo B is used (ie with B+C etc), so only half the chains (ie 5000) are unique copies of markers.

It is essential that each marker be unique; therefore, we ignore chains from different genomic sites that co-migrate on a gel. We excise the track from the gel, incubate it with an enzyme like *Hae*III that cuts frequently and run the products in a second dimension. 'Doublets' yield *Hae*III fragments whose total length is greater than that of the parent band. They, as well as bands that run too closely together to be resolved easily, will be disregarded as marker copies, even though they will all be seen subsequently. Highly-repeated sequences which yield intense bands will also be disregarded; unfortunately, sequences repeated a few times can only be eliminated later.

DETAILS OF THE GENERAL APPROACH

We now perform each of the following steps in turn with different sperm.

Step 1: DNA purification
A single sperm is encapsulated in agarose and its DNA freed of cellular protein and RNA without breaking it, so retaining marker linkage (16).
Step 2: Breakage
DNA is broken at random (eg with γ-rays) into fragments of 3.5Mbp (see later). Smaller fragments improve map resolution, larger fragments reduce the probability of having gaps in the map.
Step 3: Physical segregation
The broken DNA is split equally **without further breakage** into 3 aliquots to segregate markers. Carrier DNA can be added to facilitate this, but if it proves difficult the DNA can be run into a gel using pulsed-fields (11,12) and then the gel cut into 3 pieces. (The chain reaction is unaffected by agarose.) Once DNA has been split, no special precautions need be taken to prevent breakage.
Step 4: Pre-amplification
Just as markers were defined using 190 oligo pairs, each aliquot must now be screened with each pair in turn. But each marker is present in an aliquot in only **one** copy, if at all, so it cannot be subdivided for screening with 190 pairs. Therefore, **all** markers in the aliquot must be 'pre-amplified' to give sufficient copies (ie ~2000) to be sure that each of the 190 sub-aliquots receives at least one. This involves a chain reaction with **all** 20 oligos simultaneously, an untried step that will be discussed later.
Step 5: Screening
The mixture is now split into 190 sub-aliquots and each is screened using the chain reaction with **one** oligo pair. [As pre-amplified chains are now flanked by exact matches of oligo sequences, high stringencies can be used.] Thus, oligos A+B will be used first so that any A-B markers will be amplified to detectable levels. Then the second sub-aliquot is amplified with A+C and so on for each of the 190 pairs. Marker copies are sized and detected on a gel. If amplification is sufficient, bands will be visualised using ethidium; if not, chains can be radio-labelled during amplification and detected by autoradiography. A third attractive alternative is to use fluorescently-tagged oligos so the amplified products can be identified using an automated DNA sequencer (17,18); then up to four samples distinguished by differently-coloured oligos (eg, size-standards and the products of three screening reactions) can be combined in one gel track. Another completely different strategy which will not be discussed further here is to detect labelled chains by hybridization to arrays of immobilised markers.
Step 6: Analysis
Steps 1−5 are repeated with each of the sperm. Each marker is found with other markers in an aliquot from any cell, but only closely-linked markers are consistently associated in aliquots from successive cells. The number of times markers are found together gives a 'linkage value' which reflects their proximity. Tightly-linked markers have a linkage value close to the number of cells analyzed. Of course, **unlinked** markers have a 1 in 3 chance of ending up in the same aliquot from any cell, giving a background of 'spurious co-segregation' and false linkage. Intermediate values between these two extremes reflect the proximity of truly-linked markers.

The map is constructed in a similar way to that used classically but the problems are simpler. First, linkage is not influenced by sequence or sex-dependent differences in recombination rates. Second, linkage is detected only between closely-linked markers, so

Table 1. The relationship between marker spacing, the number of oligos and the amount of work. The work required to analyze 100 cells split into 3 aliquots is expressed in terms of gels to be run, assuming 3 screening reactions are combined and run with a marker in one track on a 16 track gel in an automatic sequencer.

