

ARTICLE

A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus

Alexandre Montpetit^{1,2}, Gino Boily¹ and Daniel Sinnett^{*,1,2,3}

¹Division of Hematology-Oncology, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada; ²Department of Biochemistry, University of Montreal, Montreal, QC, Canada; ³Department of Pediatrics, University of Montreal, Montreal, QC, Canada

Loss of heterozygosity of the short arm of chromosome 12 is a frequent event in a wide range of haematological malignancies and solid tumours. In previous studies, the shortest commonly deleted region was delimited to a 750-kb interval, defined by the markers *D12S89* and *D12S358*, in pre-B acute lymphoblastic leukaemia patients, suggesting the presence of a tumour suppressor locus. Here we report the construction of a transcriptional map that integrates the data obtained by genomic sequence analysis, EST database search, comparative analysis and exon amplification. We identified seven putative transcriptional units as well as six pseudogenes. Four of these candidate genes were already known: *ETV6*, encoding an ets-like transcription factor, *LRP6*, a member of the LDL receptor gene family, *BCL-G*, a recently identified pro-apoptotic gene and *MKP-7*, encoding a new member of the dual-specificity phosphatase family. The products encoded by the three new genes identified in this study, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*, have no clear homology to known proteins. The gene predictions were all confirmed by expression analysis using RT-PCR and Northern blot. This transcriptional map is a crucial step toward the identification of the tumour suppressor gene at 12p12.

European Journal of Human Genetics (2002) 10, 62–71. DOI: 10.1038/sj/ejhg/5200766

Keywords: tumour suppressor gene; leukaemia; chromosome 12; transcriptional units; gene expression

Introduction

The short arm of chromosome 12 is frequently rearranged in a wide variety of haematological malignancies of both myeloid and lymphoid origins.^{1–3} Most of these 12p alterations result from unbalanced translocations and deletions that lead to a loss of genetic material.⁴ The frequent loss of genetic material in tumour cells is usually indicative of the inactivation of tumour suppressor genes, which prompted the search for a suppressor locus on chromosome 12p. Loss of heterozygosity studies showed hemizygous deletions of chromosome 12p12 in 26 to 47% of pre-B acute lympho-

blastic leukaemia (ALL) cases, making it one of the most common genetic alterations found in this disease.^{5–7} The construction of both high-resolution genetic and physical maps led to the delineation of the shortest commonly deleted region within a 750 kb interval delimited by markers *D12S89* and *D12S358*.^{8–10} The existence of a putative 12p12 tumour suppressor gene is further substantiated by the observation of hemizygous deletions in a variety of haematological malignancies, as well as in certain solid tumours including breast, lung, ovarian and prostate carcinomas.^{11,12}

This region was shown to contain two known genes, *LRP6* and *ETV6*. *LRP6* encodes a member of the LDL receptor family¹³ that was recently shown to act as the WNT co-receptor.^{14–16} However Baens *et al.*¹⁰ failed to reveal any inactivating mutations in leukaemia patients. *ETV6*, also known as TEL, is a member of the ets-like family of DNA binding proteins, that was originally identified by virtue of its

*Correspondence: Dr Daniel Sinnett, Charles-Bruneau Cancer Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada. Tel: (514) 345-2990; Fax: (514) 345-4731; E-mail: daniel.sinnett@umontreal.ca

Received 12 September 2001; revised 23 November 2001; accepted 30 November 2001

fusion to the platelet-derived growth factor receptor β (PDGFR β) in chronic myelomonocytic leukaemia with the translocation t(5;12) (q33;p13).¹⁷ Recently, it has been shown that ETV6 acts as a transcriptional repressor of several matrix metalloproteinases and that its expression could counteract the transformation of cultured cells by ras.^{18,19} However, mutational analysis failed to detect deleterious mutations in the second ETV6 allele of ALL patients with hemizygous deletion at this locus^{6,20} as expected for a classical tumour suppressor gene. These observations taken together led to the suggestion that an as yet unidentified suppressor locus might be located between D12S89 and D12S358.¹¹

Here, we describe the application of a strategy that integrates the use of data mining tools, exon amplification experiments, gene expression studies and comparative genomic analysis to identify putative transcriptional units in the chromosome 12p12 locus. This strategy led to the identification of seven genes, including three uncharacterised genes, and six pseudogenes, indicating the usefulness of integrating different complementary genome-based approaches to identify candidate genes in a large chromosomal interval.

Materials and methods

Genomic clones

BAC clones RP11-525I13, RP11-267J23, RP11-757G14 and RP11-253I19 (GenBank accession numbers AC022222, AC007537, AC007621 and AC007619 respectively) were obtained from Incyte Genomics, Inc (St-Louis, MO, USA). The sequence of these BACs, of BAC RPC111-180M15 (GenBank accession number AC008115) as well as the whole ETV6 contig (GenBank accession numbers NT_000600, AC005989 and NT_000601) were selected for analysis based on their localization within or near the shortest commonly deleted region according to the chromosome 12 physical map at the Human Genome Sequencing Center, Baylor College of Medicine (<http://www.hgsc.bcm.tmc.edu>).

Sequence analysis

The genomic sequences available from the selected BAC clones were used as queries for BLAST searches and analysed using the NIX software (<http://www.hgmp.mrc.ac.uk/NIX>) that displays the results from 15 different prediction algorithms (Grail-CpG, Grail-exons, FGenes, Hexon, FEX, GenScan, etc) and BLAST analysis on many public databases (dbEST, Unigene, htgs, etc.). The programme Exofish (<http://www.genoscope.cns.fr/externe/tetraodon>) was used to identify the presence of conserved regions at the protein level between the human sequences and the genome of *Tetraodon nigroviridis*. All other sequence analysis were performed using BLAST or CD-Search programmes on the NCBI site (<http://www.ncbi.nlm.nih.gov>). Multiple sequence alignments were obtained using the ClustalW algorithm²¹ and manually

edited using the SeaView programme (<http://pbil.univ-lyon1.fr/software/seaview.html>). Alignments were displayed using the BOXSHADE programme (http://www.ch.embnet.org/software/BOX_form.html).

Exon amplification

Exon amplification was performed using the Exon Trapping System (GibcoBRL-Life Technologies, Burlington, Ontario, Canada) according to the manufacturer's recommendations. Briefly, BAC clones were partially digested with *Sau3A*. The resulting restriction fragments ranging from 2.5 to 4 kb were isolated from 1% agarose gels and subcloned into *Bam*H1-digested pSPL3. Individual subclones were pooled and transfected into 80%-confluent COS-7 cells. Total cellular RNA was extracted 24 h after transfection and used as a template for RT-PCR using vector-specific primers. The fragments containing putative trapped exons were subcloned into pAMP10 and sequenced using T7 primer.

