

ARTICLE

A comparison of genomic structures and expression patterns of two closely related flanking genes in a critical lung cancer region at 3p21.3

Tineke Timmer¹, Peter Terpstra¹, Anke van den Berg¹, Patrick MJF Veldhuis¹, Arja Ter Elst¹, Gerassimos Voutsinas¹, Miriam MF Hulsbeek¹, Tineke G Draaijers¹, Maaïke WG Looman¹, Klaas Kok¹, Susan L Naylor² and Charles HCM Buys¹

¹Department of Medical Genetics, University of Groningen, Groningen, The Netherlands

²Department of Cellular and Structural Biology, University of Texas Health Science Center, San Antonio, Texas, USA

In the search for a tumour suppressor gene in the 3p21.3 region we isolated two genes, *RBM5* and *RBM6*. Gene *RBM5* maps to the region which is homozygously deleted in the small cell lung cancer cell line GLC20; *RBM6* crosses the telomeric breakpoint of this deletion. Sequence comparison revealed that at the amino acid level both genes show 30% identity. They contain two zinc finger motifs, a bipartite nuclear signal and two RNA binding motifs, suggesting that the proteins for which *RBM5* and *RBM6* are coding have a DNA/RNA binding function and are located in the nucleus. Northern and Southern analysis did not reveal any abnormalities. By SSCP analysis of 16 lung cancer cell lines we found only in *RBM5* a single presumably neutral mutation. By RT-PCR we demonstrated the existence of two alternative splice variants of *RBM6*, one including and one excluding exon 5, in both normal lung tissue and lung cancer cell lines. Exclusion of exon 5 results in a frameshift which would cause a truncated protein of 520 amino acids instead of 1123 amino acids. In normal lung tissue, the relative amount of the shorter transcript was much greater than that in the lung tumour cell lines, which raises the question whether some tumour suppressor function may be attributed to the derived shorter protein.

Keywords: DNA/RNA binding proteins; 3p21 genes; alternative splicing; lung cancer; tumour suppressor

Introduction

According to results of cytogenetic deletion analysis, loss of heterozygosity studies and chromosome transfer

experiments, the distal region of 3p21 might be the location of a kind of broad spectrum tumour suppressor gene involved in the development of a variety of cancers.¹ We described earlier a homozygous deletion of 440 kb in the 3p21.3 region of a small cell lung cancer (SCLC) cell line, GLC20.² Because this deletion has a 370 kb overlap with the 3p homozygous deletions of two other SCLC cell lines^{3,4} and affects a region within a 2 Mb fragment showing tumour suppressor activity,⁵

Correspondence: CHCM Buys, Dept of Medical Genetics, A Deusinglaan 4, 9713 AW Groningen, The Netherlands.
Tel: 31 50 363 2925; Fax: 31 50 363 2947; E-mail: M.G.Dijkstra@med.rug.nl

Received 16 December 1998; accepted 8 February 1999

we used the region corresponding to the deletion as a starting point for gene isolation. Here we describe the genomic structures and expression patterns of two genes with RNA/DNA binding features which we found in this region and which show an overall similarity of 30% at the derived amino acid level.

Materials and Methods

Cell Lines

The lung cancer cell lines analysed in this study have been described previously.⁶

Isolation of YAC DNA

Culturing of yeast cells, generation of agarose plugs and isolation and digestion of YAC DNA were performed as previously described.⁷⁻⁹

Direct Screening of cDNA Libraries

The oligo-dT primed NALM-6 pre-B-cell library in lambda ZAP was kindly provided by Dr André Bernards (Charlestown, Mass., USA). An oligo-dT primed human lung library in lambda DR2 was purchased from Clontech (Palo Alto, USA). Screening of the cDNA libraries was performed by direct hybridisation with YAC DNA as we have described earlier.⁹

Screening of PAC Libraries

A PAC was obtained by screening the RPCI-1 and RPCI-6 libraries (Genome Technology Centre, Leiden, The Netherlands) cloned in the vectors pCYPAC2 and pPAC4 respectively with the *RBM6* cDNA as probe.

Southern and Northern Analyses

DNA analysis was performed using standard protocols. Total RNA was isolated using RNazol (Campro Scientific, Veenendaal, NL) and analysed according to standard protocols. Multi-tissue polyA⁺-filters were purchased from Clontech (Palo Alto, USA).

Sequence Analysis

cDNA clones were sequenced using an automatic fluorescent sequencer (ALF; Pharmacia, Uppsala, Sweden) with the Thermosequenase¹⁰ cycle sequencing kit¹¹ from Amersham (Buckinghamshire, UK) or using T7 polymerase with fluorescent dATP.¹² Small fragments were sequenced manually using T7-polymerase sequencing kit (Sequenase, Pharmacia, Uppsala, Sweden).

