Volume 15 Number 4 1987 Nucleic Acids Research

Progress towards construction of a total restriction fragment map of a human chromosome

Henrik Vissing[1,2], Frank Grosveld[3], Ellen Solomon[4], Graham Moore[4], Nick Lench[1], Nigel Shennan[5] and Robert Williamson[1]

[1]Department of Biochemistry, St. Mary's Hospital Medical School, University of London, London W2 1PG, UK, [2]Hagedorn Research Laboratory, Niels Steensensvej 6, DK-2820 Gentofte, Denmark, [3]Division of Molecular Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7, [4]Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2, and [5]Metabolic Unit, St. Mary's Hospital, London W2, UK

Received January 6, 1987; Accepted January 21, 1987

ABSTRACT

We present an approach to the construction of an overlapping restriction fragment map of a single human chromosome. A genomic cosmid library genome was constructed from a mouse-human hybrid cell line containing chromosome 17 as its only human genetic component. Cosmids containing human inserts were isolated by hybridisation to a human Alu sequence. DNAs from ninety-six randomly chosen cosmids were digested with either EcoRI or HindIII, end-labelled with ^{35}S-dATP and analysed using agarose gel electrophoresis. Comparison of the restriction fragment patterns revealed two pairs of overlapping clones, that were confirmed by cross-hybridisation of the overlapping fragments. The two pairs of cosmids both mapped to human chromosome 17, as shown by hybridisation to a panel of somatic cell hybrids. These data demonstrate that the generation of an overlapping cosmid map along a human chromosome is feasible, representing an intermediate step towards the complete sequencing of a human chromosome.

INTRODUCTION

A complete map of the human genome will soon be available, which will allow the approximate localisation of many phenotypes, including inherited diseases for which the biochemical defect is unknown (1). However, fine mapping of the organisation of coding genes, single copy and repetitive sequences, cross-over regions and centromeric and telomeric structures requires a complete structural analysis of the DNA sequence along the entire length of the chromosome. By combining existing techniques for mapping, landmarking, walking and sequencing, this should now be possible (2,3).

The detailed analysis of the genome of any species proceeds through a series of defined stages (4). Random

markers, often polymorphic proteins with specific functions, are first identified and located in linkage groups, or to specific chromosomes, either by fortuitous co-inheritance or by their occurrence on marker chromosomes or episomes. These markers are then supplemented by others, such as cloned DNA sequences, which are polymorphic and (when added to the previous phenotypic assignments) give a linkage map for a chromosome. However, the linked, mapped markers on a chromosome will still be separated by several million base pairs, since a recombination value of one centimorgan (1%) has been shown experimentally to correspond to approximately one million base pairs (5).

Defining the human genome at this level is of importance in understanding the inheritance of human diseases. Several groups have isolated random DNA markers which are polymorphic and can be used in segregation and linkage studies. Such anonymous, and presumably neutral, markers are being used to construct a total human linkage map (6,7).

A linkage map for the human X chromosome has been established using DNA fragments linked both to each other (8), and to diseases for which the biochemical defect is not known such as Duchenne muscular dystrophy (1). In view of the low resolution of family-based human linkage analysis, a feasability study on preparing a set of overlapping human chromosome-specific fragments to construct a complete and definitive map of a chromosome was undertaken.

We have chosen human chromosome 17 as the model to test such a strategy. Chromosome 17 is the smallest that can be obtained easily in a hybrid cell where it is the only human chromosome in a mouse background, by selection for the thymidine kinase (tk) gene in a tk^- mouse cell. Chromosome 17 is 2.8% of the human DNA content (9), and therefore it comprises 1.4% of the hybrid cell DNA, as it is present in only a single copy in the murine diploid cell. Its length could be contained in approximately 2100 cosmids, end to end, each 40kb in length. Cosmids containing human DNA inserts can be identified by hybridisation to a human Alu repetitive DNA sequence. The human Alu sequence occurs in 300,000

copies per haploid human genome (10), and assuming its distribution is random, each cosmid insert of 40kb will contain approximately four Alu repeats.