5'-Rapid amplification of cDNA ends (RACE)

Human testis Marathon-Ready cDNA (Clontech, Palo Alto, CA, USA) was used to amplify the transcript 5' ends according to the manufacturer's instructions. Briefly, 200 ng of Marathon-ready cDNA was amplified with AP1 primer and a gene-specific primer: 11962RACE (CTGTCCATGACTCATTGCTGAACA), 11962RACE2 (AGGTGGTTAGTTCTGCCAGGCGAGT), LOH2RACE (CCTTTAAGGTGGCATCGCCCTAAGT) or LOH2RACE2 (CTAATCCGGAAGTGAAGGTCTTGCG) by Touchdown PCR. The resulting RT-PCR products were analysed on 1% agarose gel, transferred to nylon membranes (Hybond N+, Amersham-Pharmacia Biotech Baie d'Urfé, Canada) and hybridised with a radioactive internal oligonucleotide to assess the specificity of the bands. The sequence of the putative 5'-end-derived exons was determined by dideoxy sequencing using the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing kit (USB, Cleveland, OH, USA) following the subcloning of the PCR products into pGEMT-easy (Promega, Montreal, Canada).

RNA expression analysis by RT-PCR

Normal human brain, testis and prostate RNAs were obtained from Clontech (Palo Alto, CA, USA). Placenta, peripheral blood and bone marrow RNAs were isolated from healthy subjects using standard procedures. Following DNase-treatment, the RNA samples were incubated at 37°C for 60 min in the presence of 15 μ M of random hexamers, 0.2 mM of each dNTPs and 200 U of MMLV reverse transcriptase (Gibco-BRL, Life Technologies, Burlington, Canada). The cDNAs were amplified using 0.2 mM dNTP, 0.4 μ M each primer and 2 U of ELONGase polymerase (Gibco-BRL, Life Technologies, Burlington, Canada) in a 60 mM Tris-SO₄ pH 9.0 buffer containing 18 mM of (NH₄)₂SO₄, 8% of DMSO and 1.5 mM of MgSO₄. The primers amplifying the entire coding regions were 11962F (GTGGGTGTGACCTGGAAGAAAT) and 11962R

(TGCACAGGACCACAGTAGACA) for the *BCL-G_L* gene; 97818F (CAAAATTGTTGAGCTGCTGAAAT) and 97818R (GGGGAAGTGCAGAAGTAGGTC) for the *BCL-G_M* gene; DUSPF (CTCGAGGGTGGGAAAAGAGACTTATTG) and DUSPR (GGATCCTTTCAGATTTACAGGGAATTTT) for the *MKP-7* gene; LOH1F (CTCGAGCCGCCGTTCTTCTGCTG) and LOH1R (GGATCCCCTTAGCTCCACACGTCCTC) for the *LOH1CR12* gene; LOH2F (GGCTACGCAAGACCTT-CAGTT) and LOH2R (ACTGCTTAGTGGGCCCTTCCCAGTC) for the *LOH2CR12* gene; FLJF (CTCGAGCCTTGACCTT-GAAGACCAAAA) and FLJR (AAGCTTAGAGACAC-CAGTTCCATCC) for the *LOH3CR12* gene. The cycling conditions were as follows: initial denaturation of 2 min at 94°C, 35 cycles of 94°C for 30 s, 60–64°C for 30 s, and 72°C for 60 s, and a final extension at 72°C for 10 min. RT-PCR products were revealed by 1 or 2% agarose gel electrophoresis. They were then excised from the gel, cloned into pGEMT-easy (Promega, Montreal, Canada) and sequenced completely.

Northern blot hybridisation

Gel-purified RT-PCR products (see above) were radiolabelled by random priming and used as probes to hybridise multiple Tissue Northern Blots (Human MTN Blot I, Human MTN Blot II and Human Immune System MTN Blot II, Clontech, Palo Alto, CA, USA). The hybridisations were carried out at 50°C in a solution containing 50% formamide, 10% dextran sulfate, 1% SDS, 1 M NaCl, 50 µg/ml fish sperm DNA and 5 × 10⁶ c.p.m./ml of a given probe. Following hybridisation, the filters were washed at high stringency with 0.1 × SSC, 0.1% SDS at 55°C and autoradiographed at –80°C with X-OMAT Kodak films.

Results

Sequence analysis

The availability of BAC sequences within the chromosome 12p12 suppressor locus, as part of the human genome sequencing project, provided the molecular framework to search for transcribed sequences. Four BAC clones, RP11-525I13, RPCI11-267J23, RP11-757G14 and RP11-253I19, mapped between markers *D12S89* and *D12S358*. Only partial and unordered sequences were available for clones 525I13 and 253I19 at the time of the writing of this paper. The genomic gap between the marker *D12S89* and the BAC 525I13 (Figure 1) was filled by using the data from the complete sequence of the *ETV6* locus.²² This genomic sequence was analysed with the NIX programme to identify putative transcribed sequences. We identified six pseudogenes based on the absence of introns and the presence of stop codons within the consensus coding sequence. In addition, we found four CpG islands and over 90 EST clusters, 38 of them corresponding to Unigene entries. Many of these ESTs appear to be artefacts as suggested by their low representation (i.e. either unique or only two members) or by the presence of a polyA tail at their 3' end in the corresponding genomic sequence without any upstream polyadenylation signal. This could have allowed oligo dT priming during RT-PCR on mRNA, but also on genomic DNA or on unspliced RNA. To restrict the number of ESTs to analyse, they were classified according to the following characteristics: (1) presence of spliced sequence, (2) homology with genes/ESTs or proteins, when translated, from other species, (3) evolutionarily conserved as shown by Exofish analysis, (4) matches with trapped exon(s) and (5) presence of a polyadenylation signal at the 3' end of the cluster. In this

Table 1 Predicted transcriptional units within the chromosome 12p12 tumour suppressor locus