Marker spacing (Mbp)	Number of oligos required	Number of oligo pairs	Work (no. of gels)
15	4	6	38
10	5	10	63
3.75	8	28	175
1	15	105	657
0.6	20	190	1188
0.23	32	496	3100

only linkages between each marker and some 15–20 others (about 0.1 million linkages) need be considered, rather than all linkages between each marker and all others (12.5 million linkages). This simplifies the initial rough ordering and subsequent 'fine-tuning' by least-squares analysis.

PROBLEMS

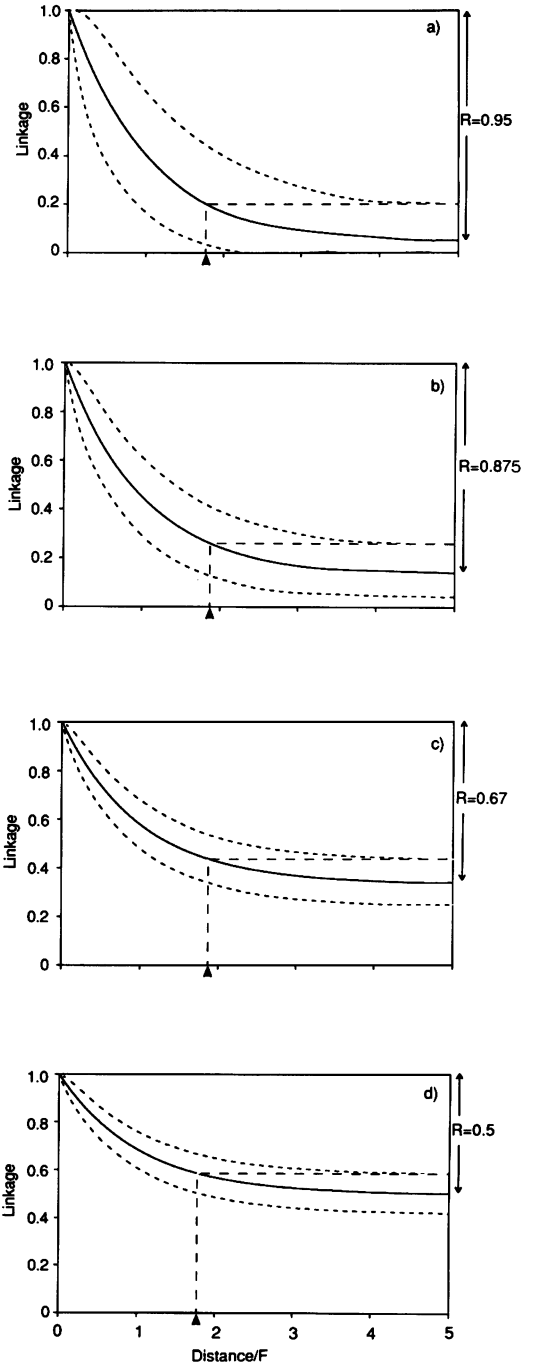
Repeats.

Some markers within repeated sequences might have escaped detection when markers were defined. They will be found in more than one aliquot from each sperm and so can be identified (as can contaminating human DNA) and disregarded. Closely-spaced markers in tandem repeats (eg L-QQ-M) will consistently co-segregate and hence appear as a single marker, locally telescoping the map (L-Q-M). Thus, in contrast to many other strategies, this one is not defeated by repeats.

Pre-amplification.

We require that all markers in each aliquot be 'pre-amplified' with 20 oligos **simultaneously**, an untried step. Combinations of many more oligos have been shown to work (19–21), but the oligos were longer and the conditions more stringent than we must use. The maximum number of oligos that can be used together, their sequences, length and amplification conditions remain to be established and constitute the bulk of any feasibility study. Fortunately, at this stage we need to amplify only about two thousand-fold (compared to the billion-fold or more usually achieved with the chain reaction), so inefficiency should not be too serious a problem. Even if we assume the worst, that only 4 oligos (A,B,C,D) can be used during pre-amplification—as has been done successfully with a single sperm (15)—this allows 6 pairwise combinations to be used to give 225 markers and a 15Mbp map. 5 oligos give a 10 Mbp map, equivalent to the best currently available (5), whilst 15 give a 1Mbp map which is probably unattainable by classical methods. Table 1 illustrates the relationship between the number of oligos used and marker spacing.

