Vol. 205, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1994 Pages 482-489

MAPPING OF TWO PHENOL SULPHOTRANSFERASE GENES, *STP* AND *STM*, TO 16p: CANDIDATE GENES FOR BATTEN DISEASE

Thomas P. Dooley[1], Hannah M. Mitchison[2], Patricia B. Munroe[2], Pamela Probst[1], Mindy Neal[1], Michael J. Siciliano[3], Zuoming Deng[3], Norman A. Doggett[4], David F. Callen[5], R. Mark Gardiner[2] and Sara E. Mole[2*]

[1]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas

[2]Department of Paediatrics, UCL Medical School, The Rayne Institute, London, UK

[3]Department of Molecular Genetics, M.D. Anderson Cancer Center, Houston, Texas

[4]Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, New Mexico

[5]Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, Adelaide, Australia

Received August 10, 1994

SUMMARY: The cytosolic phenol sulphotransferase gene (*STP*) was mapped to a region of chromosome 16, within the interval defined by human-rodent somatic cell hybrid breakpoints CY160(D) and CY12, which contains *FRA16E*. YAC and cosmid clones from this 16p interval were screened for the presence of *STP*. Two non-overlapping cosmid contigs were identified which contain *STP*-like sequences. Sequencing of these *STP*-like sequences confirmed that *STP* is contained within contig 343.1 and maps proximal to *FRA16E*, and that a related sulphotransferase *STM*, encoding the catecholamine-sulphating enzyme, is contained within contig 55.4 and maps to the adjacent hybrid interval CY12-CY180A. Thus two phenol sulphotransferase genes (*STP* and *STM*) have been finely localised to chromosome 16p12.1-p11.2, to the same region as *CLN3*, the gene for Batten disease. Both genes are therefore candidate genes for Batten disease.

The cytosolic phenol-preferring phenol sulphotransferase gene (*STP*) has been mapped to the interval CY165-CY12 (1) defined by a panel of human-rodent somatic cell hybrids containing fragments from chromosome 16 (Figure 1) (2). This contains the CY160(D)-CY12 interval to which the closest markers to *CLN3*, the gene for Batten disease, localise. Phenol sulphotransferase enzymes (PST, EC 2.8.2.1) catalyze the transfer of the sulphonate group from PAPS (phosphoadenosine phosphosulphate) to an acceptor compound (3) and have been demonstrated to

*Corresponding author and reprint requests:
Department of Paediatrics, University College London Medical School, The Rayne Institute, University Street, London, UK WC1E 6JJ. Fax: +71 380 9681.

Abbreviations used: CNS (central nervous system); NCL (neuronal ceroid lipofuscinosis); YAC (yeast artificial chromosome); CEPH (Centre d'Etude du Polymorphisme Humain); bp (base pair); kb (kilobase); PAGE (polyacrylamide gel electrophoresis); PCR (polymerase chain reaction); PAPS (phosphoadenosine phosphosulphate); PST (phenol sulphotransferase).

0006-291X/94 $5.00

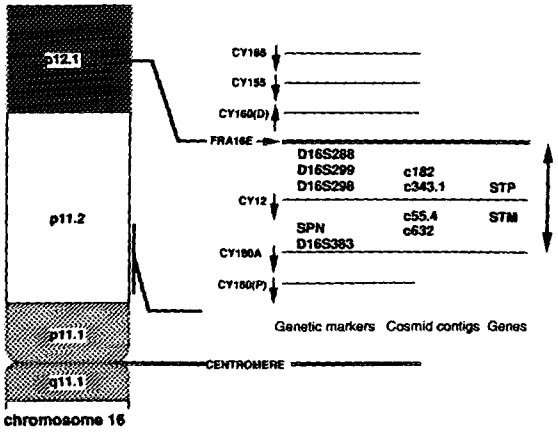

Figure 1. Diagrammatic representation of the mapping of five markers, four cosmid contigs and two phenol sulphotransferase genes, *STP* and *STM*, on the short arm of chromosome 16. The intervals are defined by a set of human-rodent somatic cell hybrids. The horizontal lines represent the region of chromosome 16 contained within each hybrid, with the small vertical arrows indicating the direction of fragment retained. The position of the *FRA16E* site relative to these breakpoints is shown. The double ended vertical arrow indicates the region known to contain *CLN3*.

sulphate a wide variety of phenolic compounds. They are expressed in many tissues including brain (3, 4). Sulphation of substrate molecules produces more highly charged, hydrophilic products. In general, phenol sulphotransferases are considered to play essential roles in the metabolism and clearance of drugs and endogenous compounds, such as dopamine.

