ARTICLE

Detailed transcript map of a 810-kb region at 11p14 involving identification of 10 novel human 3' exons

F Guillemot, C Auffray and MD Devignes

Génétique Moléculaire et Biologie du Développement, CNRS-UPR420, Villejuif, France

A limited number of genes, including the human brain-derived neurotrophic factor (*BDNF*) gene, have been identified in the human chromosome 11p14 region. Since this area is involved in a genetic disorder (WAGR syndrome) and because of interest in studying the regulation of the human *BDNF* gene, we have established a detailed transcript map of a 810-kb region clone in a yeast artificial chromosome (YAC), corresponding to a portion of this genomic locus. A set of nested deletion mutants has been generated to map genes at a mean resolution of 75 kb. Four genic markers from available mapping databases have been mapped on the YAC. Ten potential novel human exons have been isolated by a 3' terminal exon trapping procedure directly applied to purified YAC DNA. Most of these exons display polyadenylation signals and they all yield positive signals in RT-PCR experiments, confirming their status of transcribed sequences. The *BDNF* gene is now co-localised with three other genes on a 120 kb DNA fragment.

Keywords: transcript map; chromosome 11; 3' exon trapping; *BDNF* gene

Introduction

Yeast artificial chromosomes¹ (YACs) allow cloning of large fragments of genomic DNA (up to 2 Mb) and have facilitated the analysis of complex genomes by bridging physical and genetic maps. Furthermore, the ability to transfer intact YAC DNA into transgenic mice has extended the usefulness of this system to gene functional analysis.^{2,3} A YAC contig has previously been described at 11p13–p14.1⁴, which contains the human *BDNF* gene. Brain-derived neurotrophic factor⁵ (*BDNF*) is a small proteic factor that promotes the

survival and functioning of brain neurons during development and in adult life.⁶ A deficiency in this factor has been demonstrated in the brain of Alzheimer's disease patients.⁷ Reciprocally, *BDNF* administration has been shown to increase neuronal survival in numerous animal models.⁷ Thus, studying the human *BDNF* gene regulation is a critical issue for a better understanding of neurodegenerative diseases and for the design of new therapeutic approaches.

Previous mapping experiments have revealed that this gene covers at least 60 kb including a very large (> 50 kb) intron (our unpublished results). Studies involving partial versions (cDNA) of this gene in transgenic mice⁸ did not faithfully reproduce all the features of endogenous gene expression. Manipulation of the human *BDNF* gene to study its regulation should take advantage of the availability of already characterised YACs. However, good characterisation of the

Correspondence: Fabrice Guillemot, Génétique Moléculaire et Biologie du Développement, CNRS-UPR420, BP8, F-94801 Villejuif, France. Tel: 33 (0)1 49 58 35 20; Fax: 33 (0)1 49 58 35 09; E-mail: guillemo@infobiogen.fr
Received 28 October 1998; revised 4 January 1999; accepted 11 January 1999

BDNF gene environment is required prior to manipulating large genomic region. *BDNF* is also possibly associated with the WAGR syndrome⁹ (deletion at 11p13–p14). The mental retardation linked to this syndrome remains largely unexplained and identification of novel genes in this region is required better to understand this feature. In view of these issues, this report presents a transcript map of a 810-kb region of human chromosome 11p14. This map integrates already known genic markers, as well as novel exons obtained by 3' terminal exon trapping, directly performed on a non-chimeric YAC corresponding to this region. All these markers have been located along the YAC using a set of nested deletion mutants.

Materials and Methods

Yeast Strain Transfer

MegaYAC 886H12 (Centre d'Etudes du Polymorphisme Humain, CEPH-Fondation Jean Dausset, Paris)⁴ was transferred, using the *kar1*-transfer procedure¹⁰, from its original host AB1380¹ to the YPH925¹⁰ yeast strain for the YAC fragmentation procedure and to the YLBW6¹¹ yeast strain for purification of YAC DNA.

YAC Fragmentation

A set of nested deletion derivatives of YAC 886H12 was generated using the *Alu* fragmentation vectors pBP108 and pBP109 digested by *SphI* and *EcoRI* enzymes as described by Pavan *et al.*¹² About 60 transformants were analysed to check the replacement of the *URA* marker-containing YAC arm by the *HIS3* marker-containing arm of the fragmentation vectors. A subset of these clones were further analysed by STS content and pulsed-field gel electrophoresis (PFGE). A collection of nine clones containing YACs of a size from 120 kb to 745 kb was selected.

