Cosmid walking and chromosome jumping in the region of PKD1 reveal a locus duplication and three CpG islands

Gerald A.J.Gillespie[1], Gregory G.Germino[1], Stefan Somlo[1], Debra Weinstat-Saslow[1], Martijn H.Breuning[2] and Stephen T.Reeders[1,3]

[1]Yale University School of Medicine, New Haven, CT, USA, [2]University of Leiden, Leiden, The Netherlands and [3]Howard Hughes Medical Institute, Yale University, New Haven, CT, USA

ABSTRACT

The locus responsible for the most common form of autosomal dominant polycystic kidney disease (PKD1) is located on chromosome 16p13.3. Genetic mapping studies indicate that PKD1 is flanked on the proximal side by the DNA marker 26.6 (D16S125). Here we show that 26.6 has undergone a locus duplication and that the two loci are less than 150kb apart. One of the two loci contains a polymorphic *TaqI* site that has been used in genetic studies and represents the proximal boundary for the PKD1 locus. We demonstrate that the polymorphic locus is the more proximal of the two 26.6-hybridizing loci. Therefore, four cosmids isolated from the distal 26.6-hybridizing locus contain candidate sequences for the PKD1 gene. These cosmids were found to contain two CpG islands that are likely markers for transcribed regions. A third CpG island was detected and cloned by directional chromosome jumping.

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is the most common form of inherited kidney disease, affecting approximately one in 1,000 of the U.S. population. Genetic linkage studies have shown that mutations in at least two loci can lead to ADPKD (1). Approximately 95% of mutations affect a locus which has been assigned the symbol PKD1 that maps to chromosome 16p13.3. Since the original demonstration of linkage between the 3' hypervariable region of alpha-globin and PKD1 (2), a number of additional genetic markers, including phosphoglycolate phosphatase and at least 20 anonymous DNA markers, have been linked to PKD1. Somatic cell hybrid and genetic studies have localized PKD1 proximal to alpha-globin and very close to the anonymous DNA segment CMM65 (D16S84) (3). Breuning *et al* have isolated a clone, p26.6 (D16S125), which is currently the closest proximal flanking marker for PKD1 (4, 5).

In order to study the chromosomal region between the genetic markers flanking PKD1, we have isolated a number of overlapping cosmid clones that hybridize to p26.6. Initial comparison of the restriction maps of a number of these cosmids suggested that 26.6 hybridizes to two distinct loci. Both loci have been studied further and are found to lie within 150kb of one

another in 16p13.3. Because the polymorphism identified by p26.6 is an important landmark for attempts to clone the PKD1 gene, we studied the architecture of this region further. In particular we set out to determine which of the 26.6-hybridizing loci is polymorphic and the physical relationship between the two 26.6-hybridizing loci.

The cosmid clones were also used to identify two CpG islands that lie within the PKD region. A *Not*I jumping library was constructed and used to jump from one of these islands to a third CpG island in the direction of the distal flanking marker EKMDA2 (D16S83) (3).

METHODS

Probes

The probe p26.6 is a 0.4kb genomic *Pst*I fragment (4, 5).

Cosmids isolated with p26.6 were found to fall into two non-overlapping sets (cos3 and cos4/cosMB26). Single-copy subclones were isolated from a member of each group of cosmids. JA7 is a 0.7kb *Bam*HI fragment from cos4 and JA9 is a 4.3kb *Bam*HI fragment from cos3. JA7 and JA9 were used to isolate overlapping cosmids from the distal (cosMB26, cos4, cosNK14, cos7) and proximal (cos3, cos6) 26.6-hybridizing loci respectively.

Genomic Libraries

A genomic cosmid library, constructed in the vector pcos2EMBL, was a kind gift of Anna-Maria Frischauf (ICRF, London). Another genomic library constructed in the vector pWE15 (6) was also used. These libraries were screened with single-copy DNA probes labelled with ^{32}P-dCTP by the random hexamer priming method (7).