Considering chromosome 17 as 2100 contiguous cosmids, and with a definition of 25% commonality between the total sequence length of any two cosmids in order to be identified as overlapping, the probability of finding overlaps can be calculated using a Monte Carlo statistical analysis (11). The probability of finding one pair of overlapping cosmids exceeds 50% when 47 cosmids have been analysed; in order to find two overlaps 71 cosmids would have to be analysed. To identify overlaps, we have chosen agarose gel electorphoresis to analyse two parallel six-base sequence restriction enzyme digests for each cosmid. This approach has features in common with that used by Coulson et al. (12) for the nematode genome, and also provides a partial restriction map for each cosmid.

Our results indicate that, using this technique, it is possible to generate ovelaps for a human chromosome by analysis of a relatively small number of clones.

MATERIALS AND METHODS

Chemicals. All enzymes were from Boehringer Mannheim. Restriction enzyme digests were carried out in 50mM tris (pH 7.4), 10mM $MgCl_2$, 100mM NaCl, 7mM ß-mercaptoethancl, 5mM spermidine, 100 µg/ml bovine serum albumin. ^{32}P-dCTP (30TBq/mmol) was from Amersham International Ltd.; ^{35}S-dATP (18.5TBq/mmol) was from NEN Du Pont.

Cell Lines. The PCTBA1.8 cell line (13), a mouse-human cell hybrid containing human chromosome 17 as its only human genetic component, was grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 1mM hypoxanthine, 0.3µM aminopterin and 1.6mM thymidine. The cells were karyotyped immediately before harvesting for DNA isolation.

Cosmid Library. DNA was prepared from PCTBA1.8, partially digested with Sau3A and fractionated on a 1.25-5M NaCl gradient in 10mM tris (pH 7.4), 1mM EDTA (14). The 30-50kb fraction was isolated and ligated to pLTC cosmid arms. The

pLTC vector is a variant of the cosmid vector pTCF (15) in which the HpaI site has been replaced by an XhoI site. It accepts Sau3A partial digestion fragments in a BamHI site. Only inserts between approximately 35 and 45kb are accepted for packaging. Ligation and packaging were carried out as described in Grosveld et al. (15).

Screening. The cosmids were used to transfect and were plated out in E. coli ED8767 on LB plates containing 50 ug/ml ampicillin and 3 µg/ml kanamycin. Due to the presence of the neomycin gene in the vector, kanamycin is tolerated at a concentration of up to 3ug/ml by bacterial colonies containing cosmids, which enhances cosmid stability and gives an increased yield for DNA preparations. The library was plated on 20cm x 20cm Hybond^N filters (Amersham International Ltd.), at approximately 4000 colonies per filter.

Two replicas of each plate were taken and prepared for hybridisation using the protocol provided by the manufacturer. The filters were pre-hybridised in 4xSET (1xSET = 0.15M NaCl, 0.03M tris, 2mM EDTA, pH 7.4), 0.1% sodium pyrophosphate, 0.2% SDS and 50µg/ml heparin at 65°. An oligo-labelled Alu human repeat sequence removed from a low melting 1.5% agarose gel of a BamHI digest of the Blur2 plasmid (16) was then added at an activity of $5x10^5$ cpm/ml and hybridised overnight at 65°. The filters were washed at a stringency of 0.1xSSC at 65° and autoradiographed overnight. Positive colonies were picked, arranged in a grid, and rescreened. Second time positive colonies were numbered and used as a human chromosome 17-specific cosmid partial bank.

Cosmid Restriction Fragment Analysis. Cosmid DNA was prepared by the method of Grosveld et al. (14). Approximately 100ng DNA was digested with either HindIII or EcoRI followed by a direct fill of the first residue of the four base overlap using 1 unit of Klenow fragment of DNA polymerase I with 1µCi ^{35}S-dATP. One-fifth of each digest was run on a 20cm x 20cm 0.7% agarose gel in tris-acetate buffer. Each gel was then dried onto 3MM paper and autoradiographed

overnight, using EcoRI or EcoRI/HindIII digests of lambda DNA as markers.