In Silico Mapping

The region of interest was delineated by two microsatellite markers: AFM 234xc3 (D11S929) and AFM 165yf10 (D11S914) known to map immediately adjacent to it and which are present as framework markers in the radiation hybrid (RH) maps available on the Web. The RH Consortium Gene Map 98¹³ (<http://www.ncbi.nlm.nih.gov/genemap98>), the Whitehead RH map¹⁴ (http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map/) based on the Genebridge4 (GB4) panel and the Stanford Human Genome Center (SHGC) RH map, based on the G3 RH Panel¹⁵ (<http://www-shgc.stanford.edu/Mapping/rh/>) were queried using this interval. The list of markers retrieved was curated by eliminating markers not corresponding to genes, markers identifying already known genes and redundant aliases. The final list contained 38 sequence-tagged sites (STS), corresponding to unique gene transcripts, whose presence was assayed on YAC 886H12 and its deletion derivatives under standard PCR conditions (see below).

3' Exon Trapping

The protocol derives from the 3' Exon Trapping System™ kit (Life Technologies, Gaithersburg, USA), except for some modifications (see below) and is based on the strategy initially described by Krizman *et al.*¹⁶ DNA from YAC 886H12 was separated from endogenous yeast chromosomes by preparative PFGE. Gel slices were digested with beta-agarase I (NEB) and the DNA was ethanol precipitated. This purified YAC DNA was digested to completion with *EcoRI* as well as the trapping pTAG4A vector. YAC DNA (500 ng) was ligated into the pTAG4A vector (250 ng). After electroporation into DH5α *E. coli* cells, subcloned DNA was transfected into COS-7 cells using Lipofectace reagent (Life Technologies) and poly A+ RNA was extracted from these cells with the Micro-Fast Track system (Invitrogen). cDNA synthesis was performed using the SuperscriptII reverse transcriptase system (Life Technologies) with the AP2 primer (AP2: 5'-AAG GAT CCG TCG ACA TC(T)₁₇-3'). The reverse transcription mixture was used as template for a first PCR reaction, using the SV40P/UAP2 primer set (SV40P: 5'-AGC TAT TCC AGA AGT AGT GA-3', UAP2: 5'-CUA CUA CUA CUA AAG GAT CCG TCG ACA TCT-3') and under the following conditions: 3 min at 94°C, 1 min at 80°C with addition of 1 U of Taq DNA polymerase (BioTaq, Eurobio, Les Ulis, France) as a hot start, 20 cycles of amplification (30 s at 94°C, 30 s at 55°C, 2 min at 72°C) and 5 min at 72°C. An aliquot of the first PCR product was digested to completion with *EcoRI* as recommended by Krizman *et al.*¹⁷ and used as template for a second PCR performed with the primer set UAP2/dUAd2 (dUAd2: 5'-CAU CAU CAU CAU CAG TAC TCT TGG ATC GGA-3'), under the same conditions as the first PCR reaction, except that 2.5 U of Taq DNA polymerase was added and 30 instead of 20 cycles were performed. The second PCR products were directionally cloned into the UDG tail-vector pAMP1 (Life Technologies). Transformants were plated and individual clones were picked at random and transferred to four 96-well microtitre plates.

Pulsed-field Gel Electrophoresis

Total yeast DNA was prepared in agarose plugs as described.⁴ PFGEs were performed on the CHEF-DRII system (BioRad, Hercules, USA) using, unless otherwise stated, 1% agarose (Bioprobe-Quantum, Montreal, Canada) gels, run at 200 V in 0.5 × TBE buffer at 13°C. Pulse conditions were for 24 h with a ramping from 20 to 30 s when analysing YAC 886H12 after transfer to strain YPH925, for 36 h with a ramping from 30 to 60 s when analysing the YLBW6 transfer, for 26 h with a ramping from 20 to 35 s or for 24 h with a ramping from 75 to 80 s when analysing fragmented YACs of a size from 120 to 360 kb or from 400 to 745 kb, respectively. Preparative electrophoresis aimed at purifying YAC 886H12 DNA was performed on a 1% Seaplaque agarose (FMC, Rockland, USA) gel run for 36 h with a ramping from 60 to 100 s.

