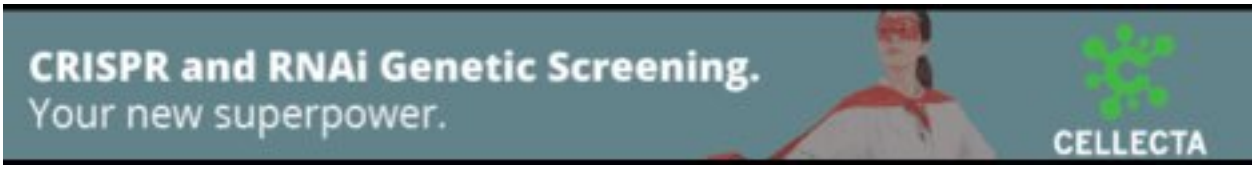
Isolation and genomic structure of a human homolog of the yeast periodic tryptophan protein 2 (PWP2) gene mapping to 21q22.3.

R G Lafrenière, D L Rochefort, N Chrétien, et al.

Genome Res. 1996 6: 1216-1226
Access the most recent version at doi:10.1101/gr.6.12.1216



LETTER

Isolation and Genomic Structure of a Human Homolog of the Yeast Periodic Tryptophan Protein 2 (PWP2) Gene Mapping to 21q22.3

Ronald G. Lafrenière,[1,4] Daniel L. Rochefort,[1] Nathalie Chrétien,[1] Catherine E. Neville,[2] Robert G. Korneluk,[2] Lin Zuo,[3] Yalin Wei,[3] Jay Lichter,[3] and Guy A. Rouleau[1]

[1]Centre for Research in Neuroscience, McGill University, and Department of Neurology, Montreal General Hospital Research Institute, Montreal, Canada H3G 1A4; [2]DNA Sequencing Core Facility, Canadian Genetic Diseases Network, Children's Hospital of Eastern Ontario, Ottawa, Canada K1H 8L1; [3]Sequana Therapeutics, Inc., La Jolla, California 92037

As part of efforts to identify candidate genes for diseases mapping to the 21q22.3 region, we have assembled a 770-kb cosmid and BAC contig containing eight tightly linked markers. These cosmids and BACs were restriction mapped using eight rare cutting enzymes, with the goal of identifying CpG-rich islands. One such island was identified by the clustering of *Not*I, *Eag*I, *Sst*II, and *Bss*HII sites, and corresponded to the *Not*I linking clone LJ104 described previously. A 7.6-kb *Hin*dIII fragment containing this CpG-rich island was subcloned and partially sequenced. A homology search using the sequence obtained from either side of the *Not*I site identified an expressed sequence tag with homology to the yeast periodic tryptophan protein 2 (PWP2). Several cDNAs corresponding to the human PWP2 gene were identified and partially sequenced. Northern blot analysis revealed a 3.3-kb transcript that was well expressed in all tissues tested. A cDNA consensus of 3157 bp was obtained, and an open reading frame potentially encoding 919 amino acid residues was identified. The predicted protein shows 42% identity and 57% similarity at the amino acid level to the yeast PWP2 protein, which is a member of the WD-repeat containing superfamily, and potentially encodes a G-protein beta subunit. The PWP2 gene is split into 21 exons, ranging in size from 53 to 516 bp, and spans an estimated 25 kb. The gene is transcribed in a 21cen→21qter direction, with its 5′ end mapping ~195 kb proximal to the 5′ end of the phosphofructokinase-liver isoform gene. Four single base-pair polymorphisms were identified using single-stranded conformation polymorphism analysis. Possible functions of the protein based on homology to other members of the WD-repeat-containing family are discussed.

[The sequence data described in this paper have been submited to GenBank under accession nos. U56085–U56089.]

The technique of positional cloning is used to identify and screen potential candidate genes based on their chromosomal colocalization with that of a known disease. Identification of CpG-rich islands in genomic fragments as markers for housekeeping genes is one method that can be used to isolate new expressed sequences from a cloned candidate region. CpG-rich islands were first described more than 12 years ago as clusters of unmethylated CpG dinucleotides and identified as HpaII (recognition sequence CCGG) tiny fragments, or HTF islands (Bird 1986). They are usually defined as regions larger than 200 bp with a G + C content above 50% and an observed/expected CpG abundance ratio above 0.6 (Gardiner-Garden and Frommer 1987). There is a wealth of data suggesting that CpG-rich islands correlate highly with the 5′ ends of housekeeping genes, and are associated to a lesser degree with other regions of genes with tissue-specific expression patterns (Larsen et al. 1992). It is therefore a useful strategy to identify these CpG-rich islands as markers for housekeeping genes in a cloned candidate region.

Several diseases have been mapped to the dis-

[4]Corresponding author.
E-MAIL bhgr@musicb.mcgill.ca; FAX 514-934-8265.

6:1216–1226 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

tal tip of human chromosome 21. These include progressive myoclonus epilepsy type 1, which is now known to be caused by mutations in the cystatin B gene (Pennacchio et al. 1996), autoimmune polyglandular disease type 1 (APECED) (Aaltonen et al. 1994), 21q22.3-linked holoprosencephaly (HPE1) (Muenke et al. 1995), a possible vulnerability locus for bipolar affective disorder (Straub et al. 1994; Gurling et al. 1995; Detera-Wadleigh et al. 1996), and a form of autosomal recessive childhood-onset nonsyndromic deafness (DFNB8) (Bonné-Tamir et al. 1996; Veske et al. 1996). As a means of identifying candidate genes for some of these disorders, we have constructed a cosmid and BAC contig spanning an estimated 770 kb, part of which has been published previously (Lafrenière et al. 1995). Subcloning and partial sequencing of potential CpG-rich islands mapping within this contig have identified a human homolog of the yeast periodic tryptophan protein 2 (PWP2) mapping to this region. We present here the cDNA sequence and exon/intron structure of the gene, document its expression pattern, screen the gene for sequence variations by single-stranded conformation polymorphism analysis, discuss the possible function of the protein based on its homology with beta-transducin-like proteins, and evaluate whether the PWP2 gene may be considered a good candidate gene for any of the diseases that have been mapped to the region.

