Molecular Psychiatry (2003) 8, 83–89
© 2003 Nature Publishing Group All rights reserved 1359-4184/03 $25.00
www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

A novel CpG-associated brain-expressed candidate gene for chromosome 18q-linked bipolar disorder

D Goossens[1], S Van Gestel[1], S Claes[1], P De Rijk[1], D Souery[2], I Massat[2], D Van den Bossche[1], H Backhovens[1], J Mendlewicz[2], C Van Broeckhoven[1] and J Del-Favero[1]

[1]*Department of Molecular Genetics, Flanders Interuniversity Institute for Biotechnology (VIB), University of Antwerp (UIA), Antwerpen, Belgium;* [2]*Department of Psychiatry, Erasme Hospital, University of Brussels (ULB), Brussels, Belgium*

We previously identified 18q21–q22 as a candidate region for bipolar (BP) disorder and constructed a yeast artificial chromosome (YAC) contig map. Here we identified three potential CpG islands using CCG/CGG YAC fragmentation. Analysis of available genomic sequences using bioinformatic tools identified an exon of 3639 bp downstream of a CpG island of 1.2 kb containing a putative transcription initiation site. The exon contained an open reading frame coding for 1212 amino acids with significant homology to the SART-2 protein; weaker homology was found with a series of sulphotransferases. Alignment of cDNA sequences of corresponding ESTs and RT-PCR sequencing predicted a transcript of 9.5 kb which was confirmed by Northern blot analysis. The transcript was expressed in different brain areas as well as in multiple other peripheral tissues. We performed an extensive mutation analysis in 113 BP patients. A total of nine single nucleotide polymorphisms (SNPs) were identified. Five SNPs predicted an amino acid change, of which two were present in BP patients but not in 163 control individuals.
Molecular Psychiatry (2003) **8,** 83–89. doi:10.1038/sj.mp.4001190

Keywords: affective disorders; gene cloning; mutation analysis; sequencing; SNP

Introduction

Bipolar (BP) disorder is a severe psychiatric condition affecting about 1% of the population. The clinical signs of BP disorder are alternating episodes of depression and mania (BP type I) or hypomania (BP type II). Genetic studies have identified several potential chromosomal loci, of which chromosome 18 is of particular interest. Berrettini *et al*[1] first reported linkage to this chromosome in a pericentromeric region on 18p. Subsequently, several studies of chromosome 18 have been published that support a pericentromeric locus,[2–4] an 18q21–q22 locus,[2–6] both loci[7,8] or other chromosome 18 loci.[9–11] Negative linkage studies with chromosome 18 have also been reported.[12] In further studies we refined the linkage region at 18q21–q22, identified by De Bruyn *et al* in a multiplex Belgian BP family,[6] to an 8.9 cM region between D18S68 and D18S979 at 18q21–q22 and constructed a physical map using yeast artificial chromosomes (YACs).[13] Our candidate region at chromosome 18q21–q22 identified in the Belgian study overlaps with that reported by Stine *et al.*[2] Several studies have described anticipation in families transmitting BP disorder,[14–16] suggesting the involvement of trinucleotide repeat expansions (TREs). Support for the involvement of CAG/CTG repeat expansions in BP disorder was obtained by the use of the repeat expansion detection (RED) method.[17] Genetic association of BP disorder with expanded CAG/CTG repeats was subsequently confirmed in several independent studies.[18–21] However, more recently, it was shown that nearly 90% of the expanded CAG/CTG repeats detected by RED may be explained by long alleles at the polymorphic CTG18.1 and ERDA1 CAG/CTG loci present also in the normal population.[22–26] Also, we have excluded the involvement of CAG/CTG repeats from within the 18q21–q22 candidate region for BP disorder,[27] using a CAG repeat YAC fragmentation method.[28] Together these data make it unlikely that CAG/CTG repeats would explain the anticipation in BP families and the association in BP cases/control populations.[26]

Considering that a number of diseases caused by an expansion of a CCG/CGG repeat show anticipation (reviewed by Margolis *et al*[16]), we originally aimed at identifying CCG/CGG repeats from within the 18q21–q22 region using a variant of the CAG repeat YAC fragmentation method. Instead, the method was more efficient in identifying CpG islands than CCG/CGG repeats. CpG islands are considered as anchor points for genes since they are frequently observed upstream of genes in their regulatory region and are expected to play a major role in transcription control of gene expression. Using this method, we identified three CpG islands and showed that at least one is associated

Correspondence: Prof. Dr C Van Broeckhoven, PhD DSc, Department of Molecular Genetics, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium.
E-mail: Christine.Vanbroeckhoven@uia.ua.ac.be
Received 8 October 2001; revised 15 April 2002; accepted 18 April 2002

with a novel brain-expressed gene, which was designated *C18orf4* by the HUGO Gene Nomenclature Committee. Since it is located in the candidate region 18q21–q22 for BP disorder identified in several independent studies,[2,5,6] we examined this gene in depth. We identified several single-nucleotide polymorphisms(SNPs) of which only two were observed in BP cases. Therefore, *C18orf4* is an interesting positional candidate gene that deserves additional studies to determine its potential role in the pathogenesis of chromosome 18q21–q22-linked BP disorder.