Rapid Amplification of cDNA Ends (RACE)

RACE-PCR was performed using human lung 5'RACE-ready total cDNA and human lung Marathon-Ready total cDNA (Clontech, Palo Alto, USA) according to the protocols provided by the manufacturer.

Determination of the Intron-Exon Structure

Intron/exon structures were determined by Vectorsite-PCR¹³ and by comparing the cDNA sequences to publicly available sequences of the 3p21 sequencing project from the Washington University Genome Sequencing Center.

SSCP Analysis

Exons or exon fragments were amplified using standard protocols. Electrophoresis was performed under two different conditions; a 6% PAA gel with 10% glycerol and a PAA MDE gel (AT Biochem, Malvern, USA).

RT-PCR

RNA was isolated from lung cancer cell lines with the use of the RNAeasy Mini Kit (Qiagen Inc, Santa Clarita, USA); cDNA was synthesised with the use of the Ready To Go You-Primed First Strand kit (Pharmacia, Uppsala, Sweden) and random primers. cDNA fragments were amplified using standard protocols and primers Gfra1F in exon 4 (5' tct taa tgc ttg gac tcc) and Gfra1R2 in exon 7 (5' ggc ttg ata cat act cca gg).

Results

Isolation of RBM5 and RBM6

The overlapping YACs 181H1 and 181B3 originate from the chromosomal segment corresponding to the GLC20 deletion and its vicinity.² Using YAC 181H1 DNA as probe in a direct hybridisation to the pre-B-cell cDNA library, followed by further screening of this library, we isolated several clones together resulting in a cDNA with a total length of 3.1 kb. In a northern blot analysis of RNA from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas two bands were detected: one band of 7.4 kb and one of 3.5 kb. Expression was seen in all tissues investigated. We named the gene concerned *RBM5* (standing for RNA Binding Motifs) for reasons described below.

Direct hybridisation of the lung cDNA library with YAC 181B3 DNA, followed by 5'-RACE PCR and 5'Marathon-RACE PCR, resulted in the isolation of several clones representing together a cDNA with a total length of 3.7 kb. In a northern blot analysis of the same tissues that expressed *RBM5*, three different bands were observed of 4.7 kb, 3.9 kb and 2.1 kb, respectively. We named the gene concerned *RBM6*, again for reasons given below.

In Southern blot analysis, *RBM5* did not hybridise to DNA from GLC20, confirming that the gene maps to the region homozygously deleted in this SCLC cell line. Although also a cDNA clone representing the 3' end of *RBM6* did not hybridise to DNA from GLC20, implying that this part of the gene is in the deletion, a cDNA clone representing the 5' end of *RBM6* did hybridise to GLC20, indicating that the *RBM6* gene overlaps one of the deletion breakpoints of GLC20. Pulsed field gel analysis was used to map the genes in the known restriction map.^{2,14} Both *RBM5* and *RBM6* hybridised to a 650 kb *NotI* fragment, a 380 kb *NruI*

fragment and a 350 kb *NruI*X*NotI* fragment, confirming that they map between the *NotI*-containing clone DD1 and *UBE1L*, and that *RBM6* overlaps the telomeric breakpoint of *GLC20* (Figure 1).

Sequence Analysis of *RBM5* and *RBM6*

The *RBM5* cDNA clones defined an open reading frame of 2460 nucleotides. In this frame, a stop codon is present at bp positions 2594–2596 and is followed by an untranslated region of 504 nucleotides. A likely ribosome-binding site¹⁵ precedes a first methionine codon at bp positions 149–151. Taking this as the transcription initiation site, a protein product of 815 amino acids can be inferred. From the lengths of the cDNA (3.1 kb) and the transcript on the northern blot (3.5 kb), some 400 bp appear to be missing at the 5' end, presumably either due to alternative processing or untranslated regions.

The sequence of *RBM6* contains an open reading frame of 3375 nucleotides. A likely ribosome-binding site precedes an initiating methionine codon at bp positions 134–136. A stop codon is present at bp positions 3503–3505 and is followed by an untranslated region of 120 nucleotides. Taking the methionine at position 134 as the start of transcription, a protein product of 1123 amino acids can be inferred. From the lengths of the cDNA (3.7) and the transcript on the northern blot (3.9 kb), some 200 bp appear to be missing at the 5' end, again presumably either due to alternative processing or untranslated sequences.