At least two highly related phenol sulphotransferase enzymes have been purified and categorized by substrate specificity. These are the phenol-preferring P-PST (32 kDa), also referred to as thermostable PST (TS PST), which sulphates simple phenolic compounds, and the monoamine-preferring M-PST (34 kDa), also referred to as thermolabile PST (TL PST), which sulphates catecholamines (3, 4, 5, 6). Purified PSTs can sulphate catecholamines (e.g. dopamine and norepinephrine), p-nitrophenol, minoxidil, and selected tyrosine-containing peptides (e.g. enkephalin and cholecystokinin (3, 4, 5). Both P- and M-PST can sulphate catecholamines *in vitro*. However the M-form of this enzyme has been implicated in this process in human brain as it has a better apparent Km for dopamine. M-PST is antigenically related to P-PST suggesting that the two enzymes are closely related in terms of amino acid sequence (4). cDNA sequences for P-PST (*STP*) (7, 8, 9) and M-PST (*STM*) (9, 10) have been isolated. The coding regions are 94.3% identical at the DNA level and 92.5% at the amino acid level.

Batten disease is the juvenile type of neuronal ceroid lipofuscinosis (NCL), neurodegenerative diseases characterised by the accumulation of autofluorescent lipopigment in neurons and other tissues and extensive death of cortical grey matter (11). A variety of biochemical changes associated with Batten disease have been reported although the basic defect is still unknown. Genetic linkage analysis has allowed the gene, *CLN3*, to be mapped within a small region (2.1 cM) on the short arm of human chromosome 16p12.1-p11.2 (12). Two markers, *D16S288* and *D16S383*, which detect recombinations flanking *CLN3*, are located in adjacent

intervals defined by a panel of human-rodent somatic cell hybrids containing fragments from chromosome 16 (Figure 1) (2). Linkage disequilibrium analysis of haplotypes in Batten families with four highly informative microsatellite markers, *D16S288*, *D16S299*, *D16S298* and *SPN*, also located within the hybrid interval CY160(D)-CY180A, have identified a core haplotype present in 73% of disease families analysed (12, 13, 14). *D16S299* and *D16S298*, the markers in strongest linkage disequilibrium with *CLN3*, are localised to interval CY160(D)-CY12. This suggests that *CLN3* is also most likely to be located in this interval. Therefore any gene which maps to this interval is a candidate for *CLN3*.

Because *STP* has been mapped to the same region as *CLN3* on chromosome 16p, is expressed in neurons in human brain and is capable of metabolizing some lipophilic compounds, and since Batten disease is characterised by the accumulation of lipophilic compounds within neurons, *STP* has been proposed as a candidate gene for Batten disease (1). We report the refinement of localisation of *STP* on chromosome 16 and the localisation of the related gene *STM* also to chromosome 16. Both genes map to adjacent intervals of the chromosome 16 hybrid panel between markers *D16S288* and *D16S383* and therefore within the region known to contain *CLN3*.

MATERIALS AND METHODS

General Methods: DNA from genomic clones was prepared using standard methods (15). PCRs were performed using standard cycling conditions, typically 1 cycle of 95°C/2-5 min; 30 cycles of 94°C/1 min, 55°C-58°C/1-2 min, 72°C/1-3 min; 1 cycle of 72°C/7-10 min. Products were visualised using ethidium bromide staining after electrophoresis in an agarose gel. Primers used were STPA (5'CTCAGGAACATGGAGCTGAT) and STPB (5'TGTGTCTTCAGGAGTCGTG) (1); STPF10 (5' TTCAAAGCCCCAGGGATT) STPR14 (5' CCTTGGCCATGTGGTAGAAG); STMF2 (5' GGTCAATGATCCAGGGGAA) and STMR6
(5' AAGAGGTAGAGAACAGGGTGGG).
Somatic Cell Hybrid Mapping; Screening of YAC and Cosmid Genomic Clones: Oligonucleotide primers were used to amplify DNA from somatic cell hybrids CY18, which contains the the whole of chromosome 16, and CY165, CY155, CY160 and CY12 which contain defined portions of chromosome 16 (Figure 1) (2) or from preparations of YAC and cosmid DNA by PCR. The products were electrophoresed on agarose gels and stained with ethidium bromide.
DNA Sequencing: Double-stranded DNA sequencing was performed on PCR amplified plasmid subclones (derived from cosmids 330A9 and 4D10) using the dideoxynucleotide chain termination method and [^{32}P]α-dATP. The reactions products were resolved on 7% PAGE, 50% urea gels, and were visualized by autoradiography (15). Alternatively an ABI automatic sequencer was used to resolve reaction products produced by following the manufacturers protocols. Oligonucleotides specific to the vector sequences and to *STP* cDNAs were used as the sequencing primers. All DNA sequences were analyzed on a Macintosh IIsi computer with MacVector version 4.1 software or ABI Sequence Navigator, and were compared to the known *STP*/P-PST (7) and *STM*/M-PST (9, 10) cDNA sequences.