Even if we wish to map more markers than the pre-amplification step permits, we can construct 'parallel' maps which can be aligned because they share some markers. The penalty is either increased work (if many markers are common to all maps) or imprecise alignment of the parallel maps (if few markers are common). Any number of groups can


a)
R=0.95
b)
R=0.875
c)
R=0.67
d)
R=0.5
Linkage
1.0
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
Distance/F

construct such maps, increasing marker density indefinitely. For example, if only 10 oligos can be used at once, a master set of 5 is selected and used in turn with many other sets of 5 to generate 'parallel' maps sharing a quarter of their markers.

Random marker loss.

We expect some markers to be lost, if only because we start with single DNA molecules that might adhere to tube walls; Li *et al.* (15) 'lost' ~20% of their 'markers' when using 4 oligos with a single sperm. If random, this loss reduces the linkage values of all marker pairs and is remedied by analyzing more cells. If a marker systematically fails to amplify, it is lost but the remainder of the map is not distorted. Using known frequencies, we calculate that <1% markers should be lost due to mutations (22) and γ-ray-induced breaks (23).

Marker loss due to polymorphisms.

Polymorphisms will cause significant losses; most obviously, markers from only one sex chromosome will be present in a sperm and will be mapped less accurately than autosomal markers. Additionally, we calculate ~6% of markers should be lost from each sperm because of autosomal polymorphisms in the oligo-binding sites (24,25). Most polymorphisms are base-substitutions and not insertions or deletions (24,25) so few losses should result from polymorphisms in marker length. The 8% of sperm that are aneuploid (26,27) will cause further losses if a chromosome is missing or if a sperm with an extra chromosome is eliminated when markers turn up in more than one aliquot. All these losses cause extra work but, again, they do not defeat us.

Although aneuploid cells cause such problems, they also provide additional data. Absence of a particular chromosome will cause the simultaneous loss of all markers on that chromosome and an additional chromosome causes their simultaneous duplication. Although this data is not 'clean' (marker loss or duplication can also occur through failure to amplify or contamination, for instance) it can nevertheless be tapped. For example, if there are a few gaps in the final map, large 'contigs' may be assigned to common chromosomes by noting that the markers they contain all tend to be lost or duplicated **together**.