Nucleotide Sequencing and Database Comparisons

Plasmid DNA was prepared using the Wizard Plus DNA purification system (Promega, Madison, USA) and sequenced at (Genome Express SA, Grenoble, France). Nucleotide sequences were aligned in order to derive a consensus sequence (GCG package). Sequence comparisons against EMBL and dbEST databases were performed using the BLASTN algorithm¹⁸ associated with the RepeatMasker2 program.¹⁹ Database matches with score > 100 and $P(n) < 10^{-4}$ were considered as significant and examined. The

sequence was declared novel when no significant hit appeared.

RT-PCR

Poly-A⁺ RNA (0.5 µg, Clontech, Palo Alto, USA) from various human tissues (see Table 3) were mixed with oligodT₍₁₂₋₁₈₎ primer (0.5 µg, Life Technologies) and reverse transcribed into cDNA using Superscript II reverse transcriptase (Life Technologies) in a total volume of 20 µl. For negative control, the same amount of RNA was submitted to similar incubation conditions except for the addition of reverse transcriptase. The cDNA amount corresponding to 12.5 ng of poly-A⁺ RNA was used as template for PCR amplification in a 15 µl reaction volume. Amplifications were performed with exon specific primer sets in 1 × PCR buffer (Platinum, Life Technologies) supplemented with 1 mM MgCl₂, 1 µM of each specific primer and 1 U of the antibody-complexed Platinum Taq DNA polymerase (Life Technologies). Cycling was performed on a Gene Amp PCR system 9600 or 2400 (Perkin-Elmer Cetus, Foster City, USA) as follows: 2 min at 94°C, 30 to 39 cycles (depending on the primers used, see text) composed of 30 s at 94°C, 30 s at the appropriate hybridisation temperature and 30 s at 72°C, and 8 min at 72°C.

Artefact Screening

Artefact screening was performed by a microplate PCR assay using primers PTAGLAC (5'-CG ATG TAA CCC ACT CGT G-3') and PTAGADENO (5'-ACC GCG AAG AGT TTG TCC T-3') in the same reaction together with a modified M13-reverse primer (5'-GGA AAC AGC TAT GAC CAT G-3'). Clones positive with these primer sets were eliminated from further analysis.

PCR Reactions

Standard PCR reactions were performed as described above for RT-PCR, with 30 cycles and annealing temperature depending on primer set, except that the BioTaq DNA polymerase (0.5 U) and the corresponding buffer supplemented with 1.5 mM MgCl₂ were used. For STS content analysis a boiled colony of YAC strain was added as template, whereas for mapping analysis, 2 µl of DNA solution extracted from overnight YAC cultures (Kristal Yeast kit, Eurobio) were used.

Results

YAC Fragmentation and Mapping of Known Genes

To obtain a detailed map of genes and exons located on YAC 886H12, we decided to produce nested deletion derivatives of the 810-kb long YAC, using *Alu* fragmentation vectors.¹² A set of nine different deletion mutants was characterised, ranging in size from 120 to 745 kb (Figure 1). The DNA of these different YACs provided an efficient way to localise sequence-tagged sites (STSs) by means of a PCR assay within 25 to 175-kb intervals (average resolution 75 kb) on YAC 886H12. At the beginning of this work, only two genes, namely

genes *BDNF*⁴ and *HSPCAL3* (our unpublished results) had been mapped on to YAC 886H12, in addition to a genomic marker called 245F2R.⁴ The presence of STSs derived from the *BDNF* and *HSPCAL3* genes and from the 245F2R marker was checked on the fragmentation panel. The new data obtained are in complete agreement with previous ones and even more precise in the case of the *HSPCAL3* and 245F2R markers. These two markers which were assigned to the same 300-kb *BssHII-NruI* fragment are now localised in two distinct but adjacent intervals of 100 and 50 kb respectively. (Figure 1).

Mapping of Genic Markers

To increase the density of our transcription map, we tested a list of 38 genic markers retrieved in region 11p14 from various genome centre Web sites. Four genic markers were found present on YAC 886H12 and further assigned to a common 175-kb interval using the YAC deletion panel (Figure 1). Two markers, SGC33830 and SHGC-36288 (no UniGene entry), are derived each from a unique cDNA clone, whereas the other two, WI-16368 and A009N46 (UniGene entries: Hs.113945 and 148136), are represented by two and three cDNA clones respectively. In these two latter cases a polyadenylation signal was found at the 3' end of the sequences. The four genic markers do not display any similarity to any identified sequence in the public databases.