Unigene/EST cluster	No. ESTs	Gene (acc.no.)	Spliced sequence	Mouse orthologue (acc.no.)	Exofish Hits (ecores)	Trapped exons	PolyA signal
Hs.169081	40	ETV6 (NM_001987)	yes	NM_007961	4	nd ^a	yes
Hs.293972	30		no		–		yes
Hs.146381	18		no		–		yes
Hs.11962	20	<i>BCL-G_L/G_S</i> (AF281254/NM_030766)	yes	BF139422	–	2	yes
Hs.97818	2	<i>BCL-G_M</i> (AY040274)	yes	–	–		yes
Hs.23672	52	LRP6 (NM_002336)	yes	NM_008514	17	2	yes
Hs.41269	93		no		–		yes
Hs.5999	40	LOH3CR12 (AK023622/NM_018050)	yes	NM_026345	–	–	yes
Hs.50860	5		no		–		yes
yx93f07	5	LOH2CR12 (AY037867)	yes	–	–	–	no
Hs.67553	7		no		–		yes
Hs.105040	43	LOH1CR12 (AY037865/AY037866)	yes	AK007423	2	–	yes
Hs.86330	3		no		–		yes
Hs.232681	4		no		–		yes
Hs.20281	147	MKP-7 (AB052156/AY038927)	yes	AB052157	3	3	yes
Hs.7341	61		no		–		yes
Hs.291547	1	–	no	–	–	–	yes
Hs.171346	2	–	no	–	–	–	yes

^anot done. BACs used in the exon trapping experiments did not cover the *ETV6* region.

study, only the EST clusters that fulfil at least one of the above characteristics were further investigated.

The EXOFISH algorithm compares genomic sequences to 150 Mb of *T.nigroviridis* sequence (corresponding to 33% of its genome) at the protein level. This algorithm has been successful in detecting two thirds of the known genes on human chromosome 22 with an estimated specificity of over 90%.²³ In this study, 33 hits (termed ecores) corresponded to four EST clusters (Table 1) and five pseudogenes (data not shown), but to no other regions. The exon trapping technique identified seven exons corresponding to three of the clusters found in the region analysed (Table 1). Six other trapped sequences did not reveal any match in the EST database, but the absence of a translation frame may suggest false positives rather than actual exons. Applying all these criteria, only 20 clusters were kept for further analysis (Table 1). None of the other clusters analysed contained an ORF that could encode a protein. Of the 20 clusters selected, eight contained spliced sequences and were considered the most interesting. Except for two clusters, *BCL-G_M* and *LOH2CR12*, they all had clear orthologues from other species in the dbEST. Unigene cluster Hs.97818 corresponded clearly to an alternative splicing form of the *BCL-G* gene (see below). The physical relationship of these predicted genes within the

commonly deleted region is presented in Figure 1. To be more informative, we have extended the map on the proximal side in order to include the known genes *CREBL2*, *GPR19* and *CDKN1B*. No other putative transcriptional units were present on the corresponding BAC clone RPCI11-180M15 (data not shown). Two of the clusters corresponded to the *ETV6*²² and *LRP6*¹³ genes that were previously characterised in detail and mapped to this region^{13,22} and as such will not be further discussed in this section.

By using the assembled ESTs and exon prediction programmes it was possible to obtain an open reading frame (ORF) starting with a methionine for all the remaining selected clusters except for *LOH2CR12* (Figure 2). The region encompassing the putative ORFs of all the clusters could be amplified by RT-PCR in multiple tissues, confirming the gene prediction (Table 2). Furthermore, many alternative splicing species could be observed by this analysis (see below).

Characterisation of the candidate genes

The human *BCL-G* gene was originally characterised as composed of six exons that encode at least two known splicing variants, *BCL-G_L* (long form, 327 amino acids) and *BCL-G_S* (short form, 252 amino acids) isoforms.²⁴ The BH3

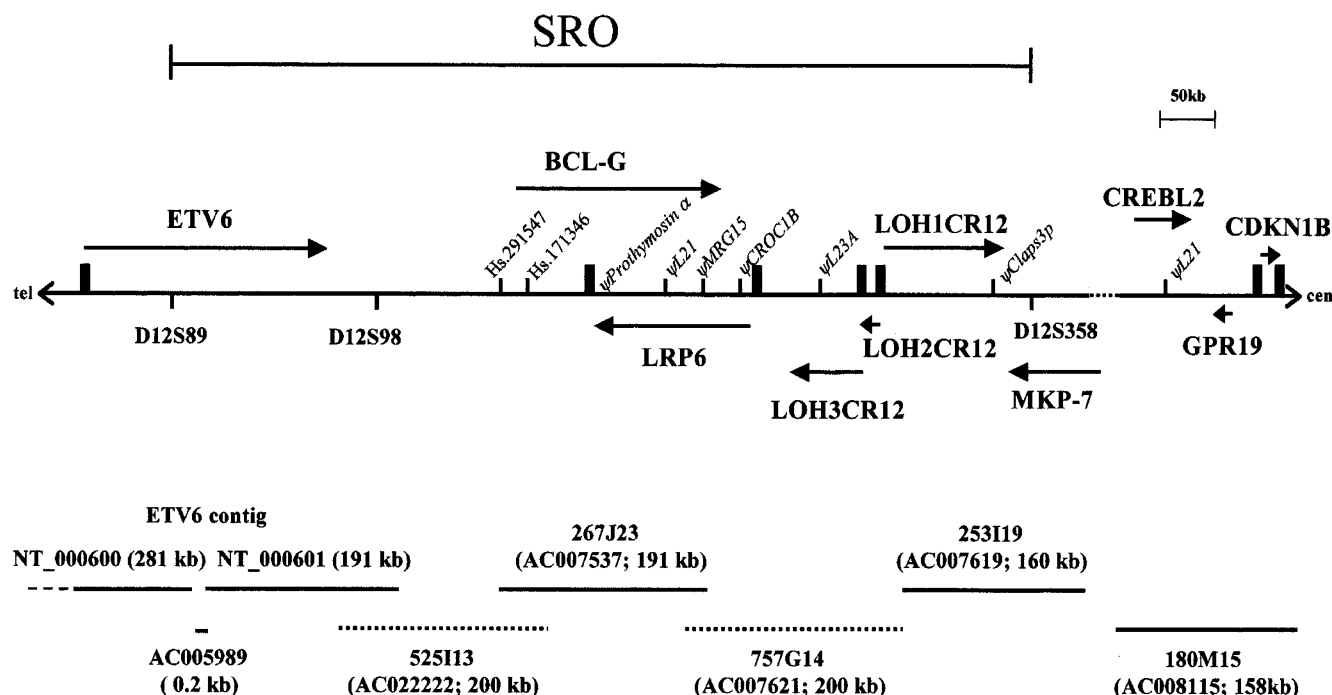


Figure 1 Physical and transcriptional map of the commonly deleted region on chromosome 12p12. The shortest region of overlapping deletion (SRO) is delimited by markers *D12S89* and *D12S358*. Filled rectangles indicate the presence of a CpG island. Arrows indicate the orientation of transcription. Other EST clusters (Hs. 291547 and Hs.171346) found in this study are also indicated on the map. Pseudogenes are indicated in italics. The position and size of the genomic clones, as well as their corresponding accession numbers are shown in the lower section. The dotted lines indicate partial and unordered genomic sequences. Tel, telomere; cen, centromere.