Materials and methods

Subjects

The pedigree and the clinical diagnoses in family MAD31 were described in detail elsewhere.[13] Case–control studies were performed on a Belgian sample of 113 BP patients and 160 controls matched for age, sex and ethnicity.[29] All individuals were interviewed using the SADS-LA after written informed consent.

Triplet repeat isolation

CCG/CGG YAC fragmentation vectors were constructed by blunt-end cloning of $(CCG)_{10}$/$(CGG)_{10}$ adapters into the *Sph*I site of the previously described pDV1 vector.[30] Sequencing determined that fragmentation vectors pDVCCG and pDVCGG have the adapter sequence in a 5′-$(CCG)_{10}$-3′ and a 5′-$(CGG)_{10}$-3′ orientation, respectively. Using these vectors, CCG/CGG repeats and flanking sequences were isolated by YAC fragmentation as described.[30]

cDNA sequencing

I.M. A.G.E. Consortium [LLNL] cDNA Clones:[31] IMAGp998A136826Q2, IMAGp998A154307Q2, IMAGp998B194346Q2, IMAGp998D126826Q2, IMAGp998D193628Q2, IMAGp998F131866Q2, IMAGp998H201815Q2, IMAGp998K235214Q2, IMAGp998L153967Q2, IMAGp998N06839Q2 were obtained from RZPD (Deutsches Ressourcenzentrum für Genomforschung GmbH, Berlin-Charlottenburg, Germany). Single colony cultures were grown and plasmid DNA was prepared by the Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA). DNA sequencing was performed by cycle sequencing using the ABI PRISM Dye Terminator Cycle Sequencing Core Kit (Perkin-Elmer, Foster, CA, USA) according to the supplier's protocol and analysed on an ABI PRISM 3700 DNA Analyser (Perkin-Elmer, Foster, CA, USA).

For the RT-PCR reactions, mRNA from *SHSY-5Y* cells was prepared using the μMACS mRNA Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany). After DNAseI treatment (Promega, Madison, WI, USA), the RT reaction was primed with oligo (dT) primers and performed with the Superscript Preamplification System for First Strand cDNA synthesis (GibcoBRL, NV Life Technologies, Merelbeke, Belgium). First-strand cDNA was used in long-range PCR reactions with TaKaRa LA *Taq* (Takara Shuzo Co., Otsu, Shiga, Japan). PCR products were reamplified with nested primers and sequenced as described above.

Expression analysis

The Human 12-lane Multiple Tissue Northern (MTN) Blot, the Human 12-lane MTN BLOT II and the Human Brain MTN Blot IV (Clontech, Palo Alto, CA, USA) were used for radioactive hybridisation in accompanying ExpressHyb solution according to the instructions of the manufacturer. Hybridisation signals of the Human 12-lane MTN Blots were detected with a phosphoimager (PhosphoImager SI, Molecular Dynamics, Amersham Pharmacia Biotech, Buckinghamshire, UK). A ZOO blot was prepared by digesting 10 μg genomic DNA to completion with *Hind*III, separating it on a 1% agarose gel and performing a Southern blot. A PCR product containing the ORF of *C18orf4* was radioactively labelled and hybridised at 65°C.

Mutation analysis

Overlapping PCR products of approximately 600 bp were generated from genomic DNA and cycle sequenced in 113 BP cases on a PE 3700 sequencer. Sequence variations were identified using the computer software program Polyphred[32] followed by visual inspection. Potential SNPs were confirmed in the BP cases by pyrosequencing using an SNP specific primer[33] on a PSQ96 apparatus (Pyrosequencing AP, Uppsala, Sweden) (Table 1), and analysed in 160 controls.

Table 1 C18orf4 SNPs

Nucleotide position	*Amino acid change*	*Frequency minor allele*		*SNP primer for pyrosequencing*
		Patients (n=113)	*Controls (n=160)*	
g.−546A>G		0.12	0.09	TGTCATTCCTCTCTCAACTT
g.859G>A	V287I	0.01	0.01	GCCAGAAAAACATACTGT
g.2017C>T	P673S	0.46	0.46	ATGCAATTCTTGTGATTGTG
g.2189A>G	Y730C	0.01	0.00	CCACCGAATCCCAG
g.2824C>T	P942S	0.02	0.01	GCCAGTTTACCCCTATTG
g.3339A>G	I1113M	0.01	0.00	GCAGCAGCCTTGAGA

Results

CCG/CGG YAC fragmentation

CCG/CGG YAC fragmentation was applied to YACs 961h9, 766f12 and 907e1 spanning 18q21.3–q22 BP candidate region.[13,27] Size determination by pulsed field gel electrophoresis (PFGE) and Southern blot hybridisation resulted in 33 sets of equally sized fragmented YAC clones. Sequencing of 112 fragmented YAC ends identified seven (out of 33) sets of fragmented YACs with identical end sequences resulting from a specific homologous recombination. One set (CCG7) was the result of fragmentation in a $(CGG)_6$ repeat. This $(CGG)_6$ repeat is part of a combined $(CGG)_6(CAG)_6$ repeat located in the 5′UTR of *CAP2* (GenBank accession number L40377). We had already shown that this repeat is polymorphic but not expanded in BP cases or associated with BP disorder.[27] A second set (CCG6) contained a $(CCG)_2$ repeat. The other five end sequence did not contain CCG/CGG repeats.