RESULTS

Identification and Sequence of PWP2 cDNAs

To search for the genes defective in inherited disorders linked to 21q22.3, we assembled a cosmid and BAC contig spanning an estimated 770 kb, part of which has been published previously (Lafrenière et al. 1995). Restriction analysis of these cosmids with eight rare cutting enzymes (*Bss*HII, *Eag*I, *Mlu*I, *Nru*I, *Sal*I, *Sfi*I, *Sst*II, and *Not*I) identified, among others, a potential CpG-rich island within cosmid Q84H11, containing *Bss*HII, *Not*I, *Eag*I, and *Sst*II sites clustered within a 1.5-kb region. The *Not*I site corresponded to that found within the *Not*I linking clone LJ104 (D21S1460, see Fig. 1A,B). A 7.6-kb *Hin*dIII fragment containing this *Not*I site was subcloned from the cosmid, and further subcloned to sequence either side of the *Not*I site. A BLASTn homology search using the sequence generated identified an expressed sequence tag (GenBank accession no. H52729) potentially encoding a protein homologous to the

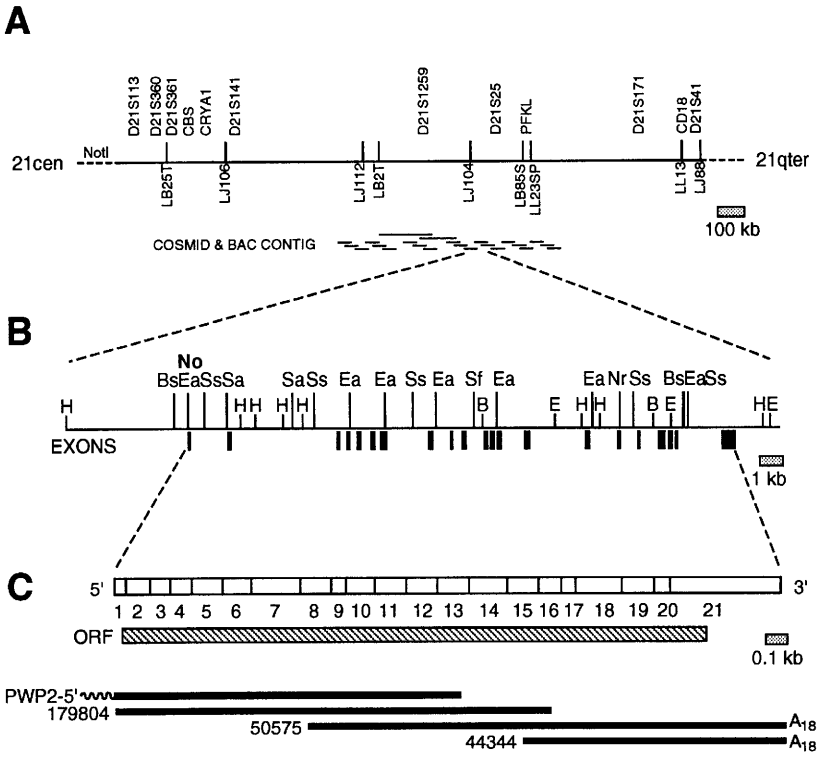


Figure 1 Physical map of the PWP2 gene and surrounding region. (*A*) A portion of the previously published (Ichikawa et al. 1993) *Not*I restriction map of 21q22.3. NotI linking clones are indicated below the horizontal line, reference markers are shown above the line. A minimally overlapping set of 21 cosmid and 2 BAC clones isolated using a walking strategy as described (Lafrenière et al. 1995) are shown. (*B*) Restriction map and genomic structure of the PWP2 gene. Restriction enzymes are given above the map (Abbreviations: H, *Hin*dIII; E, *Eco*RI; B, *Bam*HI; Bs, *Bss*HII; Ea, *Eag*I; Nr, *Nru*I; Sa, *Sal*I; Sf, *Sfi*I; Ss, *Sst*II; No, *Not*I). Black boxes below the map represent the 21 exons of the PWP2 gene. The *Not*I site corresponding to LJ104 is in bold. (*C*) Schematic representation of the consensus cDNA sequence divided by exon. The hatched box represents the full ORF as defined in Fig. 2. The relative overlaps of the four cDNA clones (PWP2-5′, 179804, 50575, and 44344) are shown. Scales for *A, B,* and *C* are shown as shaded boxes in the *lower right* corner of each.

yeast PWP2 protein (SWISS-PROT accession no. S44226). A tBLASTn search of the human dbEST using the yeast PWP2 protein sequence identified 3 cDNA clones that were available from the Integrated Molecular Analysis of Genomes and Their Expression (IMAGE) consortium EST sequencing project (Lennon et al. 1996). Clones 44344, 50575, and 179804 were purchased from Research Genetics. These clones could be hybridized to Southern blots of cosmid DNA digested with *Eco*RI, *Hin*dIII, and *Bam*HI, verifying that they mapped within the contig (data not shown). Clone 179804 was used to screen an adult blood cDNA library. One additional cDNA (PWP2-5′) was identified (see Fig. 1C).

Partial sequences from these cDNA clones were assembled into a cDNA consensus sequence of 3157 bp (Fig. 2). One large open reading frame of 2757 bp, located between the putative start ATG codon at bp 36–38 and the TAG stop codon at bp 2793–2795, encoded a putative protein of 919 amino acid residues (102 kD). However, no in-frame stop codons were located upstream of the putative start ATG, presumably because of the high G + C content of the first exon and 5′ flanking region. The partial 5′ and complete 3′ untranslated regions were 35 and 362 bp in length, respectively. A poly(A)$_{18}$ tail was found at the 3′ end of clones 50575 and 44344, with the potential polyadenylation signals CATAAA (bp 3106–3111) or GATAAA (bp 3117–3122) presumably being used. The poly(A) tail in clone 50575 started at bp 3137, whereas that in clone 44344 started after bp 3157.