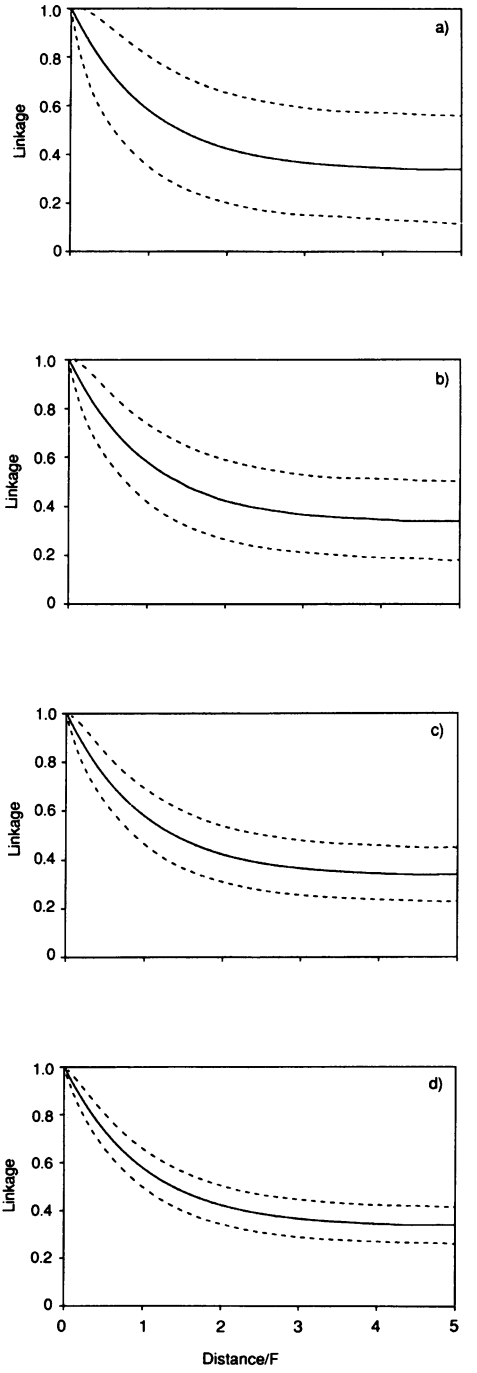
MAP QUALITY

The number of times markers co-segregate gives a 'linkage value' which reflects their proximity. Tightly-linked markers will have a linkage value close to the total number of cells analyzed. **Unlinked** markers have a 1 in 3 chance of ending up in the same aliquot from any cell, giving a background of 'spurious co-segregation' and false linkage. We now show that division into three aliquots maximises the signal due to true linkage above this background of spurious linkage.

The signal:noise ratio

Consider C cells (the number of cells analyzed), each divided into A aliquots. Tightly-linked markers have a linkage value approaching C, which we normalize to 1. Conversely, unlinked markers have a linkage value of 1/A, due to spurious co-segregation. Thus, the

Fig. 3. Linkage plots for a constant amount of work (C×A=400). A computer program was written which calculates the expected linkage values (——) between markers as a function of the distance between them, and the 99% confidence limits (--------) for these values. Distances are expressed relative to the mean fragment size, F. The program assumes that γ-ray-induced breaks occur randomly. (a) C=20, A=20. (b) C=50, A=8. (c) C=133, A=3. (d) C=200, A=2. The working range, R, and critical distance (▲) are indicated in each case.


a)
b)
c)
d)
Linkage
1.0
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
Distance/F

'working range', R (the difference in linkage between unlinked and tightly-linked markers), is given by

$$R = 1 - 1/A \qquad \text{(eqn.1)}$$

so it might seem desirable to maximise the number of aliquots. However, observed linkage values fluctuate statistically much like meiotic segregation frequencies, introducing 'noise'. The standard deviation, s, of the measured linkage values will approximately be inversely proportional to the square root of the number of cells analyzed, ie

$$s = k_1/\sqrt{C} \qquad \text{(eqn.2).}$$

where k_1 is a constant. The total amount of work, W, required is proportional to the product of C and A, ie

$$W = k_2AC \qquad \text{(eqn. 3)}$$

where k_2 is a constant. Substituting for C in eqn. 2 yields

$$s = k_1/\sqrt{(W/k_2A)} \qquad \text{(eqn. 4).}$$

Obviously, the working range of linkage values should be as large as possible relative to the error in the observed linkages. This 'signal-to-noise' ratio, R/s, is given by

$$\begin{aligned} R/s &= (1-1/A)/(k_1/\sqrt{(W/k_2A)}) \\ &= (1-1/A).\sqrt{(1/A)}.K \end{aligned} \qquad \text{(eqn.5)}$$

where $K = \sqrt{(W/(k_1^2k_2))}$. A = 3 gives the maximal ratio. In other words, the most effective strategy is to divide the DNA into three aliquots. If divided into more aliquots, the number of cells which can be analyzed (for a given amount of work) decreases, and statistical errors in linkage values become greater. If split into fewer aliquots (and more cells analyzed), then the increased background of spurious co-segregation reduces the working range of linkage values, offsetting the greater accuracy afforded by analyzing more cells.