In-depth analysis showed that three (CCG3, GenBank accession number AF480432; CCG4, GenBank accession number AF480433; and CCG6, GenBank accession number AF480434) of the seven end sequences had a high CG (70–80%) and CpG (15–20 CpGs in 200 bp) content. Primer pairs flanking these potential CpG island were used to determine their position on the YAC contig (Figure 1). BLASTN analysis[34] resulted for CCG4 and CCG6 but not CCG3, in hits with sequences of RPCI-11 BACs. CCG4 gave a hit in a contig of 27 150 bp of the working draft sequence of RPCI-11 BAC 29O13 (GenBank accession number AC022662). CCG6 is part of the complete sequence of RPCI-11 BAC 793J2 (GenBank accession number AC009802).

In silico identification of C18orf4

To identify genes associated with the potential CpG islands CCG4 and CCG6, their surrounding sequences were analysed using bioinformatic tools. Hence the 27 150 bp sequence contig of BAC 29O13 and the complete sequence of BAC 793J2 were sent for analysis to the Rummage High-Throughput Sequence Annotation Server[35] (http://gen 100.imb-jena.de/rummage/index.html). First, LCP[36] and CPG[37] recognised CpG islands containing CCG4 and CCG6 of 1.2 and 0.4 kb, respectively, confirming their potential role as CpG islands.

In the next step, the exon prediction programs Grail[38] and Genscan[39] both predicted the presence of a 3639 bp exon, 1.5 kb downstream of the 1.2 kb CpG island containing CCG4. The predicted exon contains an open reading frame (ORF), which starts at an ATG codon with an almost perfect Kozak sequence (AtCATGG), and ends with a TAA stop codon, and codes for 1212 amino acids. Other predicted features are a transcription start site (TSS) in the CpG island, 2352 bp upstream of the ORF (g.−2352; score 76.6 by Proscan[40]) and three polyadenylation signals at positions g.6671, g.6886 and g.8003 (respective scores of 4.79, 3.83 and 4.94 by PloyAH[41]) (Figure 2a). A BLASTN search of dbEST sequences revealed significant homology ($\geqslant$97%) to 24 human ESTs (Figure 2b). TBLASTX[42] searches with the ORF of the Genbank nonredundant (nr) database showed extensive homology at the protein level with SART-2 (Genbank accession number NP_037484), a squamous cell carcinoma antigen recognized by T-cells.[43] Weaker homology was found with a series of sulphotransferases. Analysis of the 1212-long amino acid sequence of the translated ORF by SMART (Simple Modular Architecture Research Tool, V3.1)[44] did not result in any known domains apart from a cleavable signal peptide at position p.1–20 and two transmembrane segments at positions p.771–791 and p.880–820 (Figure 2a). Interpro reported no significant hits, although BLASTP[42] of the Prodom database showed homology between the ORF and the chon-

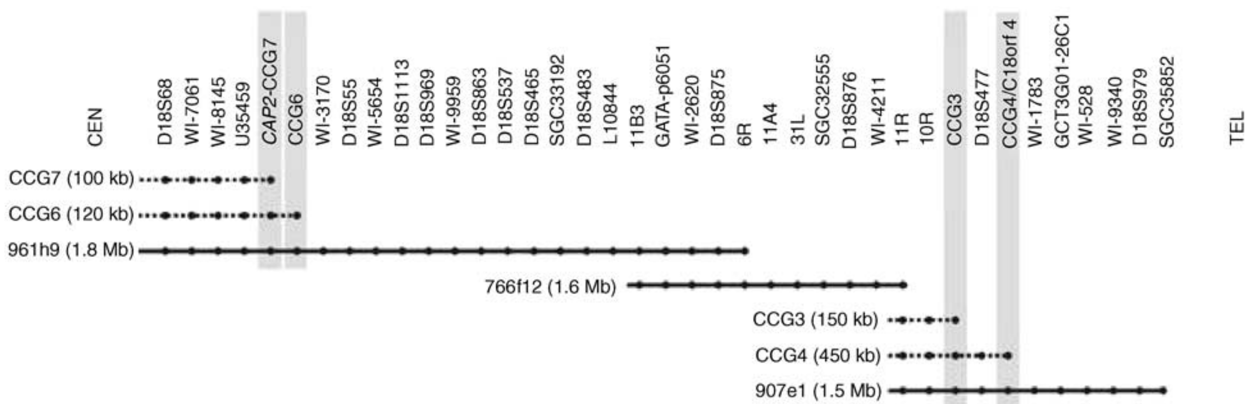

Figure 1 Minimal YAC tiling path of the 18q21.3–q22 BP candidate region.[13] The YACs are represented by solid lines and the CCG/CGG fragmentation products by dotted lines. YAC sizes given between brackets were estimated by PFGE analysis. Solid circles indicate positive STS/STR hits. Shaded boxes highlight the CCG/CGG repeat in *CAP2,* the 3 CpG islands CCG6, CCG3 and CCG4 isolated by YAC fragmentation, and *C18orf4.*

droitin-6-sulphotransferase domain (Prodom accession number PD042460).