A tBLASTn homology search of the hPWP2 protein sequence against the nonredundant DNA data base identified a homologous mouse cDNA fragment (GenBank accession no. U14418) presumably isolated as part of a chimeric clone for the murine GABA-benzodiazepine receptor beta-1 subunit mRNA (Kamatchi et al. 1995), because it is encoded on the opposite strand and corresponds to an internal fragment of the murine PWP2 homolog. Additional human ESTs, an EST of a putative *Caenorhabditis elegans* PWP2 homolog (DBJ accession no. D67118), and an EST of a putative homolog from the fungus *Pisolithus* (GenBank accession no. L41718), as well as a related human EST (GenBank accession no. R69020), were also identified using tBLASTn against dbEST. A BLASTp homology search conducted against the nonredundant protein data base using the entire human predicted protein sequence gave significant scores to the yeast PWP2 protein (SWISS PROT P25635, P = 7.3e-271), three closely linked hypothetical ORFs (PIR S19471, S19469, and S19472, P = 1.0e-128 to 6.0e-30) corresponding to the yeast PWP2 sequence with two frameshifts, and lesser scores to a number of proteins in the beta-transducin family, in particular a beta-transducin-like protein from the fungus *Podospora anserina* (GenBank accession no. L28125) and a protein kinase of unknown function from the bacterium *Thermomonospora curvata* (GenBank accession no. U23820). The human and yeast PWP2 proteins share 42% identical and 57% similar amino acid residues over the entire length of the protein, and their sequence alignment is shown in Figure 3.

A BLOCKS search using the predicted protein sequence gave significant scores to BL00678 (99.66th percentile), the signature of the WD-repeat, a motif of ~40 amino acids in length, bracketed by the dipeptides Gly-His (GH) and Trp-Asp (WD) (Neer et al. 1994). At least four well-defined WD repeats were identified within the predicted sequence (solid boxes in Fig. 3). In addition, several degenerate copies of the WD repeat were found (dashed boxes in Fig. 3). Five leucine residues (circled in Fig. 3) form a putative leucine zipper motif, but lack the basic domain of classic leucine zipper-containing transcriptional activators. However, this leucine zipper is not conserved in the yeast PWP2 protein beyond the first two leucines, so that its function remains obscure. Both the human and yeast PWP2 proteins contain two strongly acidic domains (wavy lines in Fig. 3). The first located between WD-repeats 1 and 2, and the second at the carboxy-terminal end of either protein.

Expression Pattern of PWP2 mRNA

Expression analysis using a human multiple tissue Northern blot (Clontech) showed a single transcript estimated at 3.3 kb and expressed well in all tissues tested [heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas (data not shown)]. The size of the transcript agreed well with the length of the consensus sequence derived from the isolated cDNAs suggesting that a full-length or near full-length cDNA had been isolated. The transcript could also be PCR-amplified from cDNA isolated from lymphoblastoid cell lines.

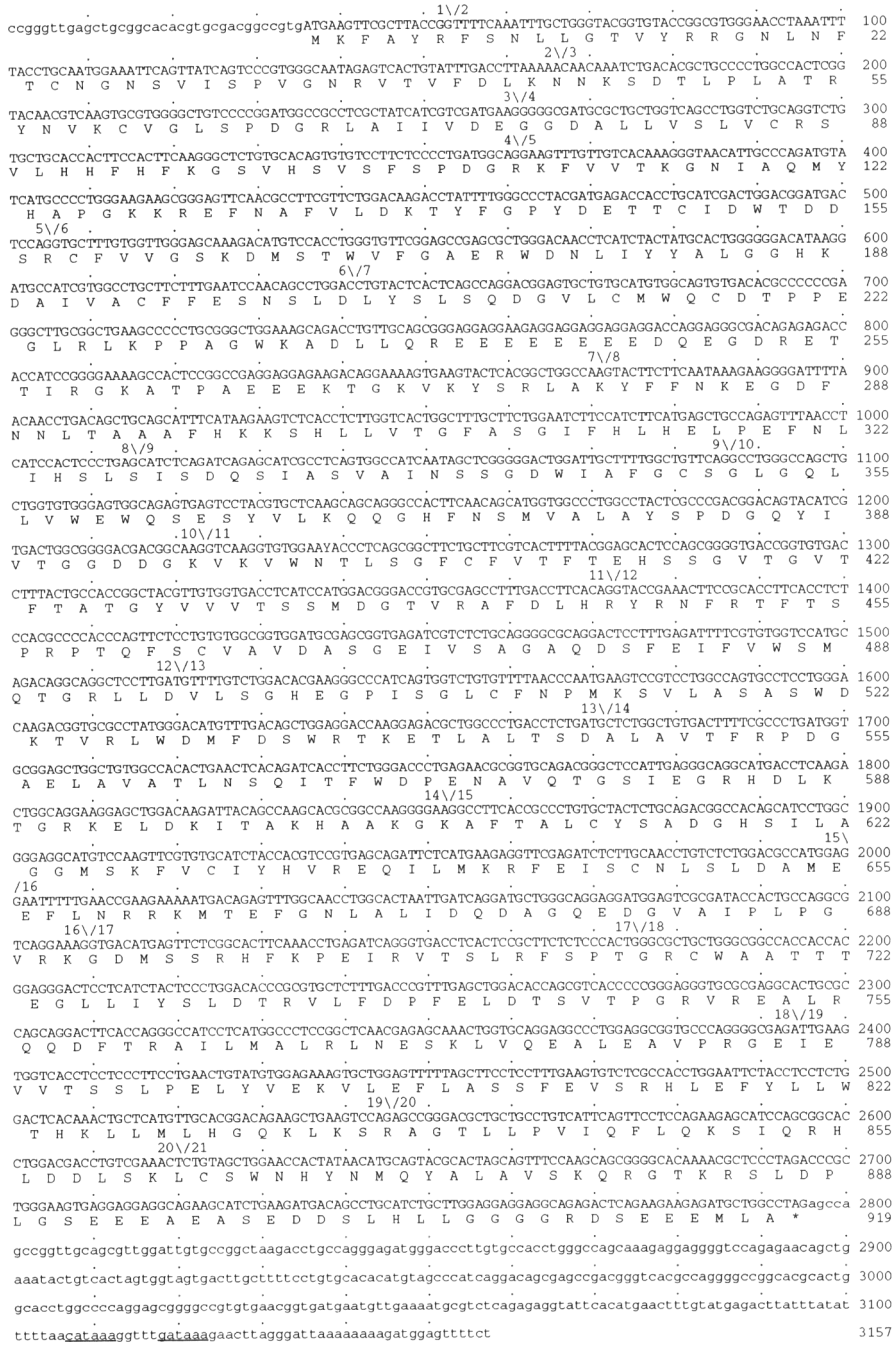


Figure 2 The cDNA and predicted protein sequence of the PWP2 gene. The 3157 bp of cDNA determined from sequence obtained from the four cDNA fragments shown in Fig. 1C. Where polymorphisms occur, the more common allele is shown (refer to Table 3). Potential polyadenylation signals (CATAAA or GATAAA) are underlined. The positions of 20 introns are indicated by (\/) symbols. The GenBank accession no. is U56085.