Exon Trapping

Further refinement of the transcription map was then obtained through implementation of the 3' terminal exon trapping method initially described by Krizman and Berget.¹⁶ In order to minimise contamination by yeast DNA during the YAC DNA purification step, YAC 886H12 was transferred from the AB1380 to the YLBW6 host strain¹¹ which provides an electrophoretic 'window' devoid of endogenous yeast chromosomes between 680 and 950 kb.

Exon-trapping procedure was performed on purified YAC DNA as described in Materials and Methods; 384 clones were isolated after exon amplification and cloning. A preliminary analysis of this mini-library involved random sequencing of 19 clones. Sequence analysis revealed three types of trapped sequences: yeast exons (four clones), trapping vector sequences (six clones), and potential 3' human terminal exons (nine clones) displaying no similarity with any sequence registered in the public databases (EMBL or dbEST). Two kinds of vector-derived sequences were found in

Figure 1 Transcript map of YAC 886H12. Already known BDNF and HSPCAL3 genes and the genomic marker 245F2R are shown. STSs retrieved by the *in silico* approach are indicated in *italics*, in addition to the 10 3' terminal exons. The intervals defined by the deletion mutant are schematised above YAC 886H12. The $\square$ refers to the 3' coding exon of BDNF gene. Short vertical lines indicate restriction sites: B = BssHII, N = NruI, Nt = NotI. The right and the left facing arrows symbolise respectively the right arm containing centromeric sequence (●) and the left arm of the YAC vector. The set of nested deletion mutants is shown below the transcriptional map. YAC mutants are partially represented, except for YAC 886.109.7. cen and tel refer to centromeric and telomeric orientations of the cloned genomic region (11p14).

our sample. The corresponding specific primers were designed so as to detect readily through a PCR assay such vector-derived artefacts in the rest of the mini-library. This screening eliminated about 80% of the clones from further analysis. All remaining clones (59) were subjected to sequencing and comparison with databases. Two new kinds of vector-derived sequences were identified (17 clones), as well as 11 clones corresponding to yeast sequences. Furthermore, two clones did not contain the 5' exon of the trapping vector in their sequence. These clones probably derived from non-specific amplification of some COS-7 endogenous mRNA during the trapping procedure. Thus 29 clones remained as potential human 3' exons, bringing to 38 the total number of clones resulting from expected splicing events between the 5' exon of the trapping vector and the genomic DNA cloned into this vector.

In summary, a total of 329 clones (86% of the mini-library) were artefact clones derived from vector-driven

self-splicing events, whereas 15 clones (4%) were yeast sequences resulting from contamination of YAC DNA during purification. Potential human 3' terminal exons represented 10% of the mini-library, but thanks to the PCR-based artefact screening procedure, these sequences corresponded to about half of the sequenced clones.

Sequence Analysis

The sequences of the 38 human clones were grouped into 17 different clusters based on sequence identities. Sequence comparison with genomic and cDNA databases indicated that only two putative exons (A2G12 and B2C1) matched with already known sequences (EMBL accession numbers H57351 and T97154, respectively). The WI-18122 marker associated with the H57351 sequence is reported to map to the long arm of chromosome 11 (Gene Map 98). Since this marker was later not found to be present on the YAC, the origin of

the A2G12 exon remains unexplained and it was not further analysed. On the contrary, the SGC33830 gene marker associated with the T97151 sequence was shown to map to YAC 886H12 (see above) and thus provides an excellent internal control for our exon trapping procedure. Sequence comparison also revealed the presence of repeat elements in about half the clusters. In five cases the repeat sequence was found adjacent to the poly-A tail. This situation does not permit differentiation between an internal repeat element and the 3' end of a transcript. Therefore these sequences (accession numbers: AJ311606, AJ131607, AJ131608, AJ131609 and AJ131610) were kept apart from the transcription map. Three other clusters were still considered as potential 3' exons since the alignment with the repeat motif was either very poor or concerned a limited portion of the sequences.