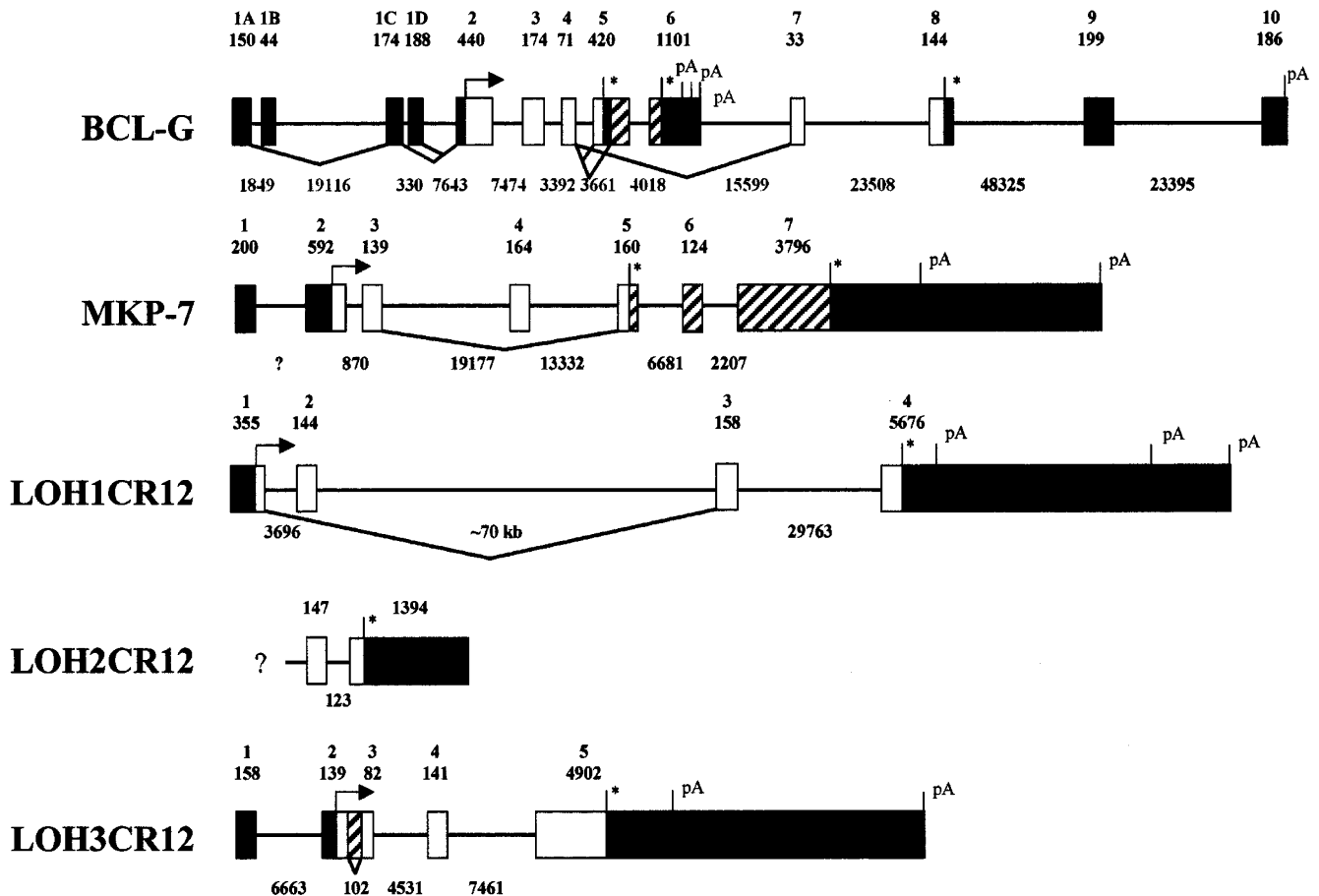


Figure 2 Genomic structure of *BCL-G*, *MKP-7*, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*. Alternative splicing is indicated by solid lines joining two exons. Coding and non-coding exons are represented by open and filled boxes, respectively. Hatched boxes indicate the presence of non-coding or coding sequence depending on the context (see text). Arrows, asterisks and pA indicate initiation sites, stop sites and polyadenylation signals, respectively. The sizes of each exon and intron, when known, are indicated above and below the genes, respectively.

Table 2 Gene expression analyses by RT-PCR

Gene	Brain	Testis	Prostate	Blood	Bone marrow	Placenta	HeLa	SJNB-7 ^b
<i>BCL-G</i>	—	+ ^a	+	—	+	—	—	—
<i>LOH3CR12</i>	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a
<i>LOH2CR12</i>	—	+	+	—	+	—	—	+
<i>LOH1CR12</i>	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a
<i>MKP7</i>	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a	+ ^a

^apresence of splicing variant(s) (see text). ^bSJNB-7: neuroblastoma-derived cell line.

domain, one of four BCL-2 homology domains (BH domain), is found in both isoforms and confers pro-apoptotic activity. The longer variant also possesses a BH2 domain, which negatively regulates this activity.²⁴ RT-PCR (Table 2) and Northern blot (data not shown) analyses revealed that while *BCL-G_S* is only expressed in the testes, *BCL-G_L* is expressed in many tissues including bone marrow, prostate, pancreas,

colon, but predominantly in the testes, which is in agreement with Guo *et al.*²⁴ Our Blast analyses and 5'RACE experiments suggest the presence of seven additional exons (Figure 2). The first three 5' upstream exons (exons 1A to 1C) are non-coding and various combinations of these exons were observed in the dbEST. Alternative splicing involving the four additional 3' exons (exons 7 to 10) could generate a

new isoform, termed *BCL-G_M* for median, whose 276 amino acid product would also lack box BH2. By RT-PCR, *BCL-G_M* was shown to be expressed only in the testes (Table 2).