RESULTS

Localisation of *STP* to hybrid interval CY160(D)-CY12: *STP* has already been mapped to the CY165-CY12 interval defined by somatic cell hybrids containing fragments of chromosome 16 (1). We refined the mapping of *STP* by PCR screening of two additional hybrids, CY155 and CY160, using primers STPA and STPB which amplify a 525 bp fragment at the 5' end of the *STP* coding region (1). The results are presented in Figure 2. Since CY155 was amplified, whereas CY160 and CY12 were not, *STP* must map within the interstitial deletion contained in CY160 but

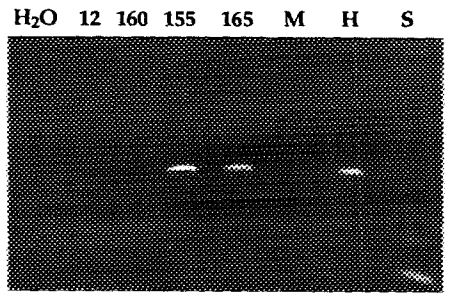

Figure 2. Mapping of *STP* within the CY160(D)-CY12 interval of a human-rodent somatic cell hybrid panel. Primers STPA and STPB were used to amplify the 525 bp *STP* genomic fragment on CY12, CY160, CY155, and CY165 hybrids containing chromosome 16p breakpoints and on control samples of mouse Cl1D (M) and human HELA (H) genomic DNAs. S is a 123-bp size marker. H_2O indicates a negative control.

above the breakpoint in CY12, that is, within the interval CY160(D)-CY12 (Figure 1). This region also contains *D16S299* and *D16S298*, the closest markers to *CLN3* (12).

Identification of genomic clones containing *STP*: YAC and cosmid genomic clones in this region of chromosome 16 were screened for the presence of *STP*. Three overlapping YAC clones known to contain the closest markers to *CLN3* (I Järvelä, personal communication) were screened by PCR using the STPA and STPB oligonucleotides. CEPH YACs containing *D16S288* only (359C10), or *D16S299* only (302G12), were negative. The YAC containing *D16S298* only (85D3), was positive (Figure 3). This suggests *STP* maps close to *D16S298*.

The predicted order of these markers is tel-*D16S288-D16S299-D16S298*-cen (16). YAC 359C10 and a cosmid clone 19C11 which both contain *D16S288* and a cosmid clone 16-129 containing marker *D16S272* (17), which maps proximal to *D16S288* in YAC 359C10, have been positioned proximal to *FRA16E* by fluorescent *in situ* hybridisation (data not shown). Therefore *STP* must also map proximal to *FRA16E*.

Cosmid contigs known to map in this region of chromosome 16 (Doggett et al., in preparation), were screened by PCR using the STPA and STPB oligonucleotides (Figure 3). A 525 bp fragment from cosmids contained in two cosmid contigs (304C7, 330A9 and 39G7 from

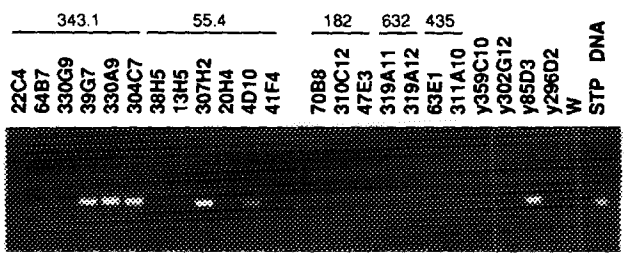

Figure 3. Mapping of *STP* to chromosome 16 cosmid contigs 343.1 and 55.4. Primers STPA and STPB were used to amplify the 525-bp genomic fragment from cosmid members of contigs 343.1, 55.4, 182, 632 and 435. Positive controls include YAC 85D3 and a plasmid subclone of an *STP* genomic fragment. Negative controls include YACs 359C10, 302G12, 296D2 and H_2O (W).