Genomic Structure of PWP2 Gene

Clones 179804 and 50575 were hybridized to Southern blots of *Eco*RI-, *Hin*dIII- and *Bam*HI-digested DNA of cosmids isolated from the region, and suggested that the PWP2 gene spanned ~25 kb and was split into at least eight exons. Cosmid Q84H11, containing the entire PWP2 gene, was partially sequenced to determine the exon–intron structure of the gene. Approximately 23 kb of genomic sequence were obtained in four separate contigs, enabling us to identify the complete structure of the gene. PWP2 is split into 21 exons (Table 1). All intronic junctions conformed to the GT...AG rule established for nuclear pre-mRNA introns (Shapiro and Senepathy 1987). Based on restriction mapping analysis, and hybridization of cDNA fragments to our cosmid contig, the PWP2 gene is transcribed in the 21cen→qter direction, with its 5′ end mapping ~195 kb proximal to the 5′ end of the phosphofructokinase liver isoform (PFKL) gene. The PWP2 gene is transcribed in the same direction as the next upstream gene (GT334, R.G. Lafrenière, Z. Kibar, D. Rochefort, F.-Y. Han, E.A. Fon, M.-P. Dubé, X. Kang, S. Baird, R. Korneluk, E. Andermann, J. Rommens, and G.A. Rouleau, in prep.) and the next downstream gene (GT335, Lafrenière et al. 1996) with ~1.5 kb between each of these genes (data not shown).

Polymorphism Screening by SSCP Analysis

We screened for base-pair polymorphisms in this gene using PCR primers derived from both the cDNA and genomic sequences (Table 2). Four polymorphic variants were identified and se-

Figure 3 Sequence homology between human and yeast PWP2 proteins. Alignment of the human PWP2 protein sequence (hPWP2) to the *Saccharomyces cerevisiae* PWP2 protein (yPWP2, SWISS-PROT accession no. P25635 is shown). Identical amino acid residues are indicated with (|) whereas functionally similar residues are indicated with (+). The well-defined WD repeats are boxed in solid lines, the less well-defined WD repeats in dashed lines. Acidic domains are indicated by wavy lines. Leucines forming a potential leucine zipper are circled.

quenced (Fig. 4; and Table 3). A polymorphic single base-pair change in exon 2 resulted in the nonconservative amino acid substitution Asn25Asp. The polymorphic base-pair substitutions in exons 9, 13, and 15 did not affect the amino acid sequence of the protein. All four polymorphisms created or destroyed restriction enzyme recognition sites, and could thus be scored as PCR-amplifiable restriction fragment length polymorphisms (RFLPs) (data not shown). Allele frequencies for these polymorphisms were determined using a collection of 18 to 45 normal individuals, and are given in Table 3.

DISCUSSION

We describe here the isolation, characterization, and genomic structure of the human PWP2 gene, which maps to 21q22.3. The human PWP2 protein shows highest homology with the previously described yeast PWP2 (R. Shafaatian, M.A. Payton, and J.D. Reis, unpubl.), and probably represents its functional human homolog. Although its exact function has yet to be determined, the yPWP2 protein has been proposed to have a role in the process of bud-site selection and cell separation. However, its structurally defining characteristic is the presence of at least four and as many as seven WD repeats, identifying it as a member of the beta-transducin superfamily of WD repeat-containing proteins. These proteins are usually involved in regulatory rather than enzymatic pathways. Several of the known WD-repeat proteins form heteromultimeric complexes, and the role of the WD-repeat domain has been postulated to be in the assembly of protein complexes involved in various aspects of cellular function (Neer et al. 1994). Although both belong to the WD-repeat family, the PWP2 protein does not share strong sequence homology to the yeast PWP1 protein (Duronio et al. 1992). Based on its strong homology to the beta-transducin family of proteins, the PWP2 protein might function as a G-protein beta subunit in a signal transduction pathway.

The PWP2 gene consists of 21 exons and

Table 1. Exon/Intron Junctions of the Human PWP2 Gene

Exon no.	Intron	EXON	Exon size (bp)	EXON	Intron	Intron size (bp)
1		CCGGGTTGAG	53	CGCTTACCGG	gtgagcgcgg	~2000
2	ttcttcacag	TTTTCAAATT	113	ACCTTAAAAA	gtaagtatgt	~4000
3	tctttatcag	CAACAAATCT	95	GTCGATGAAG	gtacttgccc	331
4	cgtcctgcag	GGGGCGATGC	100	CTGATGGCAG	gtaaggggga	346
5	ccttccatag	GAAGTTTGTT	144	ATGACTCCAG	gtgcggcctc	495
6	tcttttccag	GTGCTTTGTG	136	CAGCCTGGAC	gtatgtccct	291
7	cccctt gcag	CTGTACTCAC	230	GGCTGGCCAA	gtaggtctct	1903
8	gtcaattcag	GTACTTCTTC	144	ACTCCCTGAG	gtaagccttt	798
9	ttctctgcag	CATCTCAGAT	71	GGCTGTTCAG	gtttgtcccc	456
10	gtgtgcacag	GCCTGGGCCA	137	CGACGGCAAG	gtaggctcct	824
11	cctgcctcag	GTCAAGGTGT	149	ACCTTCACAG	gtgatgtttt	134
12	ttctctgcag	GTACCGAAAC	148	GCTCCTTGAT	gtaagcaccc	173
13	tctacggtag	GTTTTGTCTG	151	ACCTCTGATG	gtgagcacga	1075
14	ctctgtttag	CTCTGGCTGT	181	CCAAGGGGAA	gtgagtgtca	~2000
15	ctccacccag	GGCCTTCACC	148	CGCCATGGAG	gtgagccgcc	1283
16	gctgttctag	GAATTTTTGA	109	GTCAGGAAAG	gtgagcagag	780
17	ttttccccag	GTGACATGAG	66	TCTCCCACTG	gtgagcactg	968
18	aatttcccag	GGCGCTGCTG	219	AGGGGCGAGA	gtgagttggg	96
19	cctttt gcag	TTGAAGTGGT	151	TGAAGTCCAG	gtagagggtc	154
20	tcccttt cag	AGCCGGGACG	75	TGTCGAAACT	gtacgtgtgg	1970
21	tgattgacag	CTGTAGCTGG	516	CTTAGGGATT	polyA	