Sequence analysis of *RBM5* and *RBM6* revealed that the two genes show a 30% similarity at the derived amino acid level. A homology search of DNA sequences in the GenBank databases with the BlastN program revealed that *RBM5* is identical (differing by only 4 bp) to *Luca-15*, a gene recently isolated from the 3p21.3 region.¹⁶ *RBM5* also shows a similarity to *DXS8237E*¹⁷/*KIAA0122*¹⁸ on Xp11.23 of 51% at the amino acid level and a comparable similarity to the rat

equivalent *SI-1*.¹⁹ A homology search for *RBM6* revealed that this gene is identical (differing by only 3 bp) to the cDNA sequence *NY-LU-12*,²⁰ (differing by only 4 bp) to the cDNA sequence *DEF-3* (accession no. AF069517) and to the cDNA sequence *g16* (accession no. U50839). The mouse homolog *Def-3* has a similarity to *RBM6* of 82% at the amino acid level.

The derived amino acid sequences of *RBM5* and *RBM6* share functional motifs, as also reported by Güre *et al.*²⁰ These include two RNA binding motifs,²¹ two zinc finger motifs (one of the C₂C₂ type and one of the C₂H₂ type) and a bipartite nuclear signal.²² The RNA binding motifs in *RBM6* are, however, much less conserved than in *RBM5* (Figure 2). Moreover, the derived *RBM6* protein contains in its first 290 amino acids 20 repeats of 6 amino acids separated by intervals of 4 to 6 amino acids. The consensus sequence of the repeats is D-F/Y-R-G/D-R/G-D/E.

Intron/Exon Structure Comparison

The intron/exon boundaries of *RBM5* were obtained by comparing the cDNA sequence of *RBM5* with the genomic sequences from the Washington University Genome Sequencing Center. The *RBM5* gene has 25 exons ranging in size from 61 bp (exon 8) to 627 bp (exon 25) (Table 1). The intron/exon boundaries of *RBM6* were obtained by vectorette PCR with the use of the PAC isolated by hybridisation to the *RBM6* cDNA clone (exons 1–5, 8, 9, 12–20) and by comparing the cDNA sequence of *RBM6* with the genomic sequence from the Washington University Genome Sequencing Center (exons 6, 7, 10, 11). The *RBM6* gene has 20 exons ranging in size from 43 bp (exon 11) to 1279 bp (exon 2) (Table 2). From the H3938P1.contig 13 sequence from the Washington University Genome Sequencing Center it can be concluded that there is a large intron (> 34 kb) between exon 5 and exon 6 of *RBM6*. By comparing the amino acid translations of the different exons, the *RBM5* exons 4, 6, 7, 8, 10, 11, 22, 23,

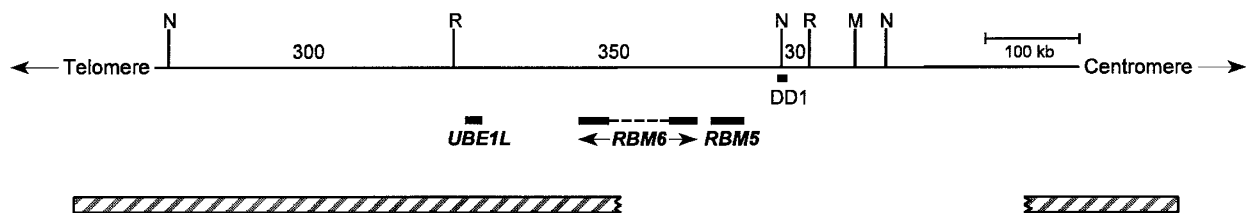


Figure 1 Long range map derived from a pulsed field gel electrophoresis analysis² of the 3p21.3 region, showing the position of the *UBE1L* gene and *DD1* probe and the newly isolated genes *RBM5* and *RBM6*. Restriction sites for *MluI*, *NotI* and *NruI* are indicated by *M*, *N*, and *R*, respectively. Numbers indicate the length of the restriction fragments in kb. The approximate position of the region homozygously deleted in *GLC20* is indicated by the deletion in the hatched bar.

and 24, respectively, appeared to have a high similarity (more than 25%) to the *RBM6* exons 3, 5, 6, 7, 9, 10, 17, 18, and 19, respectively (Figure 3).

By PCR with primer sets specific for the different exons of *RBM6*, we could determine that the homozygous deletions of the SCLC cell lines GLC20 and NCI-H740 include the larger part of *RBM6*. GLC20 contains exons 1–5 and NCI-H740 contains exons 1–6.