The fragment sizes of the cosmid bands were calculated using a GP-7 Graf/Bar digitiser (Sciences Accessories Corporation, Connecticut) and the Digicalc software program (Armand MacMurray, Massachusetts) with an IBM XT computer. The fragment patterns of the cosmids were compared using the MATCH database program (Niels Friis Nyholm, Copenhagen).

Southern Blotting. DNA was blotted onto Hybond^N filters using the method given by the manufacturer. Filters were hybridised using the conditions described above for cosmid screening. For cross-hybridisation and chromosomal mapping experiments, specific fragments were removed from low temperature melting agarose gels and oligo-labelled with ^{32}P-dCTP (17).

RESULTS

The cosmid library prepared from the cell line PCTBA1.8 consisted of approximately 10^6 recombinants. Approximately $4x10^4$ recombinants were plated and screened for cosmids containing human Alu repeat sequences. The screening gave a strong hybridisation signal from approximately 1% of the recombinants, and these were picked and re-screened. Ninety-six cosmids containing human Alu sequences after two rounds of screening were chosen at random, and DNA prepared. The DNA was digested with EcoRI and HindIII, end-labelled and analysed by agarose gel electrophoresis (Fig. 1). The migration of each fragment was determined with the digitiser, and sizes calculated using the Digicalc program. The average insert size was found to be 42kb. The average number of EcoRI fragments per cosmid was eight (not including the two fragments derived from the vector); the average number of HindIII fragments was nine.

The MATCH program made it possible to determine the number of bands which are in common between any two cosmids digested with a specific restriction enzyme. The program has an optional percent variation of fragment size, and a variable minimum number of identical bands to score any two

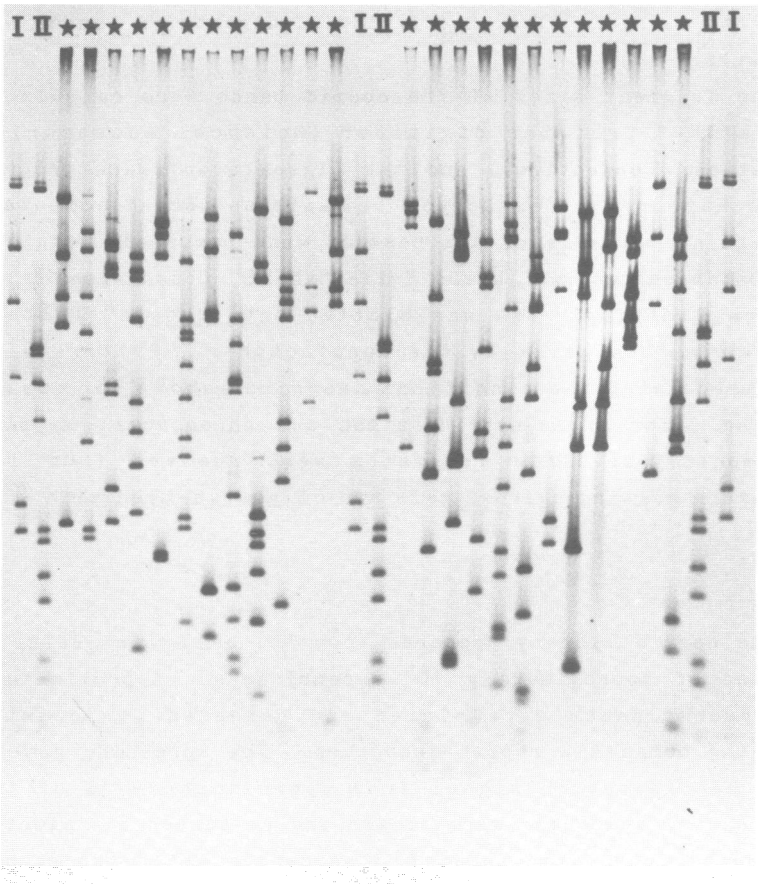

Fig. 1
Gel electrophoresis of human chromosome 17 cosmids. Cosmids were digested in this experiment with HindIII, radioactively labelled and analysed on 0.7% agarose gels. Markers are HindIII (I) and EcoRI/HindIII (II) digests of lambda DNA.

cosmids as overlapping. Comparing 96 cosmids for two or more identical bands (with 2% variation in fragment size), for two parallel digests using EcoRI and HindIII, gave two pairs of cosmids with possible overlaps.