Sequence data have been deposited in GenBank (accession nos. U56086–U56089).

spans 25 kb. Based on restriction mapping around this region, the PWP2 gene is closely flanked by other genes. The 3′ end of the GT334 gene (R.G. Lafrenière, Z. Kibar, D. Rochefort, F.-Y. Han, E.A. Fon, M.-P. Dubé, X. Kang, S. Baird, R. Korneluk, E. Andermann, J. Rommens, and G.A. Rouleau, in prep.), also known as the EHOC-1 gene (Yamakawa et al. 1995), lies ~1.5 kb centromeric of the 5′ end of the PWP2 gene. Likewise, the 3′ end of the PWP2 gene lies ~1.5 kb centromeric to the 5′ end of the GT335 gene (Lafrenière et al. 1996). All three genes span a region of ~140 kb, and are transcribed in the 21cen→qter direction. Based on the genomic structures of these three genes, there are 51 exons in this region, giving an average of one exon every 2.7 kb. Further sequencing of this region will reveal whether this high gene density can be extrapolated to the remainder of the distal tip of chromosome 21, which has already been postulated to be gene-rich (Gardiner et al. 1990; Yaspo et al. 1995).

In screening the PWP2 gene for polymorphisms, we have identified and sequenced four single base-pair changes mapping within the coding region of this gene, one of which results in a nonconservative amino acid substitution in the PWP2 protein. These polymorphisms can be used as genetic markers for haplotype analysis, and may be of interest for the eventual functional and mutational studies of the PWP2 protein.

The PWP2 gene is a candidate for inherited disorders that have been linked to the 21q22.3 region. In particular, the PWP2 gene seems an attractive candidate for bipolar affective disorder, a locus for which has been mapped to this region by several groups (Straub et al. 1994; Gurling et al. 1995; Detera-Wadleigh et al. 1996). Several lines of evidence, most notably the inhibitory effects of lithium on G-protein function (Avissar et al. 1988; Jope and Williams 1994), suggest that a possible underlying defect in this complex disorder might involve modification of G protein coupled signal transduction pathways in neurons (Manji et al. 1995a,b). Because the PWP2 protein may act as a G-protein beta subunit, it would be worthwhile to test this gene in patients with this disorder.

The PWP2 gene may also be a good candidate

Table 2. PCR Primers Used for Genomic SSCP Analysis

Exon	Primer 1 (5′–3′)		Primer 2 (5′–3′)		Amplification product size (bp)	PCR annealing temp. (°C)
1	W84:	GCGCGGACCCCCGGGAAGTGTCTC	W85:	TCAGGCGCCGTCCACCGGAAGAGC	150	65
2	W63:	TCTGCCCAGATGCCTCACTTAG	W64:	AGGAAGGCACAGAAGACACACTATCA	294	60
3	W65:	GCGCTACTTGGGGAGGGTGAG	W66:	TGGGGAGCGGATCATTTCAGA	284	57
4	W67:	CTGGGGTAGCTGTGCGGTGGTGAC	W68:	ATGCACGGCCGAACCCTCCTG	182	60
5	W69:	TTAGGGCCATTGCTGGTCTTGAAT	W70:	GCTCCCCCGAAGCCTCTGAG	223	60
6	W71:	CAAGCCCTGAGGAGGTGCCATCTC	W72:	CAGCCTGGGGCCTCCCTGAACTA	243	60
7a	W59:	GGGCTGGGGTGACCCTGTGCT	W60:	CCTCGGCCGGAGTGGCTTTTC	222	60
7b	W57:	TGCGGGCTGGAAAGCAGACCT	W58:	CCCCGCCACAGCCCACCAC	185	60
8	W73:	TCTGCGCTCCTCCTCACAACTCTT	W74:	TGGCTGGGCCCTCAGATTCA	253	60
9	W13:	GGCAGGCTGGCCTCCCACTTA	W14:	TGTAGCCGGGCAGCACTGGTC	174	55
10	W56:	GTGGGCCGGGCCAGTCTGA	W29:	GCCCGGGTCAAGCTGTGAACA	316	60
11	W30:	GGCCGAAGCCGTGGCACAC	W42:	CCAATCCGGAGCAAAAACATC	231	60
12	W43:	GGGACCCTGGCACGTGACTCT	W44:	AGCCCCAGCCCAGCCCTCAG	219	60
13	W48:	TCGGCCGCCACCTCCACTCTTC	W75:	TGGTCCCCAGGCCGCTGCTC	237	60
14	W76:	GTCTCCAGCCCAAGGGTGGTTG	w77:	GAGGCCCGATGCTGACACTCACT	284	57
15	W55:	GGCCACCGTAAGTCCATGTCTCCT	W19:	AGGGGCCCGTGCACTTAGATACA	324	60
16	W40:	AGGCGGCCAGGCCATCAGAA	W62:	AGGGGCATGAATGCGTGGTCA	228	57
17	W78:	AGCCGGGAGACCGCAGGATGA	W79:	GCCCACCGGCACCAAAAG	191	57
18	W80:	AAAAATGTTGGGAGGAATGAGC	W17:	ACCCCGGCGCCCGACACC	309	60
19	W15:	CTTCCCCTGGTTTGAGTTGGTG	W61:	ACTGCGGGGGAGACCCTCTAC	227	60
20	W37:	TGCAGCATGAAGTTGAGCCCACTG	W81:	CCCAGGCCCTGCACCCACAC	166	57
21	W45:	AGCCCCTGCTGCTTTTCTCA	W12:	GGTCTTAGCCGGGCACAATCC	318	60