contig 343.1 and 307H2 and 4D10 from contig 55.4) was amplified. No amplification from cosmids in contigs 182, 632 or 435 was observed. The presence of *STP* sequence was confirmed by hybridisation of radiolabelled *STP* cDNA to Southern blots of cosmids digested with restriction enzymes (data not shown). These results suggest that two genes related to *STP* are present in this region of chromosome 16 in two non-overlapping cosmid contigs and the sequence of these two genes are sufficiently conserved for the purified cosmid DNA to be used as template by the primers STPA and STPB and to be hybridised by the *STP* cDNA probe on Southern blots.

Localisation of *D16S48* to CY160(D)-CY12: Contig 343.1, which also contains marker *D16S48* in cosmid 304C7, had previously been reported as mapping to hybrid interval CY12-CY180A (18). The data presented here, which show *STP* is contained in contig 343.1, combined with the order of genetic markers predicted in (19), suggests that the localisation of *D16S48* should be reassigned to interval CY160(D)-CY12. This is consistent with the localisation of another cosmid contig defined at the distal end by *D16S298* and at the proximal end by the STS 80D12R, which both map to CY160(D)-CY12, and which contains within it contig 343.1 (P.E.M. Taschner personal communication).

Identification of PST genes: PstI fragment plasmid subclones were isolated from 330A9 (contig 343.1) and 4D10 and 307H2 (contig 55.4), and checked for the presence of *STP* sequence by Southern blotting. Those clones containing phenol sulphotransferase fragments were sequenced. A clone derived from cosmid 330A9 contains sequence identical to the published cDNA sequence of P-PST in coding exon 1 (7) (Figure 4). Additional regions cloned from contig 343.1 have also been sequenced and found to be identical to *STP* (data not shown). Clones derived from cosmids 4D10 and 307H2 contain sequence which is not identical to *STP*, but is identical to sequence from coding exon 1 in the M-PST clone (9, 10) (Figure 4), that is from *STM*. Additional regions cloned from contig 55.4 have also been sequenced and found to be identical to *STM* (data not shown). *STM* is 92.5% identical at the amino acid sequence level to *STP* (ie. 273 of 295 residues are identical). The sequence of *STM* is sufficiently similar to *STP* at the site of hybridisation of oligonucleotides *STP*A and *STP*B to allow amplification from purified cosmid clones from contig 55.4.

Localisation of *STM* to the interval CY12-CY180A: Primers were designed from the published sequence of P-PST and M-PST which specifically amplified either *STP* or *STM* respectively. *STP* specific primers amplified DNA from cosmid 330A9 in contig 343.1 and *STM* specific primers amplified DNA from cosmid 307H2 in contig 55.4 (Figure 5). The localisation of

```
STP     CAGGCCCGGCCTGATGACCTGCTCATCAGCACCTACCCCAAGTCCG-3'
330A9   ..........................................
STM     ..A.....A.....................A.............T.
307H2   ..A.....A.....................A.............T.
```

Figure 4. Partial DNA sequences of phenol sulphotransferase genes in contigs 343.1 and 55.4. A selected portion of the DNA sequences from the coding region in exon 1 from: (1) *STP* cDNA and genomic clones (1, 7); (2) a subclone of cosmid 330A9 from contig 343.1; (3) *STM* cDNA clone (9, 10); and (4) a subclone of cosmid 307H2 from contig 55.4. Identical sequence is indicated by '.'. The final 'G' is adjacent to an intron splice donor site.