Structural organisation of C18orf4
Based on the BLASTN EST hits, I.M.A.G.E. Consortium [LLNL] cDNA clones[31] were requested and sequenced. The cDNA sequences aligned with the genomic sequence in the presumed 5′UTR, the ORF and the presumed 3′UTR, indicating that these sequences are transcribed (Figure 2c). Alignment of the 5′UTR IMAGp998B194346Q2 cDNA sequence with the genomic sequence showed that an 865 bp fragment was missing in the cDNA. A detailed analysis of the flanking genomic sequences revealed the presence of consensus acceptor and donor splice sites, indicating that this fragment was most likely an intron. Clone IMAGp998D193628Q2 in the 3′UTR also missed a fragment of 1.9 kb when compared to the genomic sequence, but consensus splice sites flanking this sequence were absent. The cDNA clones IMAGp998L153967Q2 and IMAGp998A136826Q2 contained sequences from within the missing fragment (Figure 2c) while EST AA442543 was located entirely within the gap (Figure 2b). Two clones, IMAGp998D193628Q2 and IMAGp998A136826Q2, terminated exactly at the predicted polyadenylation signal at g.8003. Sequences of clones IMAGp998A154307Q2, IMAGp998D126826Q2 and IMAGp998F131866Q2 did not align with the genomic sequence and were not analysed further. Since cDNA clone sequencing did not result in a continuous transcript, primers were designed for RT-PCR experiments. Sequencing of overlapping RT-PCR products obtained from cultured SHSY-5Y cells confirmed the presence of a transcript of at least 9 kb, containing the ORF linked to the presumed 5′ and 3′ sequences (Figure 2d). Also, the 5′UTR intron of 865 bp was confirmed while the 3′UTR was extended until the predicted polyadenylation signal at g.8003. The novel gene was designated *C18orf4* and the consensus cDNA sequence (Genbank accession number AF480435) was defined based on the genomic sequence since no sequence differences were observed between the available genomic sequence and those obtained by cDNA and RT-PCR sequencing.