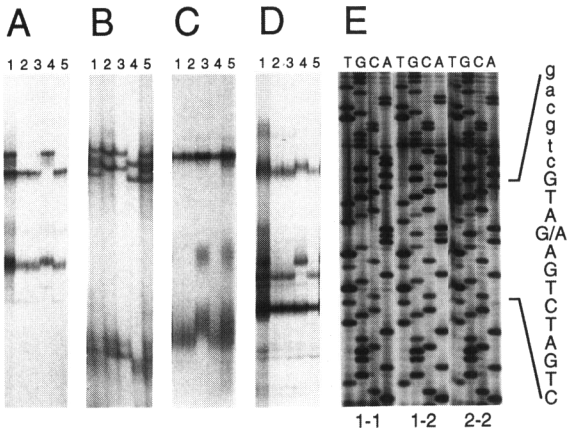


Figure 4 SSCP variants in 5 individuals are shown for polymorphisms in exons 2 (*A*), 9 (*B*), 13 (*C*), and 15 (*D*). Allele designations are (ht) heterozygotes; (hm) homozygotes. For *A,* lanes *2,3,5* are hm for allele 1; lane *1,* ht for alleles *1,2;* lane *4,* hm for allele 2. For *B,* lanes *2,3,* hm for allele 1; lanes *1,5,* ht for alleles *1,2;* lane *4,* hm for allele 2. For *C,* lanes *1,2,4,* hm for allele 1; lane *5,* ht for alleles *1,2;* lane *3,* hm for allele 2. For *D,* lanes *2,3,5,* hm for allele 1; lane *1,* ht for alleles *1,2;* lane *4,* hm for allele 2. SSCP gel conditions were as described in Methods, with condition (1) being used for *C,D,* condition (2) used for *A,* and condition (3) used for *B.* An example of sequence analysis is shown in *E* for the variant in exon 9, with the sequence of the minus strand read from bottom to top in an hm for allele 1 (1-1), an ht for alleles 1 and 2 (1-2), and a hm for allele 2 (2-2). The upper- and lowercase letters represent exonic and intronic sequences, respectively.

for 21q22.3-linked holoprosencephaly. This is based on homology to another WD repeat-containing protein, LIS-1, which constitutes the acetylhydrolase subunit of brain platelet-activating factor (Hattori et al. 1994). Haploinsufficiency for the LIS-1 gene is known to cause Miller–Dieker lissencephaly syndrome, a severe brain malformation manifested by development of a smooth cerebral surface and probably attributable to abnormal neuronal migration (Reiner et al. 1993). In cases of 21q22.3-linked holoprosencephaly (HPE1), brain malformations resulting in the formation of a single ventricle and/or microcephaly are observed. These malformations are thought to occur from impaired midline cleavage of the embryonic forebrain. Cytologically detectable terminal deletions of chromosome 21 have been observed in some HPE patients, and it has been postulated that haploinsufficiency of a gene in 21q22.3 is the genetic basis of the disease. Based on molecular analysis of DNA markers in 3 HPE patients with terminal deletions of chromosome 21, the HPE1 gene has been mapped to a 5-Mb region at the tip of chromosome 21 (Muenke et al. 1995), a region that includes the PWP2 gene. Of the handful of haploinsufficiency syndromes whose causative gene has been identified, most involve proteins that probably form multimeric complexes with defined stoichiometries (Fisher and Scambler 1994). Thus the PWP2 protein, containing WD repeats potentially involved in assembly of multimeric complexes, can be regarded as an excellent candidate for an haploinsufficiency syndrome. Further studies, such as the generation of haploinsufficient PWP2 transgenic knockout mice, will be required to assess whether the PWP2 gene is involved in this disorder.

Based on its physical localization, the PWP2 gene also remains a candidate for two other diseases that have been mapped to the 21q22.3 region. Autoimmune polyglandular disease type 1, also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) involves varying degrees of dysfunction of the parathyroid glands, adrenal cortex, gonads, pancreatic B cells, thyroid gland, and gastric parietal cells (Aaltonen et al. 1994) because of the presence of antibodies directed against adrenal proteins. And finally, a form of nonsyndromic childhood-onset deafness (DFNB8) has been linked to 21q22.3 (Bonné-Tamir et al. 1996; Veske et al. 1996). If the PWP2 protein acts in a signaling pathway, then it may be considered a candidate gene for any of the diseases mentioned above. Of course, the PWP2 gene may also play a part in the complex phenotype observed in Down syndrome patients. Further studies will be required to determine whether mutations in the PWP2 gene lead to any of these disorders, and to address the function of the protein.

METHODS

Subcloning and Sequencing

Cosmid and plasmid DNA were prepared using a modified alkaline-lysis protocol (Birnboim et al. 1979) after 16 hr of growth at 37°C in TB broth (Tartof and Hobbs 1987) supplemented with kanamycin (50 µg/ml) or ampicillin (60 µg/ml), respectively. *Eco*RI- or *Hin*dIII-digested cosmid fragments were subcloned into pBluescript II SK. Double-stranded plasmid DNA was sequenced manually using a Sequenase kit (U.S. Biochemical) or with an Applied Biosystems Inc. model 373A fluorescence sequencer available from the Network of Centres of Excellence (Canadian Ge-

Table 3. Polymorphisms in the PWP2 Gene

Exon	Primer pair	Position (codon)	Allele	Polymorphic sequence	Amino acid	Restriction site	Allele frequencies
2	W63 + W64	108 (25)	1	TGC **A**AT GGA	Asn	+ BsrDI	25/38 = 0.66
			2	TGC **G**AT GGA	Asp	− BsrDI	13/38 = 0.34
9	W13 + W14	1019 (328)	1	AGC AT**C** TCA	Ile	+ DdeI	67/90 = 0.74
			2	AGC AT**T** TCA	Ile	− DdeI	23/90 = 0.26
13	W48 + W75	1607 (524)	1	AAG AC**G** GTG	Thr	− TspRI	37/48 = 0.77
			2	AAG AC**A** GTG	Thr	+ TspRI	11/48 = 0.23
15	W55 + W19	1986 (651)	1	TCT **C**TG GAC	Leu	+ BsmAI	41/54 = 0.76
			2	TCT **T**TG GAC	Leu	− BsmAI	13/54 = 0.24

See Table 2 for primer sequences. Position refers to the position of the polymorphic basepair (or codon) in reference to the consensus cDNA sequence shown in Fig. 2.

netic Diseases Network) Sequencing Core Facility in Ottawa. Partial sequencing of cosmid Q84H11 was done using the following shotgun sequencing strategy. Briefly, cosmid DNA was sonicated and DNA fragments were repaired using nuclease BAL-31 and T4 DNA polymerase. DNA fragments of 0.8–2.2 kb were size-fractionated by agarose gel electrophoresis and ligated into the pUC9 vector. Inserts of the plasmid clones were amplified by PCR and sequenced using standard ABI dye-primer chemistry. Sequence assembly was performed using the ABI AutoAssembler.