Mutation Analysis of *RBM5* and *RBM6*

Fourteen SCLC lung cancer cell lines and one non-SCLC cell line have been subjected to northern and Southern analysis using the *RBM5* and *RBM6* clones as probes. No variations in expression, nor deletions or other rearrangements were found. In a search for possible small mutations, SSCP analysis of *RBM5* and *RBM6* was performed on 14 SCLC cell lines and two non-SCLC cell lines. For *RBM5*, band shifts were observed for exons 3, 4, 12 and 22 in several SCLC cell lines as well as in unrelated normal control samples. In GLCA2, an adenocarcinoma cell line, a band shift, not seen in DNA from blood lymphocytes from 70 controls,

was observed for exon 15. Sequence analysis of the aberrant band revealed a change at position 1341 (codon 398) of GGC to GTC, resulting in an amino acid change of glycine to valine.

Because the intron–exon boundaries of exons 1 to 5 of *RBM6* had not yet been determined when the SSCP analysis was performed, these exons were amplified with overlapping exon-specific primer sets using cDNA derived from total RNA as a template. Band shifts for exons 2, 15 and 16 were seen in some of the cell lines. Band shifts for exons 15 and 16 were, however, also observed in unrelated normal controls. The band shift for exon 2 was only seen in one of the non-SCLC cell lines and not in 70 controls. Sequence analysis of the aberrant band revealed a codon change of GAG to GAA at position 340 (codon 69) resulting in a silent glutamic acid mutation.

Alternative Splicing in *RBM6*

An earlier version of *g16* in GenBank missed 74 bp with respect to our *RBM6* sequence. When we had determined the intron–exon boundaries of *RBM6*, it

	A	B	C
Rat S-1	<u>RGFAFVEF</u> (pos.:94-101)	<u>CNKCGVQNFKRREKCFKC</u> (pos.:142-159)	<u>RGFAFIQL</u> (pos.:266-273)
Human DXS8237E	<u>RGFAFVEF</u> (pos.:252-259)	<u>CNKCGVQNFKRREKCFKC</u> (pos.:300-317)	<u>RGFAFIQL</u> (pos.:424-431)
Human <i>RBM5</i>	<u>RGFAFVEF</u> (pos.:140-147)	<u>CNKCLNLFNFKRRLKCFRC</u> (pos.:187-204)	<u>RGFAFVQL</u> (pos.:274-281)
Human <i>RBM6</i>	<u>YGYVCVEF</u> (pos.:498-505)	<u>CKRCKANIGGHRSSCSFC</u> (pos.:542-559)	<u>HTYGFIDL</u> (pos.:699-706)
Mouse Def3	<u>YGYVCVEF</u> (pos.:128-135 ^a)	<u>CKRCKASTGGHQSSCSFC</u> (pos.:172-189 ^a)	No Data
	D	E	
Rat S1-1	<u>CLLQRRQFSPKEALIRHQQLSGLH</u> (pos.:683-706)	<u>RREKYGIPEPPEPKRRK</u> (pos.:743-759)	
Human DXS8237E	<u>CLLQRRQFSPKEALIRHQQLSGLH</u> (pos.:841-864)	<u>RREKYGIPEPPEPKRRK</u> (pos.:901-917)	
Human <i>RBM5</i>	<u>CLLQRRQFPNKDALVRHQQLSDLH</u> (pos.:649-672)	<u>RREKYGIPEPPEPKRRK</u> (pos.:708-724)	
Human <i>RBM6</i>	<u>CLLQRRQFPNKVLIKHQQLSDLH</u> (pos.:957-980)	<u>RREKLSFDSPEPKRIK</u> (pos.:1016-1032)	
Mouse Def3	No Data	No Data	

Figure 2 Amino acid sequence motifs found in translations of the *RBM5* gene family. The DXS8237E nucleotide sequence (accession no. U35373) is identical to the KIAA0122 sequence (accession no. D50912), except for an insertion of 4 bp leading to a frameshift in U35373 with respect to the other sequence. For this figure the 5' incomplete translation of KIAA0122 has been taken. The rat S1-1 protein is a translation of accession no. D83948. The mouse Def-3 protein fragment is a translation of an overlapping cDNA (accession no. X96701) and a mouse EST (accession no. AA268319). Accession no. X96701 has been corrected for obvious frameshifts by deleting 3 bp. Programs used: PSORT²⁸ and PROSITE.²⁹ Numbers in parentheses give the amino acid position in the protein sequences. A: First RNA binding motif RNPI. Similarities to consensus sequence are underlined.²¹ B: Potential zinc finger domain of type C4. All sequences fit the consensus sequence C x₂ C x₁₀ C x₂ C; C: Second RNA binding motif RNPI. Similarities to consensus sequence are underlined.²¹ D: Potential zinc finger domain of type C2H2. All sequences fit the consensus sequence C x₂ C x₁₂ H x₆ H; E: Bipartite nuclear signal. All sequences fit the consensus sequence of two basic residues, 10 spacer residues, and a sequence of five residues of which at least three have a basic character.²² ^aPosition numbering starts from the beginning of the known sequence, which does not yet contain the transcription initiation site.