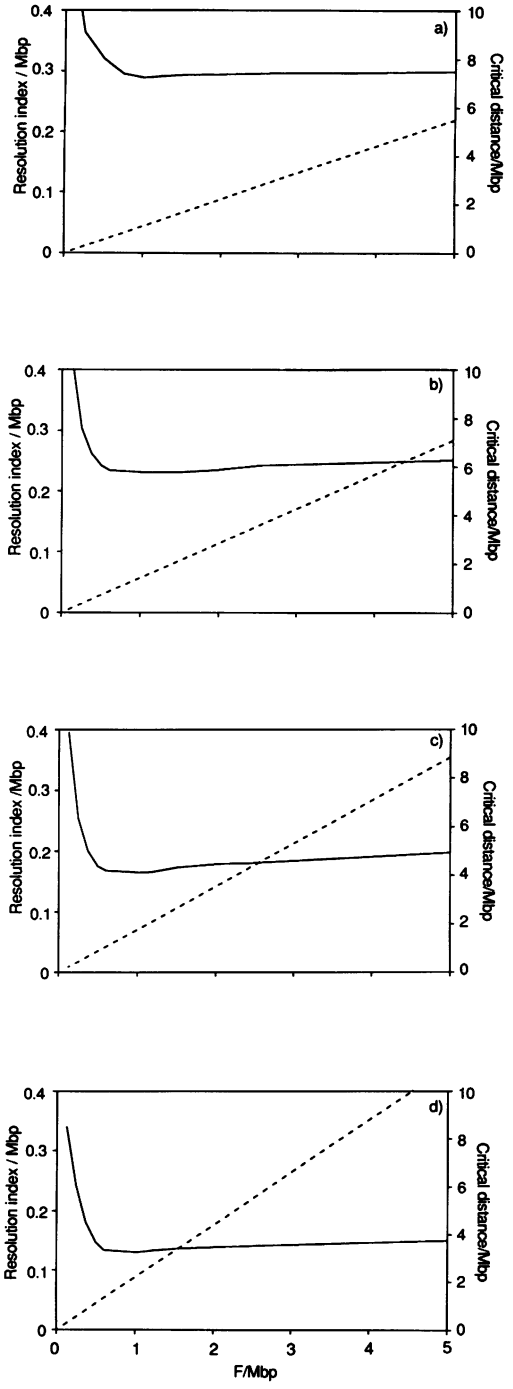
This is shown graphically in Fig. 3, where linkage values have been calculated and plotted as a function of distance between markers. Marker distance is expressed relative to F, the mean fragment size. For a given amount of work (ie $C \times A = 400$), as the number of cells increases and the number of aliquots decreases, the 'noise' of spurious co-segregation (and hence the apparent linkage measured between unlinked markers) rises, reducing the working range. At the same time, the confidence limits become narrower, as analyzing more cells reduces statistical errors. Inspection of Fig. 3 confirms that, at intermediate distances (those reliably measured), the confidence limits are narrowest relative to R, and hence the signal-to-noise ratio best, for A=3.

Equation 5 also predicts that the signal-to-noise ratio should improve as $\sqrt{W}$, and hence as $\sqrt{C}$, if A is 3. Calculation also shows this to be true; in Fig. 4 the width of the confidence limits at a given distance decreases approximately as $1/\sqrt{C}$.

Resolution and gaps

Two indices of map quality may be derived. One is the 'resolution index' (the error in

Fig. 4. Linkage plots for increasing numbers of cells calculated as for Fig. 3 (A=3 in all cases). (a) C=25, (b) C=50, (c) C=133 and (d) C=200. The width of the confidence limits at any given distance varies approximately as $1/\sqrt{C}$.


a)
b)
c)
d)
Resolution index / Mbp
Critical distance/Mbp
0
0.1
0.2
0.3
0.4
2
4
6
8
10
1
3
5
F/Mbp

placing markers at the 'typical' spacing of 600kb), the other the 'critical distance' which can be tolerated between them before linkage is lost and gaps appear in the map (Fig. 3a). F, the mean fragment size, has a complex effect on both these indices. If F is small, resolution is lost because each marker is linked to few others and the critical distance becomes so small that many gaps result. If F is large, resolution again decreases even though each marker is placed by reference to many others, because each distance is measured so inaccurately. In the extreme case where F is as large as a chromosome, any two markers on a chromosome have a linkage value of 1.0 between them, with no resolution. Effects of cell number are superimposed upon effects of F and optimisation depends on whether resolution or contiguity is emphasised (Fig. 5). However, with 100 cells, 3 aliquots and a mean fragment size of 3.5Mbp, markers spaced ~600kbp apart are placed with an error of ± ~190kbp (99% confidence level), whilst linkage is preserved across inter-marker gaps of up to 6Mbp (Fig. 5c). If markers are scattered randomly, this should leave no gaps **anywhere** in the map of the genome.