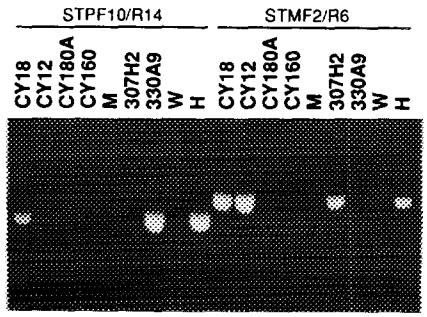

Figure 5. Mapping of *STM* within the CY12-CY180A interval of a human-rodent somatic cell hybrid panel. Primers STPF10 and STPR14 were used to specifically amplify the 1.49-kb *STP* genomic fragment on CY18, CY12, CY180A, CY160 hybrids containing chromosome 16p breakpoints and control cosmids 307H2 and 330A9. Primers STMF2 and STMR6 were used to specifically amplify the 1.67-kb *STM* genomic fragment on CY18, CY12, CY180A, CY160 hybrids containing chromosome 16p breakpoints and control cosmids 307H2 and 330A9. H and M indicate the human and mouse (A9 cell line) genomic controls. W indicates a water negative control.

the *STM* gene was determined by PCR screening of hybrids containing breakpoints in the 16p12-11.2 region using *STM* specific primers. DNA fragments were amplified from CY18 and CY12 but not from CY160 or CY180A (Figure 5). This gives a localisation for *STM* to the interval CY12-CY180A. *STP* specific primers did not amplify from CY12, CY180A or CY160, consistent with the localisation of *STP* to the interval CY160(D)-CY12 as demonstrated above.

DISCUSSION

Two phenol sulphotransferase genes *STP* and *STM* have been localised to adjacent intervals within the larger interval *FRA16E*-CY180A on chromosome 16 which contains markers known to flank the gene for Batten disease. Both *STP* and *STM* can be considered as candidate genes for *CLN3*. Haplotype studies suggest that *CLN3* is closest to markers *D16S298* and *D16S299*, which map to the interval *FRA16E*-CY12. *STP*, which maps within this interval, at a distance of less than 200 Kb proximal to *D16S298* (20)(I. Järvelä, personal communication) is therefore a more likely candidate for *CLN3* based on its positional localisation. Both *STP* and *STM* are expressed in brain neurons (9) and immunolocalisation using antibodies which recognise both forms show staining in CNS neurons (4). Since Batten disease is a neurodegenerative disease, with massive cortical death, both *STP* and *STM* are candidates based on biochemical function and tissue expression. Efforts to sequence both *STP* and *STM* from Batten patients are underway to identify changes in these genes (21). If mutations are present in the coding regions of either gene in affected individuals this would strongly suggest that either *STP* or *STM* is the genetic locus responsible for Batten disease.

Mutations affecting the catalytic activities of one of these sulphotransferases might account for the pathogenesis of Batten disease. If *STP* or *STM* are the cause of Batten Disease, then the sulphation of selected compounds including mono-, di-, and tri-cyclic phenols (22), catecholamines (e.g. dopamine and norepinephrine), phenolic amines, tyrosine containing peptides (e.g. enkephalin and cholecystokinin), or as yet unidentified substrates may be reduced in *CLN3*

patients. The resultant accumulation of these unsulphated compounds in neurons and other tissues during early development and childhood may produce the pathological effects of Batten disease.

The finding that two highly related genes, *STP* and *STM*, map to adjacent hybrid intervals suggests that an ancient gene duplication may have occurred from a single ancestral phenol sulphotransferase gene to generate at least two loci that subsequently diverged in sequence. Other intra-chromosomal duplications in the proximal region of the short arm of 16 have been proposed to account for the distribution of immunoglobulin V_H segments (23). The presence of a number of highly related genes or gene families may therefore characterise this region of chromosome 16. Efforts to characterise the genomic arrangement of the sulphotransferase genes in 16p12.1-11.2 are in progress.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health, the Children's Brain Diseases Foundation, the Batten Disease Support and Research Association, the Southwest Foundation Forum, Los Alamos LDRD, and the Division of Energy in the USA and the Medical Research Council in the UK. The authors acknowledge the technical assistance of R. D. Obermoeller.