Sequence features of the 11 trapped exons are presented in (Table 1). Most exons (7/11) are represented by more than one clone. The exon size ranges from 205 to 416 base pairs (bp) with a mean size of 315 bp. For both exons A2C4 and B2E8, a longer

sequence is found that perfectly overlaps the shorter one and is probably produced by alternative polyadenylation. Consensus poly A signals (AATAAA or ATTAAA) were found in eight of the 11 exons. Due to the rather short length of the 3' terminal putative exons, no significant open reading frame could be found in their sequences. All these complete nucleotide sequences have been deposited in the EMBL database (see Table 1 for accession numbers). Apart from exon B2C1 already mentioned above, no similarity with any entry in the public sequence databases could be found for the 10 novel sequences.

3' Exon Mapping

To validate the origin of all trapped exons, specific primer sets were designed from their sequence (Table 2). Amplifications were performed using these primers on YAC 886H12 total DNA (including yeast DNA), and on yeast and human genomic DNA as negative and positive controls respectively. This experiment assessed the presence on YAC 886H12 of the 10 new trapped 3' exons. Interestingly, the C2D9 clone is

Table 1 Features of the 11 clusters of sequences trapped from YAC 886H12 and considered as 3' terminal exons

Name	No. of clones	Size (bp)	Repeats (position in sequence, % of identity)	Poly-A signal (distance from poly-A)	EMBL accession number
A2-C4	1 short 2 long	210 325	MIR (20–164, 65%)	1st: – 2nd: attaaa (–29 bp)	AJ011593
B2-C1 ^a	7	416	no	–	T97154
B2-D7	2	207	Alu Jb (6–116, 79%)	–	AJ011594
B2-E7	1	205	no	aataaa (–16 bp)	AJ011595
B2-E8	3 short 2 long	226 306	no	1st: attaaa (–24 bp) 2nd: aataaa (–19 bp)	AJ011596
B2-F11	2	310	no	aataaa (–17 bp)	AJ011597
B2-G5	4	225	no	attaaa (–19 bp)	AJ011598
C2-A11	1	272	no	–	AJ011599
C2-B5	1	301	no	aataaa (–20 bp)	AJ011600
C2-C4	2	228	no	aataaa (–22 bp)	AJ011601
C2-D9	1	358	MIR (97–148, 71%)	aataaa (–15 bp)	AJ011602

^aT97154 is a partial cDNA sequence previously registered and associated to STS SGC33830.

Table 2 Primer sets for the trapped sequences

Primer name	Sequence 5'–3'	Product size (bp)	HT (deg C)
A2C4 S	CTT GCA GAA GGT CAC ATC GG	122	59
AS	AGC TCA GAA ATA GCA AGT GGG AG		
B2D7 S	AGA TGG GGG TCT CAC CAC CT	166	59
AS	CAC AAT TCA TGG GGA AAA GAG AA		
B2E7 S	TCC ACA GGA CAA AGA GGA AAG G	151	58
AS	ATG AAT AAA TGA AAT GTA CAT GCA CG		
B2E8 S	AAC ATG CCT GAG GGC TCT GG	175	58
AS	AGA ATT GGG TTC ACC GTG GG		
B2F11 S	TCA TTT GGG GTG ATG ACT GC	124	59
AS	CCA GAA AAT GAA GCA TTT AGG AAA		
B2G5 S	GGA GAT GTC CCA AGG CCT AA	165	59
AS	GCT TCA AGT TGT TGT GGG AGC		
C2A11 S	AGG AAG ATT TTC ACT TTT GCC C	155	58
AS	CCA TAA CAC AGC ACA TGG CTC		
C2B5 S	CAG CAT CAA AGG GAA GCC C	233	59
AS	CTT CAG GCA AGG CTT GGT G		
C2C4 S	ATG TCA AAA ACA CAG AGG GTC	197	59
AS	CAG AGA GCC TAG AGA GTA AGG AAC GT		
C2D9 S	ATC GCA TCA AGT TGT GTT GTG A	156	58
AS	TGC CAG CAT CAT TTA TTG CC		

S: sense; AS: antisense.

probably due to simultaneous trapping of two distant exons, as reported elsewhere,²⁰ since the first primer set designed from the ends of its sequence failed to give an amplification product on total human genomic DNA.

Refined mapping was performed using the YAC fragmentation panel described above. Results are schematised in Figure 1. Three of the exons are contained in a 120-kb area around the *BDNF* gene and one in the adjacent 25-kb interval, suggesting a rather dense organisation in this 145-kb area of genomic DNA.