The two cosmid pairs, cos16,24 and cos63,67, were digested with EcoRI, EcoRI/SalI and with HindIII, HindIII/SalI, in order to identify the end fragments, since

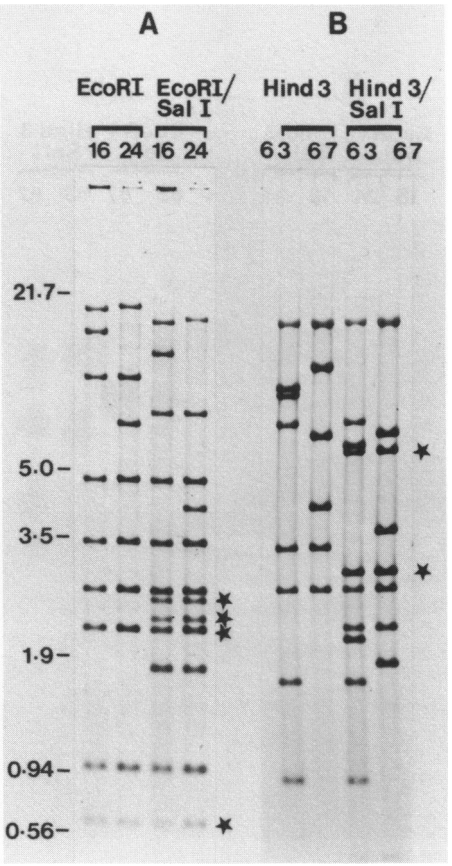

Fig. 2A,B
A) Cosmid clones cos16, 24 digested with EcoRI/SalI.
B) Cosmid clones cos63, 67 digested with HindIII and HindIII/SalI. The fragments were end-labelled, analysed on 0.7% agarose gels, and autoradiographed. (*) indicates vector-derived fragments.

the pLTC vector has SalI sites adjacent to the BamHI cloning site. Cos16 has an insert of 46kb and cos 24 of 42kb, and the two cosmids were found to share 10 of 12 bands in the EcoRI/SalI double digest (Fig. 2A). Cos63 has an insert of 39kb and cos67 of 32kb, sharing 5 of 11 bands in a HindIII/SalI double digest (Fig. 2B).