MKP-7 has 7 exons (Figure 2) and codes for a new member of the dual-specificity phosphatase family that dephosphorylates MAP kinases.²⁵ This gene shows 55% sequence identity with *DUSP8/hvh-5*, a closely related dual-specificity phosphatase.²⁵ Northern blot hybridisation with a *MKP-7* probe

revealed the expression of two mRNA species of 4.0 and 6.0 kb in all tissues tested (Figure 3). The detection of weakly hybridising 3.0 and 5.5 kb bands in some tissues, especially in the brain, could be due to cross-hybridisation with the related *DUSP8/hvh-5* species, which correspond to the observed size and distribution of this mRNA.²⁶ By RT-PCR, we also observed a splicing variant, in which exon 4 is skipped and that is present in every tissue analysed by RT-PCR, although expressed at lower levels (Table 2). The predicted protein would be truncated and lack its phosphatase domain, keeping only the conserved *cdc25* domain putatively involved in protein-protein interactions.²⁷

LOH1CR12 has four exons (Figure 2) that encode a predicted 195 amino acid protein that shows high homology with hypothetical proteins from distant species: mouse BAB25030 (95% identity), *D.melanogaster* CG11802 (35% identity and 59% similarity) and *C.elegans* F59E12.11 (27% identity and 55% similarity) (Figure 4). The function of these proteins is not known, but they have some homology to one of the spectrins repeats of the TRIO protein and to an uncharacterised region from Mekk4 protein (Figure 4). Three major mRNA species at 1.2, 1.4 and 4.5 kb were detected in all tissues tested by Northern blot (Figure 3). Bands at 1.7 and 6 kb were also detected at lower levels. RT-PCR and sequencing experiments revealed the presence of a splicing variant skipping exon 2, thus explaining the 1.2 and 1.4 kb species. Interestingly this variant was not observed in the SJNB-7 cell line (Table 2). Both bands at 4.5 and 6 kb could be explained by the alternative use of termination sites (see below).

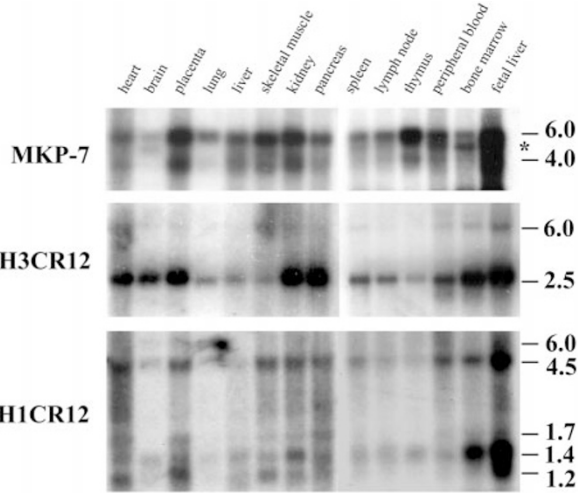


Figure 3 Northern blot analysis of *MKP-7*, *LOH1CR12* and *LOH3CR12* mRNAs. Multiple tissue Northern blot membranes were hybridised with probes specific for each gene. Asterisk indicates *DUSP8/hvh-5* cross-hybridisation (see text).

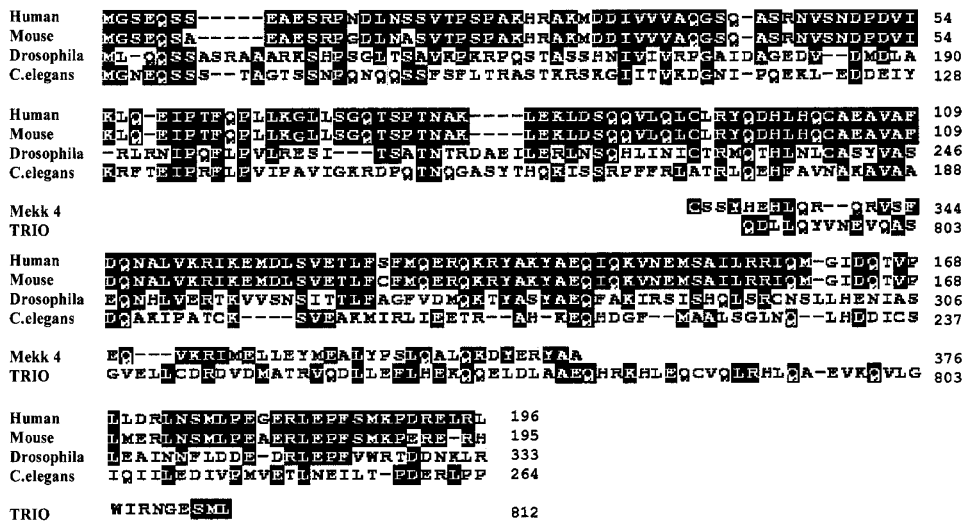


Figure 4 Amino acid alignment of hypothetical LOH1CR12 proteins from distant species. Human LOH1CR12 (GenBank acc. no. AAK71328), mouse BAB25030, *C.elegans* F59E12.11 (GenBank acc. no. T15266) and *D.melanogaster* CG11802 (GenBank acc. no. AAF48107) hypothetical proteins were aligned using ClustalW and manual refinements. Regions of high similarity with mouse Mekk4 (Swissprot no. O08648) and human Trio (Swissprot no. O75962) are also shown. Residues were shaded using BOXSHADE software according to their relative conservation (dark) or similitude (grey) across all species. Positions in the corresponding proteins are indicated on the right. The shorter human splice variant would lack amino acid 20 to 66.