Pulsed-field mapping

Long-range mapping was performed as described by Chorney *et al* (8) using orthogonal or contour-clamped homogenous electric fields. DNA, prepared in agarose plugs (9) was from the lymphoblastoid cell line 3.1.0 (kindly provided by Don Pious, University of Washington, Seattle). Pulse times were varied from 60−120 seconds to optimize separation in the desired size range. Molecular size markers were concatamers of λ DNA prepared as described (9). Pulsed-field gels were treated with 0.25M HCl for 15 minutes, rinsed, immersed in 0.5M NaOH/1.5M NaCl

for 30 minutes and then transferred in 0.5M NaOH/1.5 NaCl onto Hybond-N membranes (Amersham). Transfer was allowed to proceed for about 24 hours. Filters were rinsed in 5×SSC, air dried and U.V. irradiated for 20 seconds. Hybridization was performed in 0.5M sodium phosphate, pH7.2, 7% SDS, 1% BSA, 1mMEDTA at 65°C overnight, and blots were washed at 0.1×SSC 0.1% SDS at 65°C for 15 minutes.

Jumping Library Construction

The jumping library was prepared as described by Collins *et al* (10) except that DNA was prepared in solution instead of in agarose plugs and no size fractionation of digested DNA was performed. High molecular weight DNA was isolated (11) from 3.1.0 cells and was completely digested with *Not*I (20u/μg DNA) for 2 hours at 37°C. The enzyme was inactivated by heating to 65°C for 30 minutes. The DNA was diluted to a concentration of ~0.1μg/ml and a 50−100 fold molar excess of a selectable marker DNA fragment was added. The marker, a 219 bp suppressor tRNA cloned fragment flanked by *Not*I, *Xho*I and *Bam*HI sites, was kindly provided by Francis Collins (University of Michigan, Ann Arbor) as the plasmid pBXN-SupF. Following addition of MgCl$_2$ to 10mM and ATP to 1mM, 10μl of T4 DNA ligase were added (New England BioLabs, 400u/μl) and ligation was allowed to proceed for 12 hours at 15°C. The ligation mix was incubated with an additional 10μl of ligase for a further 12 hours. Ligation was terminated by heat inactivation. The solution was then adjusted to 100mM NaCl and the DNA was completely digested with *Eco*RI. The mix was extracted twice with phenol/chloroform, once with chloroform, and ethanol-precipitated in the presence of 20μg of carrier tRNA. Resuspended DNA (approximately 200 ng) was ligated to 4μg of *Eco*RI digested bacteriophage λ Charon 3AΔlac (12) for 16 hours at 15°C, packaged *in vitro* and transfected into *E. coli* MC1061. Only cells carrying recombinant bacteriophage which incorporate the suppressor fragment form plaques on this host.

RESULTS

Cloning and long-range mapping of two 26.6-hybridizing loci

A human genomic library in the vector pcos2EMBL was screened with the 0.4kb *Pst*I fragment, p26.6, and two independent cosmid clones were obtained. Restriction mapping of these cosmids, and cosMB26, a cosmid from which p26.6 had originally been subcloned, identified single-copy fragments that were used as walking probes. Although p26.6 had been thought to hybridize to a single locus, restriction maps indicated that the cosmids represent two non-overlapping loci. In addition, p26.6 detected two distinct *Taq*I fragment patterns in the set of cosmids (figure 1a). cosMB26 and cos4, both contain 0.4kb and 0.7kb p26.6-hybridizing *Taq*I fragments whereas cos3 contains an ~1.1kb p26.6-hybridizing *Taq*I fragment. Other restriction digests confirmed that the cosmids represent non-overlapping loci (data not shown). The walking probe JA7, a 0.7kb *Bam*HI fragment and JA9, a 4.3kb *Bam*HI fragment, were isolated from cos4 and cos3 respectively and used to screen both the pcos2EMBL library and an additional cosmid library in the vector pWE15 for overlapping clones. JA9 yielded cos6 from the pWE15 library, which overlaps with cos3 by 20kb; JA7 yielded cosNK14 from the pWE15 library and cos7 from the pcos2EMBL library (figure 1b).

Having identified two separate 26.6-hybridizing loci and having made unidirectional walks from each of these, it became

important to establish which of the two corresponded to the polymorphic locus reported by Breuning *et al* (4, 5) since this polymorphism marks the proximal boundary of the PKD1 region. It was also important to determine the relative orientation of these two loci with respect to the chromosome and the distal flanking markers so that cloning of the PKD1 region could be initiated. A second 26.6-hybridizing locus had not been anticipated since p26.6 was seen to hybridize to single *Mlu*I, *Not*I and *Bss*HII fragments of ~450kb, ~180kb and ~150kb respectively (figures 2a and 2b). The hybridization of p26.6 to single large restriction fragments but to two separate sets of cosmids suggested that the two hybridizing loci were located no further apart than 150kb on the same BssHII fragment.