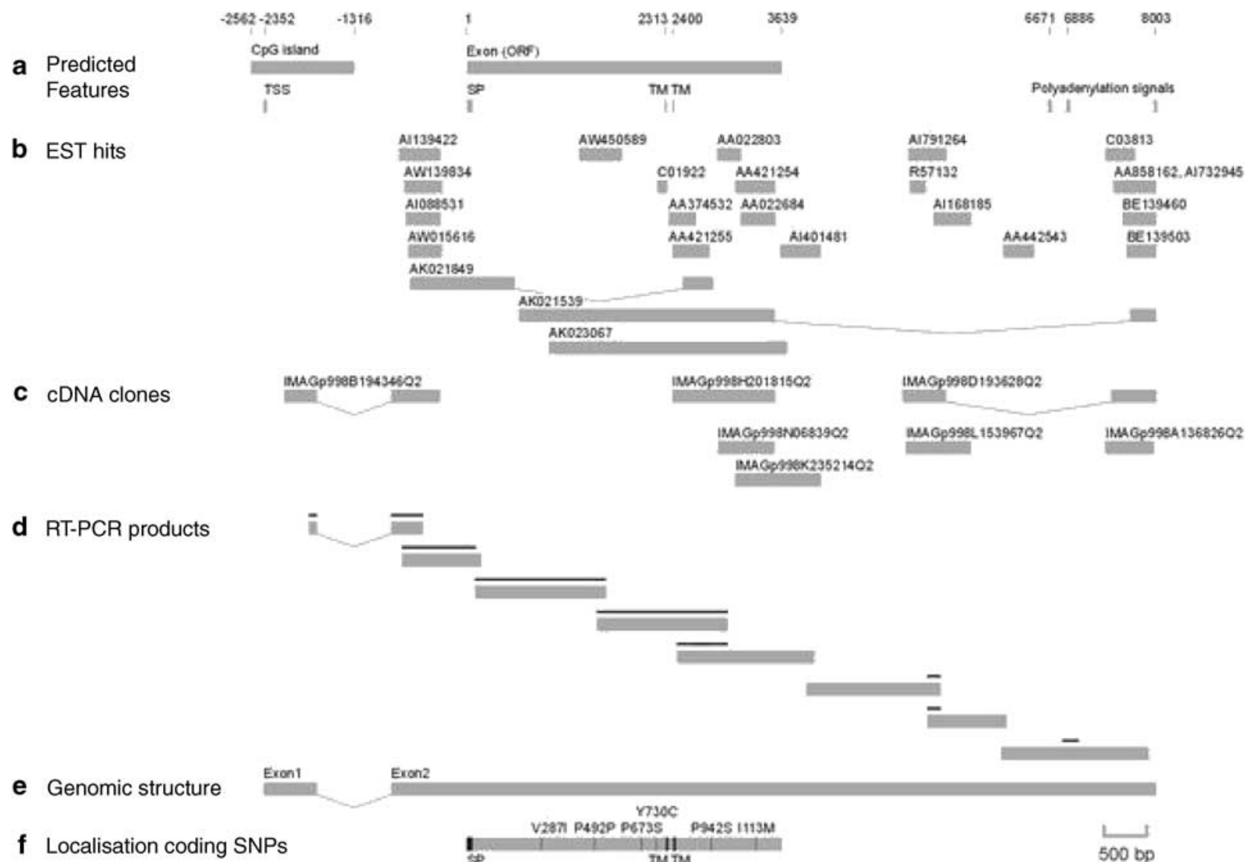

Figure 2 Feature map of *C18orf4*. (a) Predicted features by bioinformatics tools: the CpG island, the ORF or exon, the transcription start site (TSS), the three polyadenylation signals and the positions of the signal peptide (SP) and transmembrane segments (TM).(b) Alignment of EST hits. ESTs are named with their Genbank accession number.(c) Alignment of cDNA clones. I.M.A.G.E. Consortium [LLNL] cDNA Clones[31] are named with their RZPD clone ID.(d) RT-PCR products. The grey bars represent the RT-PCR product and the thin black lines represent the sequences obtained using nested primers. (e) Genomic structure.(f) Localisation of coding SNPs.

Expression pattern of C18orf4
We investigated the expression profile of *C18orf4* by hybridising Northern blots of multiple tissues (Figure 3). This showed the presence of an ~9.5 kb transcript in lung, placenta, small intestine, liver, kidney, skeletal muscle, heart, brain, uterus, trachea, thyroid, stomach, spinal cord, prostate, mammary gland, lymph node, bladder and adrenal gland. The other investigated tissues had very weak or no expression of the transcript (thymus, colon, peripheral blood leukocyte, spleen and bone marrow). After normalisation to β-actin, the brain showed the highest relative expression (data not shown). More detailed information on the regional expression in brain was obtained by Northern blot hybridisation showing expression in all investigated areas (amygdala, caudate nucleus, corpus callosum, hippocampus, substantia nigra, thalamus and total brain).

Stringent ZOO blot hybridisation experiments identified homologous sequences in the genomic DNA of dog, pig, mouse, donkey, horse and sheep (Figure 4).

Mutation analysis of C18orf4
A mutation analysis in 113 BP patients including a BPI case of the 18q21–q22-linked BP family MAD31,[13] of the complete ORF and flanking UTRs of *C18orf4* was performed by sequencing overlapping fragments generated by PCR amplification of genomic DNA. Nine SNPs were identified. Three SNPs (g.−298G>A, g.−80G>A, g.1476C>G) occurred only once in the heterozygous state in the BP patients and were not analysed further. Five SNPs predicted an amino acid change (Table 1, Figure 2f). Together with one 5′UTR SNP (g.−546A>G), these five coding SNPs (cSNPs) were analysed by pyrosequencing in 160 control individuals matched for ethnicity, sex and age (Table 1). All SNPs were in Hardy–Weinberg equilibrium. The 5′UTR and three of the cSNPs were present in control individuals and BP cases at near-equal allele frequencies. Also no differences in haplotype distributions were observed between BP patients and controls. The two cSNPs Y730C and I1113M appeared only in BP patients. The Y730C cSNP was hetero-

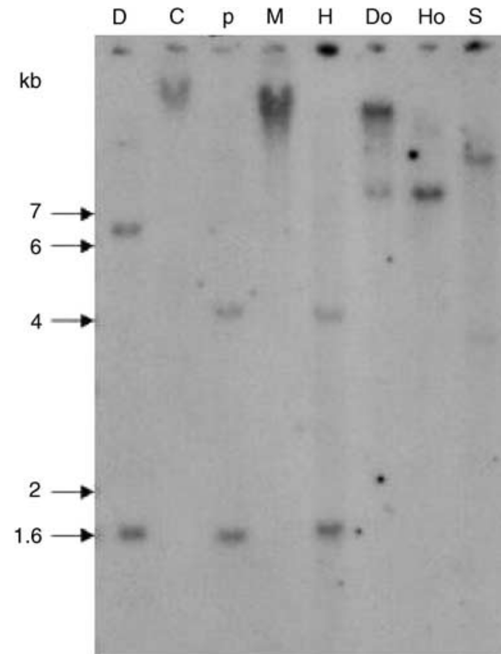

Figure 4 ZOO blot analysis of *C18orf4*. D: dog; C: chicken; P: pig; M: mouse; H: hamster; Do: donkey; Ho: horse; S: sheep.