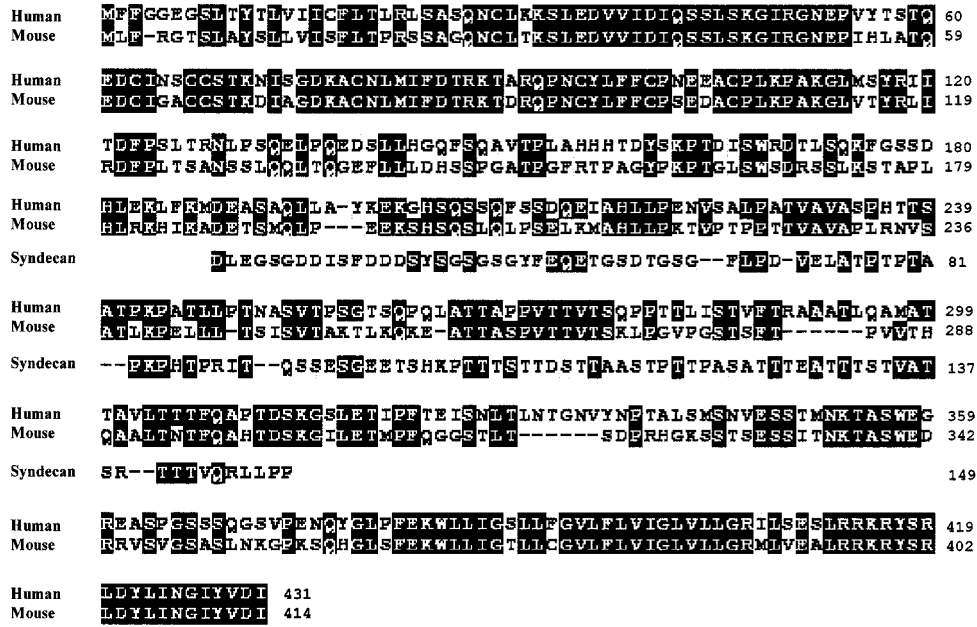


Figure 5 Amino acid alignment of hypothetical protein LOH3CR12 from mouse and human. Human LOH3CR12 (GenBank acc. no. BAB14621) and mouse RIKEN cDNA 9130403P13 (GenBank acc. no. NP_080621) proteins were aligned with syndecan domain (pfam 01034) using ClustalW and manual refinements. Dashes indicate gaps introduced to maintain optimal alignment. Residues were shaded using BOXSHADE software according to their relative conservation (dark) or similitude (grey) across all species. Positions in the corresponding human protein are indicated on the right. The shorter splice variant would lack amino acid 14 to 47.

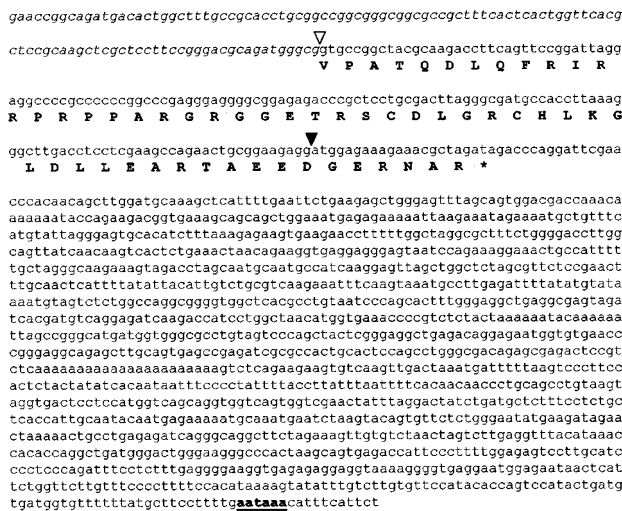


Figure 6 The partial cDNA sequence and predicted open reading frame of the LOH2CR12 gene. The observed exon/intron junction is indicated by a filled triangle. The open triangle indicates the position of the most upstream sequence that was characterised by 5'RACE and RT-PCR. Upstream genomic sequence is indicated in italic. The putative polyA signal is underlined.

The LOH3CR12 gene is constituted of five exons (Figure 2). The function of this gene product is unknown, but it possesses low homology to the syndecan, which is

characteristic of membrane heparan sulfate proteoglycans (Figure 5). Blast analysis also indicated low homology to hepatocyte growth factor activator inhibitor protein and glycoprotein 1a (data not shown). Homologues from various species were observed, including the mouse RIKEN cDNA 9130403P13 with 62% identity and 66% similarity at the protein level (Figure 5). We have also found a transcript variant (GenBank accession number AK023622) that is characterised by the presence of the 102 bp coding intron 2 that is translated in phase when left unspliced (hatched box in Figure 2). Both transcripts were present in every tissue tested by RT-PCR, the longest being the most abundant (data not shown).

The definitive structure of the LOH2CR12 gene is still unknown at this moment: only two exons were identified from the human EST database. All those ESTs, as well as the most upstream sequence obtained by 5'RACE analysis, stopped in a GC-rich region corresponding to the LOH1CR12 CpG island (Figure 6). By Blast analysis, a 500-bp mouse genomic fragment (clone Ti no.10921056) revealed clear homology to 5' exons from both LOH1CR12 and LOH2CR12. A common ORF could be constructed for LOH2CR12, although the start codon was not observed and is thus presumably located in another unidentified exon (Figure 6). No homology was found with proteins from other species. RT-PCR and sequencing experiments confirmed the gene prediction and very low levels of expression were observed in the bone marrow, prostate and testes. It also revealed that its

3' end corresponded to Unigene cluster Hs.67553 (Table 1). The weak expression could explain the absence of signal in the Northern blot analysis (data not shown).

Alternative termination

In previous studies, *ETV6* was shown to use alternative termination sites.²² Three of the EST clusters found in our study, Hs.169081, Hs.293972, Hs.146381, clearly correspond to the 3' ends of the *ETV6* transcripts of 2.4, 4.3 and 6.2 kb, respectively, when placed in the genomic context (Table 1). *LRP6* possesses two transcripts of 6 and 11 kb.¹³ Using the same reasoning, it is clear that the 3' end of clusters Hs.23672, which includes the coding sequence of *LRP6*, and Hs.41269, containing over 90 ESTs, correspond to the 3' end of the short and long transcripts, respectively. All the other identified transcribed sequences, with the exception of *LOH2CR12*, were characterised by at least two different mRNA species by Northern hybridisation (Figure 3) suggesting the presence of alternative splicing or alternative usage of termination sites. The latter is supported by the identification of large Unigene clusters with consensus 3' polyadenylation sites (but without coding potential), that could explain the 6 kb-*LOH3CR12*, 6 kb-*MKP-7* and 4.5- and 6 kb- *LOH1CR12* mRNAs respectively (see Table 1 and Figures 2 and 3). RT-PCR and sequencing experiments confirmed the existence of the longer transcripts for each gene (data not shown).