Expression Studies

Expression studies were conducted in order to ascertain the possible transcription of the trapped sequences. Primers previously designed (Table 2) were used in RT-PCR reactions performed on polyA⁺ RNA derived from 11 human tissues (Table 3). In the first instance, each assay was run for 30 cycles of amplification. Negative controls were performed to rule out genomic DNA contamination. The expression of five out of the 10 new exons was detected in all analysed tissues under standard amplification conditions, whereas in these conditions three exons (B2E7, B2E8 and B2G5) displayed a more restricted pattern of expression. The detection of A2C4 and C2A11 expression required longer cycling (36 cycles) and concerned a limited

number of tissues. However, these RT-PCR experiments indicate that all 10 trapped exons are expressed in at least one of all the human mRNA pools analysed here. Similar studies were also performed with the four genic markers retrieved from the RH map and assigned to YAC 886H12 in this study. Expression of these markers was detected in all the analysed tissues under standard amplification conditions (Table 3).

Discussion

We report here the construction of a transcript map of a 810-kb region at 11p14. In order to increase the resolution of the restriction mapping previously performed on this YAC,⁴ a set of nested deletion mutants has been generated by fragmentation.¹² This collection of nine size-reduced YACs allowed refinement of marker mapping by doubling the average resolution (ie the mean interval of localisation is decreased from 135 kb to 75 kb). In theory, this result could even be improved if a larger number of clones resulting from the fragmentation process were to be analysed and characterised. However, the number of distinct deletion mutants that can be obtained is limited by the density of *Alu* elements in the YAC insert. Although small

Table 3 RT-PCR expression analysis

Exon	Foetal liver	Placenta	Liver	Kidney	Foetal brain	Brain	Skeletal muscle	Pancreas	Heart	Lung	Testis
Potential 3' exon											
A2C4	*			*	*	*	*				*
B2D7	*	*	*	*	*	*	*	*	*	*	*
B2E7	*										*
B2E8											*
B2F11	*	*	*	*	*	*	*	*	*	*	*
B2G5	*		*	*							*
C2A11											*
C2B5	*	*	*	*	*	*	*	*	*	*	*
C2C4	*	*	*	*	*	*	*	*	*	*	*
C2D9	*	*	*	*	*	*	*	*	*	*	*
Genic markers											
A009N46	*	*	*	*	*	*	*	*	*	*	*
SGC33830	*	*	*	*	*	*	*	*	*	*	*
WI-16368	*	*	*	*	*	*	*	*	*	*	*
SHGC-36288	*	*	*	*	*	*	*	*	*	*	*

*indicates the detection of the expression in the corresponding tissue(s). Detection of C2A11 and A2C4 expression required 36 cycles of amplification but all the other exons gave results with 30 cycles of amplification. Experiments were performed in duplicate.

interspersed repetitive sequences have been estimated to occur with an average spacing of 3 kb in the human genome,²¹ their distribution depends on the genomic region analysed.²² Nevertheless, *Alu*-based fragmentation proved to be an efficient tool for YAC mapping that may provide a resolution similar to that obtained using cosmid subclones of YAC when this method is associated with partial enzymatic digestion.²³ As an interesting variation of the technique, exons or cDNAs can be cloned into fragmentation vectors instead of repetitive sequences and the recombination events with YACs can readily provide a precise information about the location and orientation of these elements.^{24,25}

In-silico retrieval of the list of available genic markers was made difficult due to the paucity of framework markers shared between maps and the lack of annotations in the Whitehead and Stanford RH maps. Three of four genic markers found on YAC 886H12 (A009N46, SGC33830 and WI-16368) in a 175-kb interval derive from a 3-cM (11 cR₁₀₀₀) interval in the Human Gene Map 98 which corresponds to about 3 Mb. The third genic marker, SHGC-36288, is surrounded in the Stanford RH Map by two markers which delineate a 26 cR₁₀₀₀₀ (about 75 kb) interval. The estimated resolution of RH maps ranges from 30 kb per cR₁₀₀₀₀ for the G3-based map to 300 kb per cR₃₀₀₀ for the GB4-based map. These values nicely flank the resolution obtained for the physical map described in

this paper thanks to the fragmentation panel and could allow further integration of the data. Indeed, it would be worth testing the presence on YAC 886H12 of the framework markers surrounding the SHGC-36288 marker in the Stanford map, and conversely assaying on the G3-RH panel the gene markers identified in this study as present on YAC 886H12.