DNA Sequence Analysis

DNA sequence data were assembled and manipulated using DNASTAR Inc. (Madison, WI) software. Sequence homology searches were conducted against nonredundant DNA and protein sequence databases using the BLAST (Altschul et al. 1990) E-mail server at the National Center for Biotechnology Information (Bethesda, MD) and the BLOCKS (Henikoff and Henikoff 1991) E-mail searcher at the Fred Hutchison Center for Cancer Research (Seattle, WA).

Northern Blot

Northern blot containing poly(A)$^+$ RNA derived from various human tissues was purchased (no. 7760-1, Clontech, Palo Alto, CA) and hybridized and washed according to conditions recommended by the manufacturer. The filter was wrapped in Saran, and exposed to BioMax MR film (Kodak) for one week at −80°C using a DuPont intensifying screen.

SSCP Analysis

Genomic DNA and RNA were isolated from EPM1 patients and controls as described previously (Cochius et al. 1993). Genomic DNA fragments were amplified by PCR in a 96-well plate using a model PTC-100 thermocycler (MJ Research, Inc.) and the following conditions: 94°C for 5 min, 35 cycles of 94°C for 30 sec, X°C for 30 sec, and 72°C for 30 sec (where X is the temperature given in Table 2), followed by a final elongation at 72°C for 7 min. Samples were diluted with formamide-containing loading buffer and denatured for 10 minutes at 90°C prior to electrophoresis under the following two sets of conditions: (1) 0.5× MDE gel in 0.6× TBE plus 5% glycerol run at 25°C for 16 hr at 3 W per gel, and (2) 9.5% polyacrylamide gel in 1× TBE run at 4°C for 16 hr at 14 W per gel. A third set of conditions were used for some of the samples: 6% polyacrylamide gel in 1× TBE run at 4°C for 16 hr at 21 W per gel. An undenatured control was loaded alongside denatured samples. Gels were then dried and exposed to BioMax MR X-ray film (Kodak) for 16 hr.

Sequencing of PCR Products

For sequencing, one primer was synthesized with 5′ biotinylation. The fragment to be sequenced was PCR-amplified using a biotinylated and nonbiotinylated primer in four 100 µl volumes. The pooled 400 µl reaction was then ethanol precipitated, the pellet was dried then resuspended in 30 µl of Tris–EDTA (10 mM:1 mM), and separated on a 0.8% LMP agarose (Life Technologies, Inc.) gel in 1× TAE, stained with ethidium bromide and the appropriate-sized band excised. The gel slice was then heated for 15 min at 75°C, allowed to cool to 50°C, and digested with 2 U beta-agarase (New England Biolabs) at 37°C for 2 hr. The DNA was then precipitated by adding 1/10 volume of 2 M potassium acetate and 2 volumes of 95% ethanol. The biotinylated strand was then separated from the nonbiotinylated strand using streptavidin-conjugated magnetic beads (Dynabeads M-280 Streptavidin, Dynal, Inc.). Separated strands were then sequenced using the Sequenase kit (U.S. Biochemical) and [^{35}S]dATP (Amersham).

ACKNOWLEDGMENTS

The authors would like to thank the families and clinicians

that are actively participating in our study of EPM1, P.J. de Jong for the gift of the gridded chromosome 21-specific cosmid library, I. Lopes-Cendes for helpful discussions, and Z. Kibar, C. Marineau, and L. Lebuis for technical assistance. R.G.L. is supported by a studentship from the Medical Research Council (MRC) of Canada. G.A.R. is supported by the MRC and the Fonds de Recherche en Santé du Québec.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aaltonen, J., P. Björses, L. Sandkuijl, J. Perheentupa, and L. Peltonen. 1994. An autosomal locus causing autoimmune disease: Autoimmune polyglandular disease type 1 assigned to chromosome 21. *Nature Genet.* **8:** 83–87.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Avissar, S., G. Schreiber, A. Danon, and R.H. Belmaker. 1988. Lithium inhibits adrenergic and cholinergic increases in GTP binding in rat cortex. *Nature* **331:** 440–442.

Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321:** 209–213.

Birnboim, H.C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7:** 1513–1523.

Bonné-Tamir, B., A.L. DeStefano, C.E. Briggs, R. Adair, B. Franklyn, S. Weiss, M. Korostishevsky, M. Frydman, C.T. Baldwin, and L.A. Farrer. 1996. Linkage of congenital recessive deafness (gene DFNB10) to chromosome 21q22.3. *Am. J. Hum. Genet.* **58:** 1254–1259.

Cochius, J.I., D.A. Figlewicz, R. Kälviäinen, U. Nousiainen, K. Farrell, G. Patry, B. Söderfeldt, M. Frydman, P. Lerman, F. Andermann, E. Andermann, and G.A. Rouleau. 1993. Unverricht-Lundborg disease: Absence of nonallelic genetic heterogeneity. *Ann. Neurol.* **34:** 739–741.

Detera-Wadleigh, S.D., J.A. Badner, L.R. Goldin, W.H. Berrettini, A.R. Sanders, D.Y. Rollins, G. Turner, T. Moses, H. Haerian, D. Muniec, J.I. Nurnberger, Jr., and E.S. Gershon. 1996. Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q. *Am. J. Hum. Genet.* **58:** 1279–1285.

Duronio, R.J., J.I. Gordon, and M.S. Boguski. 1992. Comparative analysis of the beta transducin family with identification of several new members including PWP1, a nonessential gene of Saccharomyces cerevisiae that is divergently transcribed from NMT1. *Proteins* **13:** 41–56.

Fisher, E. and P. Scambler. 1994. Human haploinsufficiency—One for sorrow, two for joy. *Nature Genet.* **7:** 5–7.

Gardiner, K., M. Horisberger, J. Kraus, U. Tantravahi, J. Korenberg, V. Rao, S. Reddy, and D. Patterson. 1990. Analysis of human chromosome 21: Correlation of physical and cytogenetic maps; gene and CpG island distributions. *EMBO J.* **9:** 25–34.