Even if gaps do appear in our map, perhaps because we fortuitously have no markers in long blocks of repeated sequences, we **can** nevertheless close them by analyzing extra cells with DNA broken into larger fragments.

MAPPING A SINGLE CHROMOSOME

More limited approaches are also possible, eg mapping single flow-sorted mitotic chromosomes. Their DNA is sufficiently intact to retain marker linkage (J. Fantes and PRC, unpublished) but the **two** copies of each marker present per cell complicate analysis. Each segregates independently, doubling linkage values and quadrupling the background of false linkage. However, the smaller scale means that fewer markers need be defined, allowing fewer oligos to be used, which may be become important if we are limited by the number of oligos that can be used during pre-amplification. For example, 4 oligos give 225 markers, which would be spaced every 444kb on a 100Mbp chromosome. The oligos would, of course, have to be slightly shorter (ie 8−9 bases long).

MAPPING KNOWN LOCI

A simplified strategy can be applied to map, at almost any desired resolution, small numbers of markers. This entails no untried steps. Consider two loci, α and β, separated by an unknown distance and for which sequences are available. Pairs of oligos specific for each locus are synthesised (ie $\alpha 1$, $\alpha 2$, and $\beta 1$, $\beta 2$) which yield locus-specific chains of known size. DNA from a single sperm is divided into 3 aliquots, each aliquot 'pre-amplified' **at high stringency** with all four oligos simultaneously, split into two sub-aliquots, each amplified with one of the two oligo pairs and the chains sized electrophoretically. After analyzing a number of cells, the linkage value for the two loci, and hence map distance, is calculated as before. By measuring a number of inter-locus spacings in this way, a local map may be constructed.

This approach is analogous to that suggested by Li *et al.* (15) who successfully amplified two loci simultaneously from single human sperm and then used allele-specific probes to identify which of two alleles was present at each of the two loci. They propose to go on

Fig.5. Resolution index (——) and critical distance (------) plotted as functions of mean fragment length, F, for (a) C=25, (b) C=50, (c) C=100 and (d) C=200. A=3 in all cases.

to linkage map using meiotic recombination to segregate markers. Replacing meiotic breakage and segregation by physical procedures—the **only** technical difference in our approach—affords considerable advantages. First, any locus for which some sequence information is available may be mapped: there is no need for polymorphisms in 'informative families'. Second, our strategy is not limited by the infrequency of meiotic recombination; by varying the break frequency, markers separated by only a few tens of kilobases, or a few megabases, can be mapped with equal ease.

Errors in measuring inter-marker distances can be read directly from linkage plots such as those in Figs. 3 and 4 where positive and negative errors are the distances (measured horizontally) from the mean linkage value to the upper and lower confidence bounds, respectively. For example, with 100 cells, $A=3$ and $F \approx 2$Mbp, the distance between two markers can be shown to be overestimated by 60% or underestimated by 50% (99% confidence), if the markers are between 0.5 and 2Mbp apart. At the 90% confidence level, the error is approximately +30%/−25%, adequate for determining marker order. The same accuracy can be attained for markers between 50−200kbp apart, by reducing F to ~0.2Mbp (in which case DNA can be broken mechanically). Such experiments entail 600 gel tracks (100 cells×3 aliquots×2 oligo pairs), or about 12 gels with ~50 tracks plus markers per gel.

Even less work can still provide useful information on whether markers are tightly-linked. If only 25 cells are analyzed (3 gels!), errors in measuring inter-marker spacing become very large but linkage can still be detected over distances less than the critical distance of 1.1F (Fig. 4a). As F is under control, we can set the critical distance (in the range from a few kilobases to several megabases) and determine whether one marker is within that distance from another. This means, for example, we can easily determine whether a marker is within range of a chromosome 'walk'.