Hybridization of p26.6 to *Nru*I and *Mlu*I fragments of ~700kb and ~450kb respectively was observed in the cell line 3.1.0 (figure 2b). Double digestion with *Mlu*I and *Nru*I yielded a fragment of ~300kb (figure 2b). A second probe, pGGG12, a subclone from a 50kb cosmid contig containing the linking clone N54 (13) has also been shown to hybridize to the same ~700kb *Nru*I fragment. N54 lies on the same *Cla*I fragment as CMM65 (data not shown), which is distal to the polymorphic locus recognized by 26.6 (3, 13). JA7, a single-copy fragment isolated from cos4, was also found to hybridize to the 700kb *Nru*I fragment and is within 4kb of an *Mlu*I site in the cosmid. This *Mlu*I site was found to be cleaved in 3.1.0 genomic DNA (figure 3a) and therefore allowed cos4 to be assigned to the more distal 26.6-hybridizing locus (figure 2c).

Mapping of the polymorphic 26.6-hybridizing locus

To determine whether the proximal or distal 26.6-hybridizing cosmids contain the polymorphic locus, the following experiments were conducted. Hybridization of p26.6, a 0.4kb *Pst*I fragment (containing an internal *Taq*I site) to a *Taq*I digest of a panel of genomic DNA samples from different individuals resulted in either three bands (1.1kb, 0.7kb, and 0.4kb), four bands (6kb, 1.1kb, 0.7kb, and 0.4kb), or three bands (6kb, 0.7kb, and 0.4kb). In family studies the 1.1kb band has been shown to be allelic with the 6kb band while the 0.7kb and 0.4kb bands are constant (4, 5). In order to determine which cosmid contig contains the polymorphic 26.6 locus, the cosmids were digested with *Taq*I, blotted, and hybridized with 26.6. Cosmids MB26 and cos4 were found to contain the invariant 0.7kb and 0.4kb bands, whereas cosmid 3 was found to contain the 1.1kb band (figure 1b), which is an allele of the polymorphic system. Therefore, the more proximal 26.6-hybridizing locus, 26.6PROX, represented by cos3, is the polymorphic locus. The exact distance between the two 26.6-hybridizing loci has not as yet been determined.