CpG islands

The presence of a CpG island usually indicates the 5' extremity of an housekeeping gene. The analysis of the genomic sequences with Grail/CpG predicted 4 CpG islands in the *D12S89-D12S358* interval. Three of them seem to be associated with the 5' ends of *LRP6*, *LOH3CR12* and *LOH1CR12*. *ETV6* also possesses a 5' CpG island,²² which is located outside the interval analysed (Figure 1). As predicted, all of them are expressed in every tissue analysed in contrast with *BCL-G* and *LOH2CR12* whose expression is more restricted (Table 2). The 5' end of *MKP-7* could not be identified since it lies in a gap between the BACs 253I19 and 180M15 (Figure 1), but Baens *et al.*¹⁰ placed a CpG island in that region. The other CpG island is associated with the pseudo-prothymosin gene and probably does not reflect the presence of a transcribed gene. Indeed, it has been shown that this CpG island is methylated *in vivo*.¹⁰

Discussion

Here we report the construction of a detailed transcriptional map of the putative chromosome 12p12 tumour suppressor locus using a combination of different approaches (comparative analysis, exon trapping and gene prediction algorithms). We identified seven distinct candidate transcriptional units, including *ETV6*, *LRP6*, *BCL-G* and *MKP-7* that were characterised previously.^{13,22,24,25} *LOH1CR1*, *LOH2CR12* and *LOH3CR12* are three novel genes characterized in this study.

ETV6 is a nuclear phosphoprotein that is widely expressed in all normal tissues²⁸ and plays an important role in angiogenesis²⁹ as well as in the normal development of the haematopoietic system.³⁰ Unlike most Ets-like proteins, *ETV6* acts as a transcriptional repressor recruiting proteins involved in the histone-deacetylase pathway.^{31,32} It can also block the transformation process induced by the over-expression of ras in cultured cells.¹⁹ *LRP6* is a member of the LDL-receptor family that has been recently shown to be an essential component of the Wnt pathway.¹⁴⁻¹⁶ This pathway is involved in many different cancers and developmental defects in humans.³³ *BCL-G* is a new member of the BCL-2 family that possesses proapoptotic activity.²⁴ *BCL-G* has multiple splicing variants generating three proteins of different lengths. *BCL-G_L* contains both conserved boxes BH3 and BH2, whereas *BCL-G_S* lacks box BH2. The latter is a more potent apoptosis inducer than the longer variant indicating that Box BH2 has an inhibitory role.²⁴ We found a third variant, *BCL-G_M*, that also lacks box BH2 thus predicting a function similar to *BCL-G_S*. Considering their function, *ETV6*, *LRP6* and *BCL-G* are good candidate genes, but the failure to detect any inactivating mutations in these genes in leukaemia patients does not support a role as classical tumour suppressor genes.^{10,20,24}

MKP-7 is a new member of the dual-specificity ser/thr and tyr phosphatases. Phosphatases play important roles in the regulation of intracellular signalling and have been associated with either oncogenic or tumour suppressor activity.³⁴ For instance, PTEN, another dual-specificity phosphatase, is often inactivated in human cancers, particularly endometrial carcinomas and glioblastomas, and germline mutations have been reported in patients affected with either of these two tumour predisposing syndromes.³⁵ *MKP-7* was recently shown to bind and inactivate p38 MAPK and JNK/SAPK involved in the transmission of cellular growth signals to the nucleus.²⁵

LOH1CR12 and *LOH3CR12* have no strong sequence similarity with proteins of known function, but some speculations can be made. *LOH1CR12* has homologues at least in mouse, *C.elegans* and *D.melanogaster* and Blast analysis revealed some sequence similarities at the protein level with a region encompassing one of the spectrin repeats of TRIO and other closely related proteins. TRIO is a guanine nucleotide exchange factor that regulates the actin cytoskeleton organization, cell motility and cell growth via activation of Rho GTPases.^{36,37} Spectrin repeats are usually found in proteins associated with the cytoskeleton. Although the specific function of *LOH1CR12* cannot be inferred, the fact that it is conserved throughout evolution points to an important role for this protein. Homologues of protein *LOH3CR12* are present in various mammals including the mouse, but no homologues have been identified in lower vertebrates. It possesses a syndecan domain and has also low homology with glycoprotein 1a and hepatocyte growth factor activator inhibitor. All of them are membrane-bound

or transmembrane proteins. Syndecans are transmembrane heparan sulphate proteoglycans that play an important role in cell-matrix and cell-cell interactions.³⁸ Finally, the structure of *LOH2CR12* is still incomplete and the characterisation of the remaining 5' exons will be necessary to infer a putative function for its encoded protein.

Previous expression analysis has demonstrated that *ETV6* and *LRP6* are present in a wide range of tissues.^{13,28} In this study, we showed that *LOH1CR12*, *LOH3CR12* and *MKP-7* also have a broad range of expression. Only *BCL-G* and *LOH2CR12* have a more restricted pattern of expression. *BCL-G* variants, *BCL-G_S* and *BCL-G_M* are only expressed in the testes while *BCL-G_L* is expressed predominantly in the testes and also at lower levels in other tissues including prostate, pancreas and the bone marrow, which is consistent with the work done by Guo *et al.*²⁴ The expression of *LOH2CR12* was observed only by RT-PCR analysis in certain tissues including testes, prostate and bone marrow. The fact that all the genes in the commonly deleted region from ALL patients are expressed in the bone marrow makes them interesting tumour suppressor gene candidates.

The observation that more EST clusters were identified than the number of genes in the region was puzzling at first. One explanation is the fact that most of the genes identified in our study seem to use alternative termination sites, giving rise in the database of multiple independent clusters, since most of the ESTs are primed at their 3' end. Whether the latter phenomenon is specific to the genomic context of this locus is not known, but many examples of alternative termination exist in the literature.³⁹ The exact role of the alternative 3'UTRs length is unclear since it does not affect the coding sequence, but it could be associated with the relative mRNA stability. For instance, the gene *eIF-2a* uses at least two polyA sites generating two common transcripts of 1.6 and 4.2 kb.⁴⁰ The smaller transcript is more readily translated *in vitro* while the longer transcript is more stable. The relative ratios of these transcripts vary with each tissue. The majority of the other clusters, if not all, correspond to artefacts that could be explained by internal or genomic polyT priming during the reverse transcription reaction. Only two small clusters in our analysis, represented by only one or two ESTs, could not be ruled out as artefacts using our criteria, indicating that we probably identified all the genes in the locus.

The construction of a transcription map of this locus and the identification of seven candidate genes provide the molecular framework to identify the tumour suppressor gene(s) at 12p12, a locus frequently deleted in haematological malignancies as well as in many different solid neoplasias.

Accession numbers

Sequence data from this article have been deposited in the EMBL/GenBank Libraries under accession numbers AY040274, AY037865, AY037866, AY037867 and AY038927.

Acknowledgments

We thank Zeina Saikali for critical reading of this manuscript. This work was supported by the Canadian Institutes of Health Research (CIHR). Daniel Sinnett is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ). Alexandre Montpetit and Gino Boily have received studentships from the CIHR and from the FRSQ, respectively.