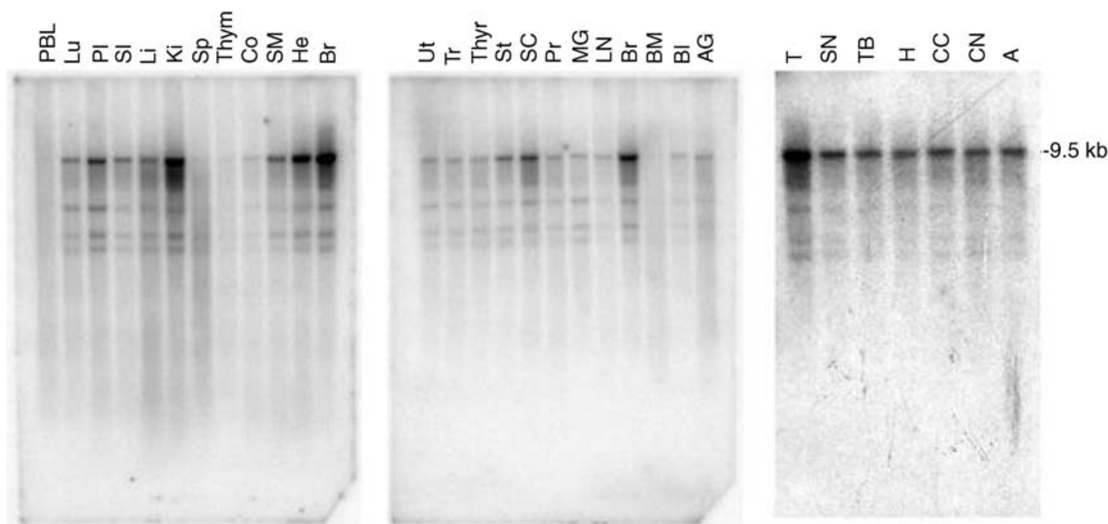

Figure 3 Northern blot analysis of *C18orf4*. A: amygdala; AG: adrenal gland; Bl: bladder; BM: bone marrow; Br: brain; CC: corpus callosum; CN: caudate nucleus; Co: colon; H: hippocampus; He: heart; Ki: kidney; Li: liver; LN: lymph node; Lu: lung; MG: mammary gland; PBL: peripheral blood leukocyte; Pl: placenta; Pr: prostate; SC: spinal cord; SI: small intestine; SM: skeletal muscle; SN: substantia nigra; Sp: spleen; St: stomach; T: thalamus; TB: total brain. Thym: thymus; Thyr: thyroid; Tr: trachea; Ut: uterus. The size of the *C18orf4* transcript was estimated using the size marker present on the Northern blots.

zygous in one BPI patient with a positive family history. The I1113M cSNP was observed twice in the heterozygous state in a sporadic BPI patient and a familial BPII patient. The BPI patient of family MAD31 was homozygous for all SNPs except for P942S. Segregation analysis indicated that the minor allele of P942S was comprised in the risk haplotype segregating in the family (data not shown).

Discussion

We originally aimed at identifying CCG/CGG repeats from within the 18q21–q22 candidate region for BP disorder using YAC fragmentation.[28] However, instead, we isolated three sequences (CCG3, CCG4 and CCG6) that had several characteristics of CpG islands. The CCG4 sequence contained in a 1.2-kb CpG island was associated with a predicted exon of 3639 bp containing an ORF coding for a 1212 amino acid protein. This novel CpG-associated gene was designated *C18orf4*, of which the putative protein contains a predicted cleavable signal peptide and two potential transmembrane domains. This suggests that both the N- and C-termini are on the same side of the membrane in which the protein is embedded. *C18orf4* shows a significant homology with the SART-2 protein, a squamous cell carcinoma antigen recognised by T-cells. Weaker homology was observed with a series of sulphotransferases. A sulphotransferase SULT1A1 that inactivates dopamine and other phenol-containing compounds by sulphation was also identified in a microarray expression study in rats treated with methamphetamines.[45] At the genomic level the gene sequence contained a predicted transcription start site at −2.4 kb within the CpG island, an intron of 865 bp in the 5′UTR and a polyadenylation signal at g.8003. The ORF was entirely contained within exon 2 of 3639 bp. *C18orf4* is highly conserved among different species and is expressed as a 9.5-kb transcript in multiple tissues including all brain regions. Therefore, *C18orf4* can be considered a positional and functional candidate gene for BP disorder linked to 18q21–q22.[2,5,6]

By direct sequencing we identified nine SNPs in *C18orf4* in BP patients. Five SNPs predicted an amino acid change. Two cSNPs (Y730C and I1113M) were present in the heterozygous state only in BP patients. They were, however, not observed in patients of the 18q21–q22-linked BP family MAD31. In this family, only the P942S SNP was observed that segregated with the risk haplotype. Since this SNP was also observed in controls (minor allele frequency 0.01), it is unlikely that this variant *C18orf4* explains the disease aggregation in this family.