Two CpG islands are present within the cosmids spanning 26.6DIS

As a first step toward the isolation of expressed sequences we attempted to locate CpG islands from within the cos4 contig. In so doing, we identified two *Not*I sites within cos4 and showed that these are cleaved in 3.1.0 genomic DNA. We examined possible clustering of non-methylated CpG sites within this region using *Eco*RI digestion of 3.1.0 DNA in combination with a variety of endonucleases that include the dinucleotide CpG as part of their recognition sequence. Figure 3a shows that a number of sites for these enzymes are clustered, for example, within the region of JA7, defining a putative CpG island (figure 3b). Similar results (data not shown) were obtained for the second, more proximal *Not*I site within cos4.
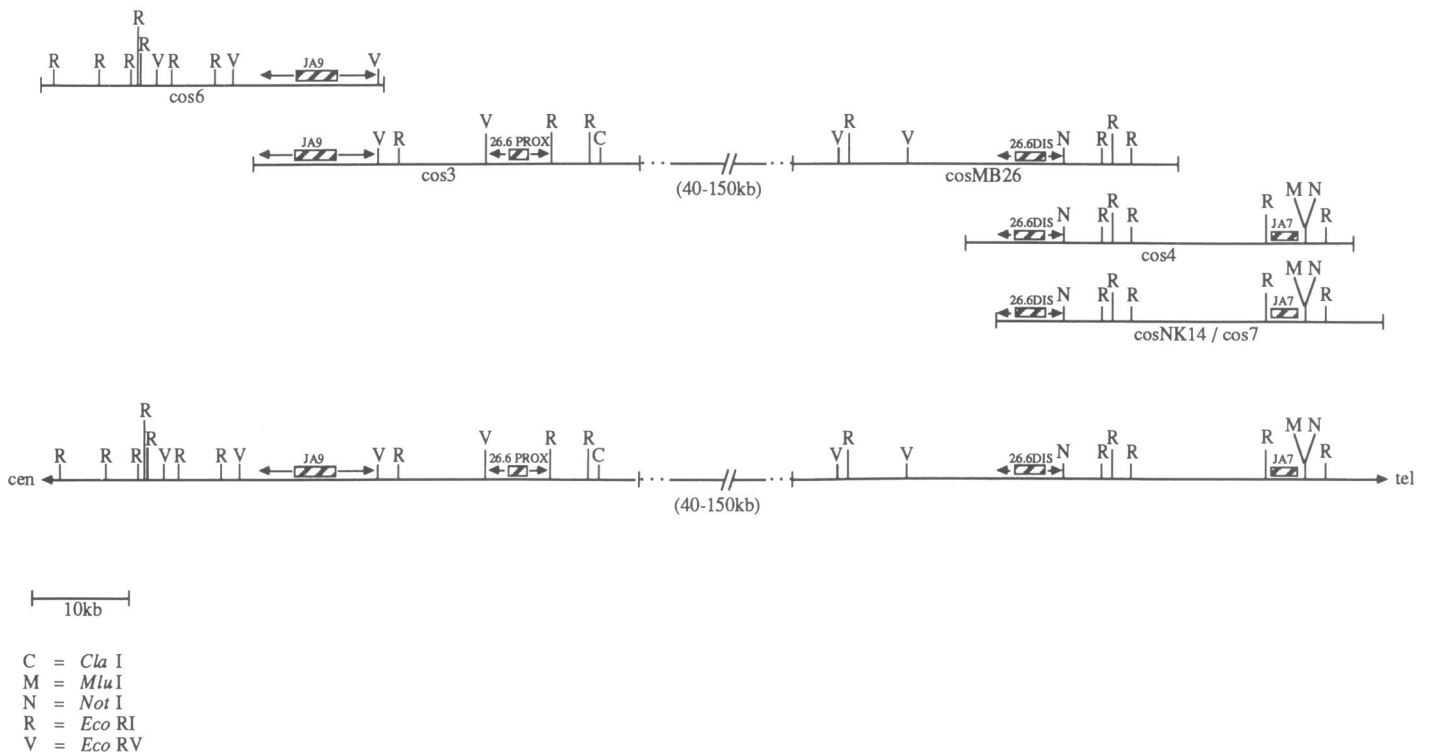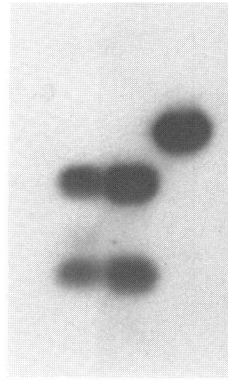
Figure 1a: Pattern of hybridization of the probe p26.6 to *Taq*I-digested cosmid DNA from cos3, cos4 and cosMB26. **Figure 1b:** Restriction maps of cosmids from the 26.6-hybridizing loci.

Directional jumping from 26.6DIS

Having identified the order of 26.6PROX and 26.6DIS, further attempts were made to extend the walk beyond cos7 in a distal direction towards the opposite flanking markers. These attempts were unsuccessful. To overcome this problem, a *Not*I jumping library was constructed. Such a library allows chromosome jumps to be made between *Not*I sites that are cleaved in genomic DNA. Because 89% of *Not*I sites in genomic DNA lie within CpG islands and because CpG islands mark the 5' ends of genes in many cases (14), a *Not*I jumping library also allows one to jump from one gene to the next (15). Not all CpG islands contain *Not*I sites, however, so *Not*I jumping libraries do not contain all islands. JA7, previously shown to lie within 4kb of an *Mlu*I site, was found to be close to a *Not*I site which is also cleaved in genomic DNA (figure 3a). However, screening of the library with JA7 would result in a proximal jump to an already cloned region since JA7 lies proximal to the *Not*I site (figure 3b).

Therefore, to make a distal jump, a single-copy 0.7kb *Sma*I fragment, JA20, was isolated from the *Not*I-*Eco*RI fragment immediately distal to the *Not*I site (figure 3b). A single positive jumping clone, λJA21, was obtained. Pulsed-field mapping data (G.G. Germino, in preparation) indicate that the first *Not*I site distal to JA20 that is cleaved in genomic DNA lies ~40kb away. This *Not*I site should, therefore, represent the end point of the jump.