Eleven human exons have been isolated in this work by 3' terminal exon trapping. A high level of background clones (86%) has been identified as resulting from splicing events in the vector sequence. This mechanism is probably favoured by the lack of an efficient termination signal near the cloning site of pTAG4A, allowing transcription of vector sequence beyond this site. Vector artefacts could then be obtained by the combination of genomic inserts devoid of splicing and polyadenylation signals and the use of such cryptic elements in the vector sequence. An improved 3' exon trapping method has been described¹⁷ and is based on direct transformation into COS-7 cells of the uncloned products of a ligation between digested genomic DNA and double-digested pTAG4A vector. This strategy, which requires more genomic DNA, seems to circumvent the background problem since 1–6% of vector-derived background has been reported.^{17,29}

Only one of the 11 human trapped exons matches a sequence previously registered in public databases.

Since the method chosen to isolate exons does not rely on the expression level of genes, the under-representation of the trapped exons in cDNA databases is likely to be related to the low abundance of corresponding mRNAs in the tissues used for cDNA library construction.

Eight of the 11 trapped exons contain a consensus poly A signal (ATTAAA or AATAAA). This 3' terminal exon hallmark is considered by Krizman *et al*,¹⁷ to be a good warranty for trapped sequence validity. However, the lack of poly A signal in the B2D7 and C2A11 clusters does not exclude their terminal exonic nature since about 10–20% of the genes do not display such a consensus sequence.²⁶ On the other hand, in the absence of true functional poly-A signal, these exons may result from inappropriate priming with oligodT during the RT-PCR. For example, in the case of A2C4, the shorter sequence could be a truncated version of the longer transcript. In contrast, the two forms of B2E8 exon both display a consensus poly-A site suggesting that they resulted from a bona fide alternative polyadenylation.

Despite the fact that all the trapped sequences described here are expressed elements (according to RT-PCR data) it cannot be asserted that they correspond to 10 distinct novel genes. The method used to isolate these exons suggests that some of them may result from inappropriate priming with oligodT during RT-PCR as mentioned above or from the utilisation of cryptic poly A sites as described by others.²⁰ Furthermore, many genes have been described to have multiple poly-A sites leading to production of transcripts with non-overlapping 3' terminal exonic sequences.²⁶ This could be, for instance, the case of the five genic markers B2D7, WI-16368, SHGC-33830, SGC-36288 and A009N46 mapping to the same interval and displaying the same expression pattern, or for the C2D9, B2F11 and C2C4 exons which are in the same situation. The cloning of the complete cDNAs corresponding to the trapped exons should determine if some of them are different versions of a single gene.

In conclusion, the present investigation contributes to the refinement of the 11p14 transcript map by assignment of four already registered gene transcripts and identification of 10 novel human 3' exons. The novel exons have been mapped on a 810-kb area including the *BDNF* gene and lying in the WAGR syndrome locus. Further expression studies consisting of northern blot analysis are underway to determine precise tissue transcription pattern of some of these

exons and the length of the transcripts. We are now interested in cloning the complete cDNAs in order to explore the function of the corresponding novel genes. Such studies could increase the number of genes which may account for some features of the WAGR phenotype (such as mental retardation) and, more globally, will contribute to better characterisation of the genic environment of the human *BDNF* gene, thus providing a more comprehensive basis for exploration of gene function without dissociation from its genomic context.

Acknowledgements

We are grateful to R Reeves for the gift of the 'window' yeast strains and the fragmentation vectors. We also thank R Joudani for technical assistance. This work was supported by Association Française contre les Myopathies (AFM) grants. FG is grateful to AFM, France Alzheimer Association and Ligue Nationale contre le Cancer for fellowships.