CONCLUSIONS

The approach has a number of advantages, circumventing the drawbacks of other strategies. It is general and efficient. The 600kb map discussed needs 190 chain reactions and gel tracks to define markers, plus ~57000 reactions and tracks during screening (100 cells×3 aliquots×190 oligo pairs), equivalent to ~1200 gels (Table 1). [It is only prudent to double this to allow for marker loss.] A 10Mbp map, equivalent to the best currently available, requires ~126 gels (Table 1). Most importantly, we control the breakage frequency. This allows us to select a compromise between maximum resolution and the minimum number of gaps in the map. If the 5000 markers are distributed randomly, we can map each to within +/−190kbp; even if some are distributed so non-randomly that we are left with gaps in the map, analysis of larger fragments allows these to be closed.

We suggest that the general approach warrants the effort to make it work. We also suggest that the limited approach is immediately useful for mapping existing markers in local regions as it requires only a few hundred chain reactions and several gels.

ACKNOWLEDGEMENTS

We thank Dr. J. Fantes, Dr. J.S. Rollet, Mr. D.C. Handscomb, Dr. B. Mullen and Prof. R.M. May for their help. This research was supported by the Cancer Research Campaign and the Leukaemia Research Fund.

*To whom correspondence should be addressed

REFERENCES

1. Olson, M.V. *et al.*, (1986). Proc. Natl. Acad. Sci. USA *83*, 7826–7830.
2. Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986). Proc. Natl. Acad. Sci. USA *83*, 7821–7825.
3. Poustka, A. *et al.*, (1986). Cold Spring Harbour Symp. quant. Biol. *51*, 131–139.
4. Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Am J. Hum. Genet. *32*, 314–331.
5. Donis-Keller H. *et al.* (1987). Cell *51*, 319–337.
6. White, R. and Lalouel, J-M. (1988). Ann. Rev. Genet. *22*, 259–279.
7. Wyman, A.R., Wolfe, L.B. and Botstein, D. (1985). Proc Natl. Acad. Sci. USA *82*, 2880–2884.
8. Drmanac, R., Petrovic, N., Glisin, V. and Crkvenjakov, R. (1986). Nuc. Acids Res. *14*, 4691–4692.
9. Brown, W.R.A. and Bird, A.P. (1986). Nature, London, *322*, 477–481.
10. Smith, L.M. *et al.*, (1986). Nature *321*, 674–679.
11. Schwartz, D.C. and Cantor, C.R. (1984) Cell *37*, 67–75.
12. Carle, G.F. and Olson, M.V. (1984). Nuc. Acids Res. *12*, 5647–5664.
13. Coulson, A., Waterson, R., Kiff, J., Sulston, J. and Kohara, Y. (1988). Nature *335*, 184–186.
14. Saiki, R.K. *et al.*, (1988). Science *239*, 487–491.
15. Li, H. *et al.*, (1988). Nature *335*, 414–417.
16. Cook, P.R. (1984). EMBO J. *3*, 1837–1842.
17. Landegren, U., Kaiser, R., Caskey, C.T. and Hood, L. (1988). Science *242*, 229–237.
18. Carrano, A.V. *et al.*, (1989). Genomics *4*, 129–136.
19. Lee, C.W., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. and Caskey, C.T. (1988). Science *239*, 1288–1291.
20. Girgis, S.I., Alevizaki, M., Denny, P., Ferrier, G.J.M. and Legon, S. (1988). Nuc. Acids Res. *16*, 10371.
21. De, R.K. and Srinivasan, A. (1989). Nuc. Acids Res. *17*, 2142.
22. Searle, A.G. (1974). Adv. Radiat. Biol. *4*, 131–207.
23. Friefelder, D. and Trumbo, B. (1969). Biopolymers *7*, 681–693.
24. Jeffries, A. (1979). Cell *10*, 1–10.
25. Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S. and Schmidtke, J. (1985). Hum. Genet. *69*, 201–205.
26. Martin, R.H., Lin, C.C., Balkan, W. and Burm, K. (1982). Am. J. hum. Genet. *34*, 459–468.
27. Edwards, R.G. (1986). Human Reproduction *1*, 185–198.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.