Two previously characterized somatic cell hybrids (N-OH1 and 23HA) were used to show that λJA21 resulted from a legitimate intramolecular ligation (circularization). Each hybrid cell-line contains a single der(16) chromosome as the only human component (3). Hybrid N-OH1 contains a der(16)t(1;16)-(p36.3;p13.3) chromosome in which the translocation breakpoint has been mapped to a 6kb interval within 16p13.3. Hybrid 23HA contains a der(16)t(16;22)(p13.3;q12.2) chromosome in which the 16p13.3 translocation breakpoint lies proximal to the breakpoint in N-OH1 (3). The orientation of the breakpoints with
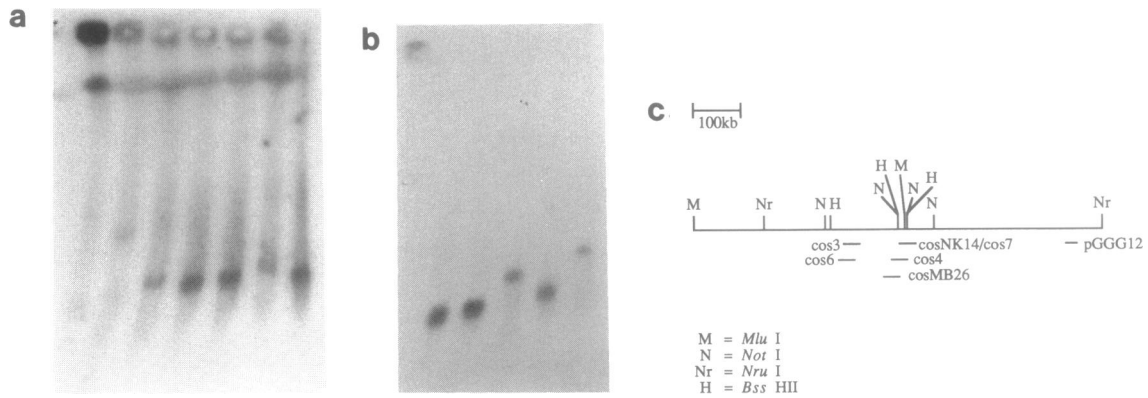
Figures 2a and 2b: Hybridization of probe p26.6 to 'infrequent cutter' digests of human genomic DNA from the cell-line 3.1.0, separated by pulsed-field gel electrophoresis. **Figure 2c:** A long-range restriction map of the region encompassing the 26.6PROX and 26.6DIS loci.
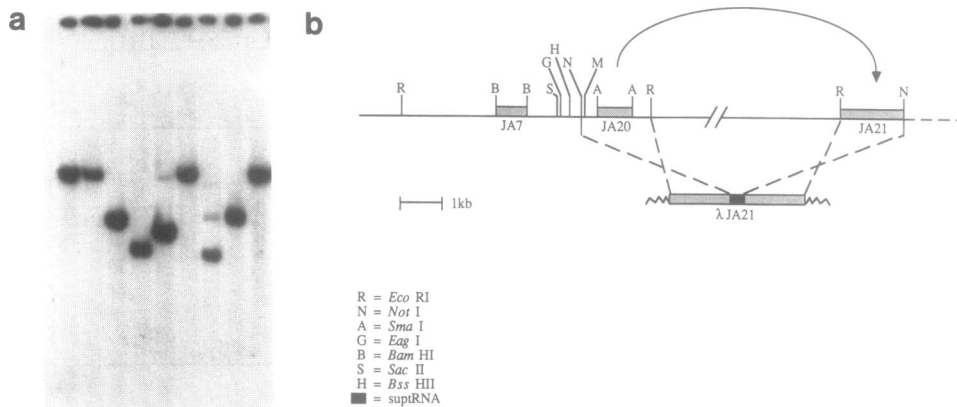