The common bands obtained from cos16 and cos 24 also

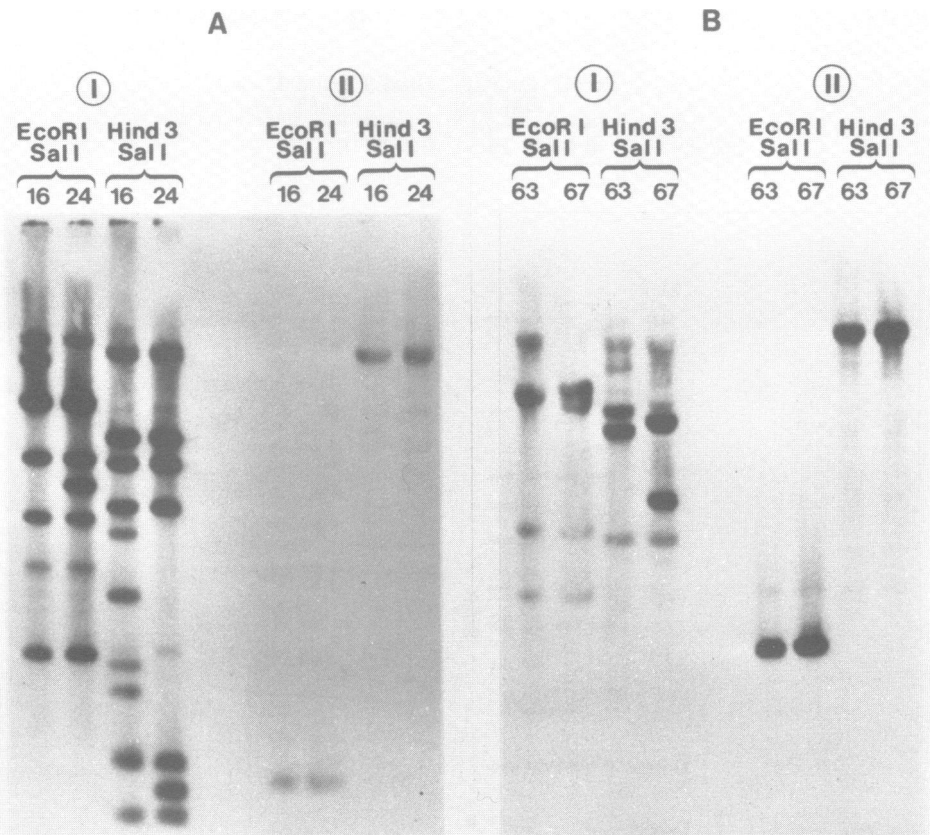

Fig. 3A,B
A) Cosmids 16,24 (A) and cos63,67 (B) digested with EcoRI/SalI and HindIII/SalI, and blotted onto Hybond^{N} filters. (I) Hybridised with human Alu repeat sequence. (II) Hybridised with (A) cos16 1.0kb EcoRI fragment. (B) cos63 1.7kb EcoRI fragment.

share a similar distribution of repetitive sequences (Fig. 3A). To confirm the overlap definitively, the cos16 EcoRI band migrating at 1.0kb was excised, labelled and hybridised to both cos16 and cos24 digests. Cross-hybridisation was seen to the same bands from both cosmids, showing that the overlap is real. In a similar fashion, the 1.7kb EcoRI band from cos63 showed cross-hybridisation to the same bands from cos67 (Fig. 3B). The two pairs of clones cos16,24 and