REFERENCES

1. Dooley, T. P., Obermoeller, R. D., Leiter, E. H., Chapman, H. D., Falany, C. N., Deng, Z. and Siciliano, M. J. (1993) Genomics 18, 440-443.
2. Callen, D. F., Doggett, N. A., Stallings, R. L., Chen, L. Z., Whitmore, S. A., Lane, S. A., Nancarrow, J. K., Apostolou, S., Thompson, A. D., Lapsys, N. M., Eyre, H. J., Baker, E. G., Shen, Y., Holman, K., Phillips, H., Richards, R. I. and Sutherland, G. R. (1992) Genomics 13, 1178-1185.
3. Falany, C. N. (1991) Trends Pharmacol Sci 2, 255-259.
4. Zou, J., Pentney, R. and Roth, J. A. (1990) J Neurochem 55, 1154-1158.
5. Campbell, N. R. C., Van Loon, J. and Weinshilboum, R. M. (1987) Biochem Pharmacol 36, 1435-1446.
6. Falany, C. N., Vazquez, M. E., Heroux, J. A. and Roth, J. A. (1990) Arch Biochem Biophys 278, 312-318.
7. Wilborn, T. W., Comer, K. A., Dooley, T. P., Reardon, I. M., Heinrikson, R. L. and Falany, C. N. (1993) Mol Pharmacol 43, 70-77.
8. Zhu, X., Veronese, M. E., Sanson, L. N. and McManus, M. E. (1993) Biochem Biophys Res Commun 192, 671-676.
9. Zhu, X., Veronese, M. E., Bernard, C. C. A., Sansom, L. N. and McManus, M. E. (1993) Biochem Biophys Res Commun 195, 120-127.
10. Wood, T. C., Aksoy, I. A., Aksoy, S. and Weinshilboum, R. M. (1994) Biochem Biophys Res Commun 198, 1119-1127.
11. Santavuori, P. (1988) Brain Dev 10, 80-83.
12. Mitchison, H. M., Taschner, P. E. M., O'Rawe, A. M., De Vos, N., Phillips, H. A., Thompson, A. D., Kozman, H. M., Haines, J. L., Schlumpf, K., D'Arigo, K., Boustany, R.-M. N., Callen, D. F., Breuning, M. H., Gardiner, R. M., Mole, S. E. and Lerner, T. J. (1994) Genomics 22, 465-468.
13. Mitchison, H. M., Thompson, A. D., Mulley, J. C., Kozman, H. M., Richards, R. I., Callen, D. F., Stallings, R. L., Doggett, N. A., Attwood, J., McKay, Y. R., Sutherland, G. R. and Gardiner, R. M. (1993) Genomics 16, 455-460.
14. Lerner, T. J., Boustany, R.-M. N., MacCormack, K., Gleitsman, J., Schlumpf, K., Breakefield, X. O., Gusella, J. F. and Haines, J. L. (1994) Am J Hum Genet 54, 88-94.
15. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor).
16. Shen, Y., Kozman, H., M., Thompson, A., Phillips, H. A., Holman, K., Nancarrow, J., Lane, S., Chen, L.-Z., Apostolou, S., Doggett, N. A., Callen, D. F., Mulley, J. C., Sutherland, G. R. and Richards, R. I. (1994) Genomics 22, 68-76.

17. Lerner, T., Wright, G., Leverone, B., Dackowski, W., Shook, D., Anderson, M. A., Klinger, K., D., C. and Landes, G. (1992) Mamm Genome 3, 92-100.
18. Stallings, R. L., Doggett, N. A., Callen, D., Apostolou, S., Chen, L. Z., Nancarrow, J. K., Whitmore, S. A., Harris, P., Mitchison, H., Breuning, M., Saris, J. J., Fickett, J., Cinkosky, M., Torney, D. C., Hildebrand, C. E. and Moyzis, R. K. (1992) Genomics 13, 1031-1039.
19. Kozman, H. M., Keith, T. P., Donis-Keller, H., White, R. L., Weissenbach, J., Dean, M., Vergnaud, G., Kidd, K., Gusella, J., Royle, N. J., Sutherland, G. R. and Mulley, J. C. (1994) Genomics, In press.
20. Jarvela, I. E., Mitchison, H. M., Callen, D. F., Lerner, T. J., Doggett, N. A., Taschner, P. E. M., Gardiner, R. M. and Mole, S. E. (1995) Am J Med Genetics, submitted.
21. Munroe, P. B., Mitchison, H. M., Dooley, T. P., Gardiner, R. M. and Mole, S. E. (1995) Am J Med Genet, submitted.
22. Parker, M. H., McCann, D. J. and Mangold, J. B. (1994) Arch Biochem Biophys 310, 325-331.
23. Tomlinson, I. M., Cook, G. P., Carter, N. C., Elaswarapu, R., Smith, S., Walter, G., Buluwela, L., Rabbits, T. H. and Winter, G. (1994) Hum Mol Genet 3, 853-860.