References

- 1 Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987; **236**: 806–812.
- 2 Peterson KR, Clegg CH, Li Q, Stamatoyannopoulos G: Production of transgenic mice with yeast artificial chromosomes. *Trends Genet* 1997; **13**: 61–66.
- 3 Rubin EM, Smith DJ: Optimizing the mouse to sift sequence for function. *Trends Genet* 1997; **13**: 423–426.
- 4 Rosier MF, Goguel AF, Martin A *et al*: YAC contig around the human *BDNF* gene (11p13): integration of the physical, genetic, and cytogenetic maps in relation to WAGR syndrome. *Genomics* 1994; **24**: 69–77.
- 5 Leibrock J, Lottspeich F, Hohn A *et al*: Molecular cloning and expression of brain-derived neurotrophic factor. *Nature* 1989; **341**: 149–152.
- 6 Snider WD: Functions of the neurotrophins during nervous system development: what the knockouts are teaching us. *Cell* 1994; **77**: 627–638.
- 7 Connor B, Dragunow M: the role of neuronal growth factors in neurodegenerative disorders of the human brain. *Brain Res* 1998; **27**: 1–39.
- 8 Timmusk T, Lendahl U, Funakoshi H, Arenas E, Persson H, Metsis M: Identification of brain-derived neurotrophic factor promoter regions mediating tissue-specific, axotomy-, and neuronal activity-induced expression in transgenic mice. *J Cell Biol* 1995; **28**: 185–199.
- 9 Hanson IM, Seawright A, van Heyningen V: The human *BDNF* gene maps between FSHB and HVBS1 at the boundary of 11p13–p14. *Genomics* 1992; **13**: 1331–1333.
- 10 Spencer F, Hugerat Y, Simchen G, Hurko, O Connelly C, Hieter P: Yeast kar1 mutants provide an effective method for YAC transfer to new hosts. *Genomics* 1994; **22**: 118–126.

- 11 Hamer L, Johnston M, Green ED: Isolation of yeast artificial chromosomes free of endogeneous yeast chromosomes: construction of alternate hosts with defined karyotypic alterations. *Proc Natl Acad Sci USA* 1995; **92**: 11706–11710.
- 12 Pavan WJ, Hieter P, Sears D, Burkoff A, Reeves RH: High-efficiency yeast artificial chromosome fragmentation vectors. *Gene* 1991; **106**:125–127.
- 13 Schuler GD, Boguski MS, Stewart EA *et al*: A gene map of the human genome. *Science* 1996; **274**: 540–546.
- 14 Hudson TJ, Stein LD, Gerety SS *et al*: An STS-based map of the human genome. *Science* 1995; **270**: 1945–1954.
- 15 Stewart EA, McKusick KB, Aggarwal A *et al*: An STS-based radiation hybrid map of the human genome. *Genome Res* 1997; **7**: 422–433.
- 16 Krizman DB, Berget SM: Efficient selection of 3'-terminal exons from vertebrate DNA. *Nucleic Acids Res* 1993; **21**: 5198–5202.
- 17 Krizman DB, Hofmann TA, DeSilva U, Green ED, Meltzer PS, Trent JM: Identification of 3'-terminal exons from yeast artificial chromosomes. *PCR Methods Appl* 1995; **4**: 322–326.
- 18 Atschul SF, Madden TL, Schaffer AA *et al*: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
- 19 Smit A, Green P: at <http://ftp.genome.washington.edu:80/cgi-bin/RepeatMasker>.
- 20 Elkahoul AG, Krizman DB, Wang Z, Hofmann TA, Roe B, Meltzer PS: Transcript mapping in a 46-kb sequenced region at the core of 12q13.3 amplification in human cancers. *Genomics* 1997; **42**: 295–301.
- 21 Moyzis RK, Torney DC, Meyne J *et al*: The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* 1989; **4**:273–289.
- 22 Liu P, Siciliano J, Seong D *et al*: Dual *Alu* polymerase chain reaction primers and conditions for isolation of human chromosome painting probes from hybrid cells. *Cancer Genet Cytogenet* 1993; **65**: 93–99.
- 23 Cook GP, Tomlinson IM: High resolution restriction mapping of YACs using chromosome fragmentation. *Nucleic Acids Res* 1996; **24**: 1585–1586.
- 24 Das Gupta R, Morrow B, Marondel I *et al*: An integrated approach for identifying and mapping human genes. *Proc Natl Acad Sci USA* 1993; **90**: 4364–4368.
- 25 Mazzearella R, Pilia G: Recombination trapping: an *in-vivo* approach to recover cDNAs encoded in YACs. *Gene* 1998; **212**: 315–322.
- 26 Edwalds-Gilbert G, Veraldi KL, Milcarek C: Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* 1997; **25**:2547–2561.