Figure 3a: Double-digests of genomic DNA, blotted and hybridized with pJA7, reveal several sites for 'infrequent cutting' enzymes. Partial digestion with *Bss*HII and *Sst*II was observed. **Figure 3b:** The restriction map of the CpG island adjacent to JA7 and a schematic representation of the chromosome jumping strategy.
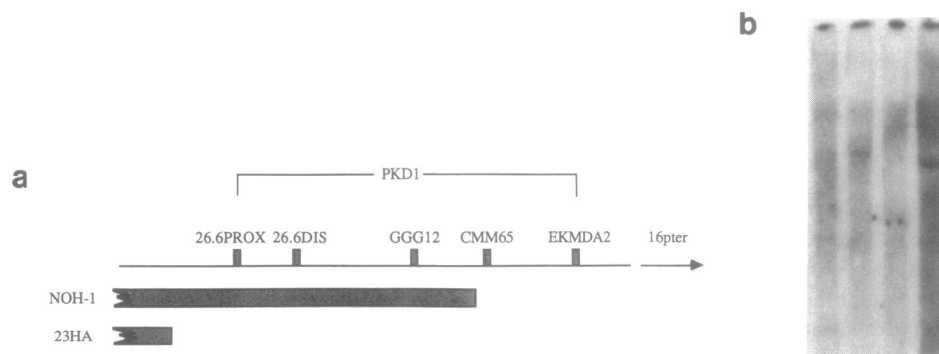

Figure 4a: The location of the chromosome 16 translocation breakpoints in the N-OH1 and 23HA somatic cell hybrid lines with respect to probes from the PKD1 region of chromosome 16p13.3 (not to scale). The shaded bars represent regions of chromosome 16 that are present in the hybrids. The region defined by the flanking markers for PKD1 is bracketed. **Figure 4b:** Hybridization of the end-point of the jump clone, pJA21, to *Bam*HI-digested somatic cell hybrids DNA demonstrates that the end-point of the jump is present in genomic DNA and in the hybrid N-OH1 but not in the hybrid 23HA, placing it between the hybrid translocation breakpoints.

respect to markers defining the PKD1 region (3, 13) is summarized in figure 4a. A 1.5kb *Not*I-*Eco*RI fragment, pJA21, representing the distal end of the jump clone, was isolated from λJA21 and used to probe *Bam*HI-digested hybrid DNA (figure 4b). pJA21 lies between the hybrid breakpoints as does the starting point of the jump, JA20 (not shown). pJA21 was hybridized to genomic DNA digested with several infrequent-

cutting restriction endonucleases and was found to lie close to or within a CpG island.

DISCUSSION

There are several critical steps in the reverse genetic approach to the identification of genes that are mutated in human disease.

In the absence of cytogenetic abnormalities that mark the position of the disease gene, perhaps the most important step is the correct identification of close flanking markers on either side of the disease mutation. Since the disease gene and its associated mutations can only be tracked through an abnormal phenotype, genetic markers flanking the mutation are the only means of defining the physical boundaries of the region containing the disease locus. When the closest flanking markers are still some distance away from the disease mutation, analysis of large numbers of families will reveal several crossovers. As the refinement of mapping of the disease locus improves, fewer crossovers are available to define the region of interest. When only one or two crossovers define the position of the disease, it becomes critical to characterize the clones defining the crossovers in detail and to orientate them physically with respect to the chromosome.

In the case of PKD1, the closest proximal marker is defined by the probe p26.6 (4, 5). The observation, reported here, that p26.6 recognizes not one but two loci on chromosome 16, indicates the need to localize both loci to determine which of them is polymorphic and to determine their relative orientations. We found that both 26.6-hybridizing loci, 26.6PROX and 26.6DIS, are within 150kb of each other and that the proximal locus, 26.6PROX, contains the *Taq*I polymorphism that has been used in family studies (4, 5). Therefore, 26.6DIS lies up to 150kb within the PKD1 region as the latter is defined by 26.6PROX, the closest proximal flanking marker. Since these observations were made, *in situ* hybridization of cosMB26 to interphase nuclei has revealed two very close but distinct signals (J.G. Dauwerse, unpublished).

The cosmids that define the 26.6DIS locus have been shown to contain at least two CpG islands and the end point of the chromosome jump resides in a third. In all reported examples, CpG islands have been found to mark the positions of genes. Although the literature may reflect a reporting bias, it is highly likely that cos4 and cos7 and pJA21 contain parts of at least three transcribed sequences and thus provide a focus for the search for candidate genes for PKD1. Indeed the CpG island detected by JA7 has allowed the isolation of a number of clones from a cystic kidney epithelial cell cDNA library (G.A.J. Gillespie, in preparation).