In conclusion, our study did not support a role for expanded CCG/CGG repeats from within the 18q21–q22 region in BP disorder. However, we identified within the 18q21–q22 candidate region one novel CpG-associated gene *C18orf4* that is widely expressed, including in the brain. Owing to its localisation on chromosome 18, in a region that we and others previously linked to BP disorder,[13,10] as well as the identification of two cSNPs present only in BP cases, we cannot exclude a potential role for *C18orf4* in BP disorder. The predicted sequence similarities, although suggestive, do not provide substantial clues on its potential function in BP pathogenesis.

Therefore, studies of the *C18orf4* SNPs will be needed in independent BP case/control samples and chromosome 18q21–q22 families to confirm that *C18orf4* is a candidate gene for chromosome 18q21–q22-linked BP disorder.

Websites

Rummage High-Throughput Sequence Annotation Server: http://gen100. imbjena.de/rummage/index. html
PolyAH: http://dot.imgen.bcm.tmc.edu: 9331/gene-finder/gf.html
SMART: http://smart.embl-heidelberg.de
Psort: http://psort.nibb.ac.jp/form2.html
Interpro: http://www.ebi.ac.uk/interpro/
Prodom: http://www.toulouse.inra.fr/prodom.html

Acknowledgements

The work described in this paper was funded in part by the Fund for Scientific Research-Flanders, Belgium (FWO) and the EU-BIOMED Grants CT97-2466 and BMH4-CT97-2307. DG has a PhD fellowship from FWO.

References

1 Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger Jr JI *et al*. Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. *Proc Natl Acad Sci USA* 1991; **91**: 5918–5921.
2 Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C *et al*. Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. *Am J Hum Genet* 1995; **57**: 1384–1394.
3 Maier W, Hallmayer J, Zill P, Bondy B, Lichtermann D, Ackenheil M *et al*. Linkage analysis between pericentrometric markers on chromosome 18 and bipolar disorder: a replication test. *Psychiatry Res* 1995; **59**: 7–15.
4 Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G *et al*. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci USA* 1999; **96**: 5604–5609.
5 McMahon FJ, Hopkins PJ, Xu J, McInnis MG, Shaw S, Cardon L *et al*. Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. *Am J Hum Genet* 1997; **61**: 1397–1404.
6 De Bruyn A, Souery D, Mendelbaum K, Mendlewicz J, Van Broeckhoven C. Linkage analysis of families with bipolar illness and chromosome 18 markers. *Biol Psychiatry* 1996; **39**: 679–688.
7 Nothen MM, Cichon S, Rohleder H, Hemmer S, Franzek E, Fritze J *et al*. Evaluation of linkage of bipolar affective disorder to chromosome 18 in a sample of 57 German families. *Mol Psychiatry* 1999; **4**: 76–84.
8 Turecki G, Grof P, Cavazzoni P, Duffy A, Grof E, Martin R *et al*. Lithium responsive bipolar disorder, unilineality, and chromosome 18: a linkage study. *Am J Med Genet* 1999; **88**: 411–415.
9 Coon H, Jensen S, Hoff M, Holik J, Plaetke R, Reimherr F *et al*. A genome-wide search for genes predisposing to manic-depression,