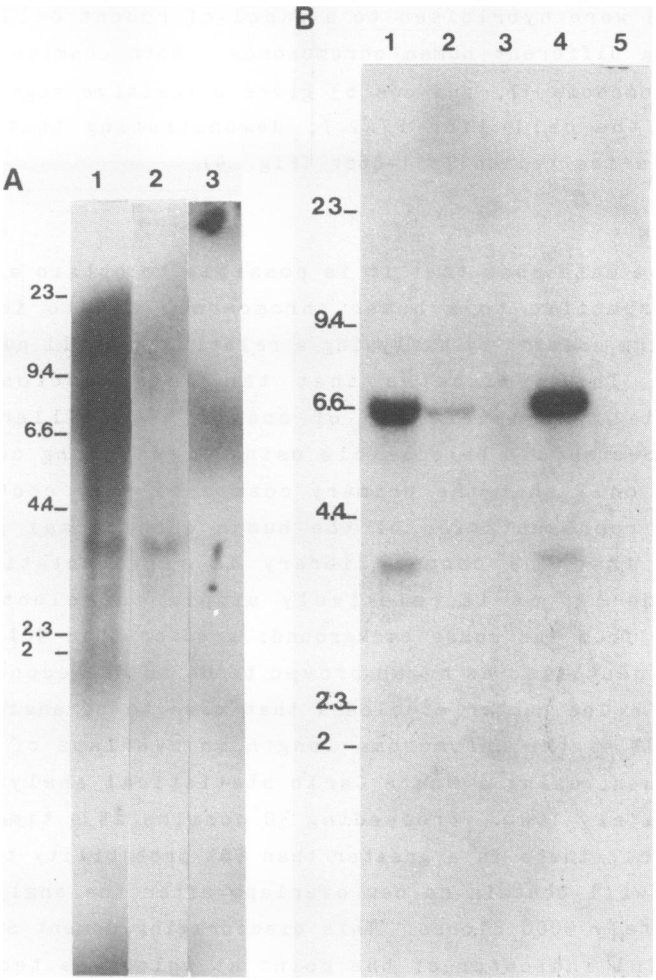

Fig. 4

Chromosomal location of human inserts from cos16 and cos63. (A) 1.7kb EcoRI fragment from cos16 hybridised to DNAs each digested with EcoRI from: lane 1, Molt4 (total human); lane 2, PCTBA1.8 (17 only); lane 3, 1R (total mouse). (B) 1.7kb EcoRI fragment from cos63 hybridised to DNAs each digested with HindIII from: lane 1, PCTBA1.8 (17 only); lane 2, P7A2.7 (17p11-qter only); lane 3, Hor1I (15, 11q, Xp); lane , Molt4 (total human); lane 5, 1R (total mouse).

cos63,67 have overlaps of approximately 80% and 60% respectively.

In order to confirm that the clones obtained are specific to human chromosome 17, unique fragments from cos16

and cos63 were hybridised to a panel of rodent cell lines containing different human chromosomes. Both cosmids map to human chromosome 17, and cos 63 gives a positive signal with DNA from the cell line P7A2.7, demonstrating that it is located in the region17p11-qter (Fig. 4).

DISCUSSION

These data show that it is possible to obtain a cosmid library specific to a human chromosome, and to identify overlapping cosmids by analysing a relatively small number of clones. This indicates that the construction of a substantially complete map of one of the smaller human chromosomes should be possible using overlapping cosmids, provided only that the primary cosmid library contains a complete representation of the human chromosomal genome. Assuming that the cosmid library is representative for chromosome 17, it is relatively simple to select human sequences from the mouse background; greater than 90% of our cosmids identified as human proved to be so on second round screening. The number of clones that need to be analysed to obtain 50% of the chromosome length in overlaps of two or more clones, using a Monte Carlo statistical analysis, is approximately 1600. Proceeding 30 cosmids at a time on an agarose gel, there is a greater than 50% probability that the next gel will contain no new overlaps after the analysis of approximately 5000 clones. This discouraging moment seems to be a useful indicator of the point at which the technique ceases to be rewarding.

The number of gaps remaining at this point will be determined by a number of factors: statistical chance, any stretches greater than approximately 40kb which contain no Alu sequences, and any sequences which are not clonable in the vector/host system used. Several groups have found that certain human DNA sequences are difficult to clone in lambda or cosmid libraries (Grosveld, personal communication). This may be because they contain sequences which interfere with host or vector viability, or because they are bounded by particularly sensitive Sau3A cleavage sites. However, this

cannot be a frequent problem, as screening cosmid libraries with a plasmid or phage recombinant probe usually yields a corresponding sequence at approximately the frequency expected. Any such gaps in the overlap map would have to be bridged using other techniques, such as chromosomal walking using directional walking vectors (18).

Approaches which simplify mapping include any technique which allows the alignment of overlapping cosmid sets, including reverse and orthogonal pulse field gel electrophoresis (19,20), and jumping libraries which could link DNA sequences separated by several hundred kilobases (2). In addition, anonymous markers, as well as coding genes such as thymidine kinase, myosin and human growth hormone, which are located on chromosome 17 (4), can be used as landmarks to orientate contiguous overlapping cosmids (contigs). Finally, direct segregation analysis for probe ordering is possible using the family resources of the Centre d'Etude Polymorphisme Humain.

We conclude that construction of a complete physical overlap map for a single human chromosome is feasible, and can be done in a relatively short period of time. This method seems to be particularly useful when applied to small regions (of the order of $5x10^6$ base pairs) of the human genome, available in cell hybrids generated by chromosome mediated gene transfer (21). In this case, only a few hundred cosmids would have to be analysed.

Other investigators have used random mapping in order to generate a physical map for other, simpler genomes. Coulson et al. have put 60% of the nematode genome ($8x10^7$bp) into contigs of two or more overlapping cosmids using a slightly different approach to fragment analysis (12). This supports the view that physical map construction for a human chromosome is feasible, since the size of chromosome 17 ($6x10^7$bp) is similar to that of the nematode genome.