Gardiner-Garden, M. and M. Frommer. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Gurling, H., C. Smyth, G. Kalsi, E. Moloney, L. Rifkin, J. O'Neill, P. Murphy, D. Curtis, H. Petursson, and J. Brynjolfsson. 1995. Linkage findings in bipolar disorder. *Nature Genet.* **10:** 8–9.

Hattori, M., H. Adachi, M. Tsujimoto, H. Aral, and K. Inoue. 1994. Miller-Dieker lissencephaly gene encodes a subunit of brain platelet-activating factor. *Nature* **370:** 216–218.

Henikoff, S. and J.G. Henikoff. 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* **19:** 6565–6572.

Ichikawa, H., F. Hosoda, Y. Arai, K. Shimizu, M. Ohira, and M. Ohki. 1993. A NotI restriction map of the entire long arm of human chromosome 21. *Nature Genet.* **4:** 361–366.

Jope, R.S. and M.B. Williams. 1994. Lithium and brain signal transduction systems. *Biochem. Pharmacol.* **47:** 429–441.

Kamatchi, G.L., P. Kofuji, J.B. Wang, J.C. Fernando, Z. Liu, J.R.J. Mathura, and D.R. Burt. 1995. GABAA receptor beta 1, beta 2, and beta 3 subunits: Comparisons in DBA/2J and C57BL/6J mice. *Biochim. Biophys. Acta* **1261:** 134–142.

Lafrenière, R.G., P.J. de Jong, and G.A. Rouleau. 1995. A 405-kb cosmid contig and HindIII restriction map of the progressive myoclonus epilepsy type 1 (EPM1) candidate region in 21q22.3. *Genomics* **29:** 288–290.

Lafrenière, R.G., D.L. Rochefort, Z. Kibar, E.A. Fon, F.-Y. Han, J. Cochius, X. Kang, S. Baird, R.G. Korneluk, E. Andermann, J.M. Rommens, and G.A. Rouleau. 1996. Isolation and characterization of GT335, a novel human gene conserved in *E. coli* and mapping to 21q22.3. *Genomics* (in press).

Larsen, F., G. Gundersen, R. Lopez, and H. Prydz. 1992. CpG islands as gene markers in the human genome. *Genomics* **13:** 1095–1107.

Lennon, G.G., C. Auffray, M. Polymeropoulos, and M.B. Soares. 1996. The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33:** 151–152.

Manji, H.K., G. Chen, H. Shimon, J.K. Hsiao, W.Z.

Potter, and R.H. Belmaker. 1995a. Guanine nucleotide-binding proteins in bipolar affective disorder: Effects of long-term lithium treatment. *Arch. Gen. Psychiatry* **52:** 135–144.

Manji, H.K., W.Z. Potter, and R.H. Lenox. 1995b. Signal transduction pathways: Molecular targets for lithium's actions. *Arch. Gen. Psychiatry* **52:** 531–543.

Muenke, M., L.J. Bone, H.F. Mitchell, I. Hart, K. Walton, K. Hall-Johnson, E.F. Ippel, J. Dietz-Band, K. Kvaloy, C.-M. Fan, M. Tessier-Lavigne, and D. Patterson. 1995. Physical mapping of the holoprosencephaly critical region in 21q22.3, exclusion of SIM2 as a candidate gene for holoprosencephaly, and mapping of SIM2 to a region of chromosome 21 important for Down Syndrome. *Am. J. Hum. Genet.* **57:** 1074–1079.

Neer, E.J., C.J. Schmidt, R. Nambudripad, and T.F. Smith. 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371:** 297–300.

Pennacchio, L.A., A.-E. Lehesjoki, N.E. Stone, V.L. Willour, K. Virtaneva, J. Miao, E. D'Amato, L. Ramirez, M. Faham, M. Koskiniemi, J.A. Warrington, R. Norio, A. de la Chapelle, D.R. Cox, and R.M. Myers. 1996. Mutations in the gene encoding Cystatin B in progressive myoclonus epilepsy (EPM1). *Science* **271:** 1731–1734.

Reiner, O., R. Carrozzo, Y. Shen, M. Wehnert, F. Faustinella, W.B. Dobyns, C.T. Caskey, and D.H. Ledbetter. 1993. Isolation of a Miller-Dieker lissencephaly gene containing G protein B-subunit-like repeats. *Nature* **364:** 717–721.

Shapiro, M.B. and P. Senepathy. 1987. RNA splice junctions of different classes of eukaryotes: Statistics and functional implications in gene expression. *Nucleic Acids Res.* **15:** 7155–7175.

Straub, R.E., T. Lehner, Y. Luo, J.E. Loth, W. Shao, L. Sharpe, J.R. Alexander, K. Das, R. Simon, R.R. Fieve, B. Lerer, J. Endicott, J. Ott, T.C. Gilliam, and M. Baron. 1994. A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. *Nature Genet.* **8:** 291–296.

Tartof, K.D. and C.A. Hobbs. 1987. Improved media for growing plasmid and cosmid clones. *Bethesda Res. Lab. Focus* **9:** 12.

Veske, A., R. Oehlmann, F. Younus, A. Mohyuddin, B. Muller-Myhsok, S. Qasim Mehdi, and A. Gal. 1996. Autosomal recessive non-syndromic deafness locus (DFNB8) maps on chromosome 21q22 in a large consanguineous kindred from Pakistan. *Hum. Mol. Genet.* **5:** 165–168.

Yamakawa, K., S. Mitchell, R. Hubert, X. Chen -N., S. Colbern, Y.-K. Huo, C. Gadomski, U.-J. Kim, and J.R. Korenberg. 1995. Isolation and characterization of a candidate gene for progressive myoclonus epilepsy on 21q22.3. *Hum. Mol. Genet.* **4:** 709–716.

Yaspo, M.-L., L. Gellen, R. Mott, B. Korn, D. Nizetic, A.-M. Poustka, and H. Lehrach. 1995. Model for a transcript map of human chromosome 21: Isolation of new coding sequences from exon and enriched cDNA libraries. *Hum. Mol. Genet.* **4:** 1291–1304.

Received July 24, 1996; accepted in revised form September 9, 1996.