assuming autosomal dominant inheritance. *Am J Hum Genet* 1993; **52**: 1234–1249.
10 Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P *et al*. Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22–q23. *Nat Genet* 1996; **12**: 436–441.
11 Ewald H, Wang AG, Vang M, Mors O, Nyegaard M, Kruse TA. A haplotype-based study of lithium responding patients with bipolar affective disorder on the Faroe Islands. *Psychiatr Genet* 1999; **9**: 23–34.
12 Van Broeckhoven C, Verheyen GR. Report of the chromosome 18 workshop. *Am J Med Genet* 1999; **88**: 263–270.
13 Verheyen GR, Villafuerte SM, Del-Favero J, Souery D, Mendlewicz J, Van Broeckhoven C *et al*. Genetic refinement and physical mapping of a chromosome 18q candidate region for bipolar disorder. *Eur J Hum Genet* 1999; **7**: 427–434.
14 McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JRJ. Anticipation in bipolar affective disorder. *Am J Hum Genet* 1993; **53**: 385–390.
15 Nylander PO, Engstrom C, Chotai J, Wahlstrom J, Adolfsson R. Anticipation in Swedish families with bipolar affective disorder. *J Med Genet* 1994; **31**: 686–689.
16 Maroglis RL, McInnis MG, Rosenblatt A, Ross CA. Trinucleotide repeat expansion and neuropsychiatric disease. *Arch Gen Psychiatry* 1999; **56**: 1019–1031.
17 Schalling M, Hudson TJ, Buetow KH, Housman DE. Direct detection of novel expanded trinucleotide repeats in the human genome. *Nat Genet* 1993; **4**: 135–139.
18 Lindblad K, Nylander PO, De bruyn A, Souery D, Zander C, Engstrom C *et al*. Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. *Neurobiol Dis* 1995; **2**: 55–62.
19 O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA *et al*. Expanded CAG repeats in schizophrenia and bipolar disorder. *Nat Genet* 1995; **10**: 380–381.
20 O'Donovan MC, Guy C, Craddock N, Bowen T, McKeon P, Macedo A *et al*. Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. *Psychol Med* 1996; **26**: 1145–1153.
21 Oruc L, Lindblad K, Verheyen GR, Ahlberg S, Jakovljevic M, Ivezic S *et al*. CAG repeat expansions in bipolar and unipolar disorders. *Am J Hum Genet* 1997; **60**: 730–732.
22 Lindblad K, Nylander PO, Zander C, Yuan QP, Stahle L, Engstrom C *et al*. Two commonly expanded CAG/CTG repeat loci: involvement in affective disorders? *Mol Psychiatry* 1998; **3**: 405–410.
23 Verheyen GR, Del-Favero J, Mendlewicz J, Lindblad K, Van Zand K, Aalbregtse M *et al*. Molecular interpretation of expanded RED products in bipolar disorder by CAG/CTG repeats located at chromosomes 17q and 18q. *Neurobiol Dis* 1999; **6**: 424–432.
24 Guy CA, Bowen T, Jones I, McCandless F, Owen MJ, Craddock N *et al*. CTG18.1 and ERDA-1 CAG/CTG repeat size in bipolar disorder. *Neurobiol Dis* 1999; **6**: 302–307.
25 Vincent JB, Petronis A, Strong E, Parikh SV, Meltzer HY, Lieberman J *et al*. Analysis of genome-wide CAG/CTG repeats, and at *SEF2-1B* and *ERDA1* in schizophrenia and bipolar affective disorder. *Mol Psychiatry* 1999; **4**: 229–234.
26 Goossens D, Del-Favero J, Van Broeckhoven C. Trinucleotide repeat expansions: do they contribute to bipolar disorder? *Brain Res Bull* 2001; **56**: 243–257.
27 Goossens D, Villafuerte SM, Tissir F, Van Gestel S, Claes S, Souery D *et al*. No evidence for the involvement of CAG/CTG repeats from within 18q21.33–q23 in bipolar disorder. *Eur J Hum Genet* 2000; **8**: 385–388.
28 Del-Favero J, Goossens D, De Jonghe P, Benson K, Michalik A, Van den Bossche D *et al*. Isolation of CAG/CTG repeats from within the chromosome 2p21–p24 locus for autosomal dominant spastic paraplegia (SPG4) by YAC fragmentation. *Hum Genet* 1999; **105**: 217–225.
29 Souery D, Lipp O, Mahieu B, Mendelbaum K, De MV, Van Broeckhoven C *et al*. Association study of bipolar disorder with candidate genes involved in catecholamine neurotransmission: DRD2, DRD3, DAT1, and TH genes. *Am J Med Genet* 1996; **67**: 551–555.
30 Del-Favero J, Goossens D, Van den Bossche D, Van Broeckhoven C. YAC fragmentation with repetitive and single-copy sequences: detailed physical mapping of the presenilin 1 gene on chromosome 14. *Gene* 1999; **229**: 193–201.
31 Lennon G, Auffray C, Polymeropoulos M, Soares MB. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 1996; **33**: 151–152.
32 Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997; **25**: 2745–2751.
33 Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res* 2000; **10**: 1249–1258.
34 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
35 Taudien S, Rump A, Platzer M, Drescher B, Schattevoy R, Gloeckner G *et al*. RUMMAGE—a high-throughput sequence annotation system. *Trends Genet* 2000; **16**: 519–521.
36 Huang X. An algorithm for identifying regions of a DNA sequence that satisfy a content requirement. *Comput Appl Biosci* 1994; **10**: 219–225.
37 Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. *Genomics* 1992; **13**: 1095–1107.
38 Uberbacher EC, Mural RJ. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci USA* 1991; **88**: 11 261–11 265.
39 Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997; **268**: 78–94.
40 Prestridge DS. Predicting Pol II promoter sequences using transcription factor binding sites. *J Mol Biol* 1995; **249**: 923–932.
41 Salamov AA, Solovyev VV. Recognition of 3′-processing sites of human mRNA precursors. *Comput Appl Biosci* 1997; **13**: 23–28.
42 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
43 Nakao M, Shichijo S, Imaizumi T, Inoue Y, Matsunaga K, Yamada A *et al*. Identification of a gene coding for a new squamous cell carcinoma antigen recognized by the CTL. *J Immunol* 2000; **164**: 2565–2574.
44 Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* 2000; **28**: 231–234.
45 Niculescu AB, Segal DS, Kuczenski R, Barrett T, Hauger RL, Kelsoe JR. Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiol Genomics* 2000; **4**: 83–91.