It appears likely that during the next few years human gene mapping will move from sets of linked probes placed several million base pairs from one another to the existence of complete maps of overlapping DNA fragments. These will

represent a primary resource with which to study both simple and complex inheritance, in addition to the eventual sequencing of the human genome (22).

ACKNOWLEDGEMENTS

This work was funded by the Medical Research Council and the Cystic Fibrosis Research Trust. We are grateful to Dr. Denise Sheer (I.C.R.F.) for growing and characterising the PCTAB1.8 cell line, and to John Sulston, Alan Coulson and Peter Little for communication of data prior to publication nd for helpful discussions.

REFERENCES

1. Davies,K.E., Pearson,P.L., Harper,P.S., Murray,J.M., O'Brien,T., Sarfarazi,M. and Williamson,R. (1983) Nucl.Acids Res.8, 2303-2312.
2. Poustka,A. and Lehrach,H. (1986) Trends in Genetics 2, 174-179.
3. Williamson,R., Bell,G., Bell,J., Bates,G., Davies,K.A., Estivill,X., Farrall,M., Kruyer,H., Law,H.Y., Lench,N., Scambler,P., Stanier,P., Wainwright,B., Watson,E. and Worrall,C. (1986) Cold Spring Harbor Symposium 51, in press.
4. McKusick,V.A. (1986) Medicine 65, 1-33.
5. Hartley,D., Davies,K.E., Drayna,D., White,R.L. and Williamson,R. (1984) Nucl.Acids Res. 12, 5277-5285.
6. Solomon, E. and Bodmer, W.F. (1979) Lancet i, 923.
7. Botstein,D., White,R.L., Skolnick,M. and Davis,R.W. (1980) Am.J.Hum.Genet. 32, 314-331.
8. Drayna,D., Davies,K.E., Hartley,D., Mandel,J-L., Camerino,G., Williamson,R. and White,R. (1984) Proc.Natl. Acad.Sci. U.S.A. 81, 2836-2839.
9. Mendelsohn,M.L., Mayall,D.H., Bogart,E., Moore,D.H. and Perry,B.H. (1973) Science 179, 1126-1129.
10. Schmid,C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
11. Mosteller,F., Rourke,R.E.K. and Thomas,G.B. Jr. (1970) Probability with Statistical Applications (Addison-Wesley, Reading, Massachusetts), p. 105.
12. Coulson,A., Sulston,J., Brenner,S. and Karn,J. (1986) Proc.Natl.Acad.Sci. U.S.A., in press
13. Bai,Y., Sheer,D., Hiorns,L., Knowles,R.W. and Tunnacliffe, A. (1982) Ann.Hum.Genet. 46, 337-347.
14. Grosveld,F.G., Dahl,H.H.M., de Boer,E. and Flavell,R.A. (1981) Gene 13, 227-237.
15. Grosveld,F.G, Lund,T., Murray,E.J., Mellor,A.L., Dahl, H.H.M. and Flavell,R.A. (1982) Nucl.AcidsRes. 10, 6715-6732.
16. Deninger,P.L., Jolly,D.J., Rubin,C.M., Friedmann,T. and Schmid,C.W. (1981) J.Mol.Biol. 151, 17-33.

17. Feinberg,A.P. and Vogelstein,B. (1984) Anal.Biochem. 137, 266-267.
18. Little,P.F.R. and Cross,S.H. (1985) Proc.Natl.Acad.Sci. U.S.A. 82, 3159-3163.
19. Carle,G.F., Frank,M. and Olson,M.V. (1986) Science 232, 65-67.
20. Smith,C.L. and Cantor,C. (1986) Nature 319, 701-702.
21. Kamarck,M.E., Barker,P.E., Miller,R.L. and Ruddle,F.H. (1984) Exptl.Cell Res. 152, 1-14.
22. Robertson,M. (1986) Nature 322, 11.