References

- Berger R, Bernheim A, Le Coniat M *et al*: Abnormalities of the short arm of chromosome 12 in acute nonlymphocytic leukemia and dysmyelopoietic syndrome. *Cancer Genet Cytogenet* 1986; **19**: 281–289.
- Raimondi SC, Williams DL, Callihan T, Peiper S, Rivera GK, Murphy SB: Nonrandom involvement of the 12p12 breakpoint in chromosome abnormalities of childhood acute lymphoblastic leukemia. *Blood* 1986; **68**: 69–75.
- Chan LC, Kwong YL, Liu HW, Lee CP, Lie KW, Chan AY: Deletion 12p in de novo acute myeloid leukemia. An association with early progenitor cell. *Cancer Genet Cytogenet* 1992; **62**: 47–49.
- Raimondi SC: Current status of cytogenetic research in childhood acute lymphoblastic leukemia. *Blood* 1993; **81**: 2237–2251.
- Stegmaier K, Pendse S, Barker GF *et al*: Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* 1995; **86**: 38–44.
- Takeuchi S, Bartram CR, Miller CW *et al*: Acute lymphoblastic leukemia of childhood: identification of two distinct regions of deletion on the short arm of chromosome 12 in the region of TEL and KIP1. *Blood* 1996; **87**: 3368–3374.
- Baccichet A, Qualman SK, Sinnett D: Allelic loss in childhood acute lymphoblastic leukemia. *Leuk Res* 1997; **21**: 817–823.
- Baccichet A, Sinnett D: Frequent deletion of chromosome 12p12.3 in children with acute lymphoblastic leukaemia. *Br J Haematol* 1997; **99**: 107–114.
- Aissani B, Sinnett D: Fine physical and transcript mapping of a 1.8 Mb region spanning the locus for childhood acute lymphoblastic leukemia on chromosome 12p12.3. *Gene* 1999; **240**: 297–305.
- Baens M, Wlodarska I, Corveleyn A, Hoornaert I, Hagemeyer A, Marynen P: A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by *ETV6* and *CDKN1B*: hypermethylation of the *LRP6* CpG island in two leukemia patients with hemizygous del(12p). *Genomics* 1999; **56**: 40–50.
- Aissani B, Bonan C, Baccichet A, Sinnett D: Childhood acute lymphoblastic leukemia: is there a tumor suppressor gene in chromosome 12p12.3? *Leuk Lymphoma* 1999; **34**: 231–239.
- Kibel AS, Freije D, Isaacs WB, Bova GS: Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer* 1999; **25**: 270–276.
- Brown SD, Twells RC, Hey PJ *et al*: Isolation and characterization of *LRP6*, a novel member of the low density lipoprotein receptor gene family. *Biochem Biophys Res Commun* 1998; **248**: 879–888.
- Pinson KI, Brennan J, Monkley S, Avery BJ, Skarnes WC: An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* 2000; **407**: 535–538.
- Tamai K, Semenov M, Kato Y *et al*: LDL-receptor-related proteins in Wnt signal transduction. *Nature* 2000; **407**: 530–535.
- Wehrli M, Dougan ST, Caldwell K *et al*: Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* 2000; **407**: 527–530.
- Golub TR, Barker GF, Lovett M, Gilliland DG: Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* 1994; **77**: 307–316.
- Van Rompaey L, Dou W, Buijs A, Grosveld G: Tel, a frequent target of leukemic translocations, induces cellular aggregation and influences expression of extracellular matrix components. *Neoplasia* 1999; **1**: 526–536.

- 19 Fenrick R, Wang L, Nip J *et al*: TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol* 2000; **20**: 5828–5839.
- 20 Stegmaier K, Takeuchi S, Golub TR, Bohlander SK, Bartram CR, Koeffler HP: Mutational analysis of the candidate tumor suppressor genes TEL and KIP1 in childhood acute lymphoblastic leukemia. *Cancer Res* 1996; **56**: 1413–1417.
- 21 Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–4680.
- 22 Baens M, Peeters P, Guo C, Aerssens J, Marynen P: Genomic organization of TEL: the human ETS-variant gene 6. *Genome Res* 1996; **6**: 404–413.
- 23 Roest Crollius H, Jaillon O, Bernot A *et al*: Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet* 2000; **25**: 235–238.
- 24 Guo B, Godzik A, Reed JC: Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. *J Biol Chem* 2001; **276**: 2780–2785.
- 25 Tanoue T, Yamamoto T, Maeda R, Nishida E: A Novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem* 2001; **276**: 26629–26639.
- 26 Martell KJ, Seasholtz AF, Kwak SP, Clemens KK, Dixon JE: hVH-5: a protein tyrosine phosphatase abundant in brain that inactivates mitogen-activated protein kinase. *J Neurochem* 1995; **65**: 1823–1833.
- 27 Denu JM, Dixon JE: Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol* 1998; **2**: 633–641.
- 28 Poirel H, Oury C, Carron C *et al*: The TEL gene products: nuclear phosphoproteins with DNA binding properties. *Oncogene* 1997; **14**: 349–357.
- 29 Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, Orkin SH: Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J* 1997; **16**: 4374–4383.
- 30 Wang LC, Swat W, Fujiwara Y *et al*: The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* 1998; **12**: 2392–2402.
- 31 Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J: TEL is a sequence-specific transcriptional repressor. *J Biol Chem* 1999; **274**: 30132–30138.
- 32 Chakrabarti SR, Nucifora G: The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun* 1999; **264**: 871–877.
- 33 Polakis P: Wnt signaling and cancer. *Genes Dev* 2000; **14**: 1837–1851.
- 34 Dahia PL: PTEN, a unique tumor suppressor gene. *Endocr Relat Cancer* 2000; **7**: 115–129.
- 35 Bonneau D, Longy M: Mutations of the human PTEN gene. *Hum Mutat* 2000; **16**: 109–122.
- 36 Debant A, Serra-Pages C, Seipel K *et al*: The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc Natl Acad Sci USA* 1996; **93**: 5466–5471.
- 37 Seipel K, Medley QG, Kedersha NL *et al*: Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth. *J Cell Sci* 1999; **112**: 1825–1834.
- 38 Rapraeger AC: Molecular interactions of syndecans during development. *Semin Cell Dev Biol* 2001; **12**: 107–116.
- 39 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.
- 40 Miyamoto S, Chiorini JA, Urcelay E, Safer B: Regulation of gene expression for translation initiation factor eIF-2 alpha: importance of the 3' untranslated region. *Biochem J* 1996; **315**: 791–798.