Cosmid 'walking' is frequently halted by the failure to detect overlapping clones in any given cosmid library. It is our experience that when one genomic cosmid library fails to contain an overlap, the use of additional cosmid and bacteriophage libraries increases the chance of finding an overlap by less than 50%. Two well-characterized libraries were used in this study but neither allowed a walk to be made from cos7. Although a systematic study to determine the reasons for poor representation of certain sequences has not been conducted, two plausible explanations can be put forward. One is that the combination of vector, insert and host is biologically unstable. Another possibility is that variation in the spacing and cleavage rate of cloning sites in the genome is such that there is marked non-randomness of the position of insert fragment ends.

Chromosome jumping provides a means to overcome difficulties in cosmid overlap walking. In this context, since the size of the jump is immaterial (provided that it is sufficient to 'step over' the poorly represented region), the problems inherent in making large jumps (>200kb) are not limiting. Moreover, if preliminary restriction mapping suggests that jumping to the far end of a defined complete-digest fragment fulfills the purpose, the extra steps involved in size fractionation and control of partial digests can be omitted. Finally, the use, in the construction of the library, of enzymes that are frequently represented in and relatively specific for CpG islands allows a jump to be made directly to the site of a 'new' CpG island.

ACKNOWLEDGMENTS

REFERENCES

1. Kimberling, W.J., Fain, P.R., Kenyon, J.B., Goldgar, D., Sujansky, E., and Gabow, P.A. (1988) *N.E.J.M.*, **319**:913–918.
2. Reeders, S.T., Breuning, M.H., Davies, K.E., Nicholls, D.R., Jarman, A.J., Higgs, D.R., Pearson, P.L., and Weatherall, D.J. (1985) *Nature*, **317**, 542–544.
3. Germino, G.G., Barton, N.J., Lamb, J., Higgs, D.R., Harris, P., Scherer, G., Nakamura, Y. and Reeders, S.T. (1990) *Am. J. Hum. Genet.*, **46**, 925–933.
4. Breuning, M.H., Snijdewint, F.G.M., Brunner, H., Verwest, A., IJdo, J.W., Saris, J.J., Dauwerse, J.G., Blonden, L., Keith, T., Callen, D.F., Hyland, V.J., Xiao, H.G., Scherer, G., Higgs, D.R., Harris, P., Bachner, L., Reeders, S.T., Germino, G., Pearson, P.L. and van Ommen, G.J.B. (1990) *s, S.T., and Frischauf, A-M. (1990J. Med. Genet.*, in press.
5. Breuning, M.H., Snijdewint, F.G., Smits, J.R., Dauwerse, J.G., Saris, J.J. and van Ommen, G.J. (1990) *Nucleic Acids Res.*, **18**, 3106.
6. Evans, G.A. and Wahl, G.M. (1987) *Meth. Enzymol.*, **152**, 604–610.
7. Feinberg, A.P. and Vogelstein, B. (1983) *Analy. Biochem.*, **132**, 6–13.
8. Chorney, M.J., Sawada, I., Gillespie, G.A., Srivastava, R., Pan, J. and Weissman, S.M. (1990) *Mol. Cell. Biol.*, **10**(1), 243–253.
9. Smith, C.L., Lawrence, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M. and Collins, F.S. (1987) *Meth. Enzymol.*, **151**, 461–489.
10. Collins, F.S., Drumm, M.L., Cole, J.L., Lockwood, W.K., Vande Woude, G.F., and Iannuzzi, M.C. (1987) *Science*, **235**, 1046–1049.
11. DiLella, A.G. and Woo, S.L.C. (1987) *Meth. Enzymol.*,**152**, 189–212.
12. Blattner, F.R., Williams, B.G., Brechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Greenwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) *Science*, **196**, 161–169.
13. Himmelbauer, H., Germino, G.G., Ceccherini, I., Romeo, G., Reeders, S.T., and Frischauf, A-M. (1990) *Am. J. Hum. Genet.*, in press.
14. Lindsay, S. and Bird, A.P. (1987) *Nature*, **327**, 336–338.
15. Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M., and Lehrach, H. (1987) *Nature*, **325**, 353–355.