Table 1 Intron–exon boundaries of the *RBM5* gene

(1; 1352)	taatctctctgtctcctag	AAAAAATAAA	(1; >95)	TTCCTTGGG	gtaagtgcggcgctgtctg
(2; 1592)	aaaattatttccctggcag	AGTGAGTAGA	(2; 70)	CAGACAAAAG	gtaagttactacagtacgtg
(3; 1504)	ttttgtttattgtaactcag	AGAGAGCGTG	(3; 166)	CAGTCCAGAG	gtgagtgaccagcggctgta
(4; 6094)	taatgtaaaccctgtgcag	ATTTCGAGAAA	(4; 156)	AGAGAGCGAT	gtaaggggaaatgacagtta
(5; 480)	tatctgaatgctgtctctag	GTGTAAGCCG	(5; 70)	AGGAAAACAG	gtgagagcttctgtagtcc
(6; 2473)	aattactcttttcttacag	AAAAAGTTGG	(6; 74)	AGCCAATCAG	gttgcttactcacaagtc
(7; 1081)	gataaaactttttttcag	TGCTGCCTTA	(7; 84)	TTGTAACAAG	gtaagcatatgttctccca
(8; 768)	tggatcaacttccctctag	ACTCTGAACA	(8; 61)	GACAAGTTTG	gtaagactggattcagctgt
(9; 406)	ttgtctcatttaccctag	CGATCATTCT	(9; 66)	TACTGTGATA	gtaagttcatacacgatctt
(10; 1054)	tccttcacattctcctcag	GATGCTTCTC	(10; 161)	CTCTGCAATG	gtgaggttctcatcgattct
(11; 617)	ttgatctttgtggtttcag	AGACTTGGTC	(11; 98)	GTGCCAGAAA	gtgagtggtctcattgtcct
(12; 499)	ctcattttgatgtttgtag	TCTCAAAGTG	(12; 88)	ATCCACCCAG	gtaagatcgaggatcttttt
(13; 84)	aatggtggactttctcag	TATTCACAGG	(13; 78)	ATATGCTCAG	gtaggtagatttttagcagea
(14; 1298)	acactgtgtttatcatcag	GCACAGCAGT	(14; 73)	TCTGCATCAG	gtagtaaactcatctcct
(15; 690)	gaagctgctgttcttccag	ACTGAGGAAG	(15; 86)	TGGCTCTCCG	gtaatcctgtttgcctatat
(16; 215)	gctggattttgtttgcag	CAGTACCTGA	(16; 85)	ACCAAATATG	gtaagccaacctcatgggg
(17; 2602)	ttttgtttgttttctag	TACTACTATA	(17; 92)	CAACTCGCAA	gtaaatgtgctgctttcctc
(18; 405)	tggtttgtttgacattag	ATTGCCAAAG	(18; 162)	AGCCCAGCAG	gttagaacatgacccatatt
(19; 80)	gcttgcctctgtttccag	GGAGCCTTAG	(19; 150)	TGAGAAGAAG	gtaatagcaggaatggccag
(20; 1176)	tgctcctctttgtctccag	AGGGGTCTGG	(20; 72)	TCCCCTCAA	gtaagggagtagaccaccagt
(21; 298)	cttcaaatctttgtgtag	CAAAACATGG	(21; 180)	CCTTCACAAG	gtggccatgcttctgagctg
(22; 1093)	gtgttttctcccatgtag	ATGAAATACC	(22; 75)	GGAGAGAGAG	gtgaatgggaaactgtgcca
(23; 78)	tggtgtttgccactggcag	GAATTACGAG	(23; 98)	CCGGCACTGT	gtatgtgatgtgcacatttt
(24; 952)	cagtctctgtcttccag	GCTCAAGTTC	(24; 130)	TCCCATGAG	gtaagcagtggtgtagcaggc
			(25; 627)	ATACCTTGAG	

Intron sequences are indicated by lower case letters; exon sequences by capitals. Numbers of the introns (or exons) and their total length (including the sequences shown) are indicated in parentheses. The length of the introns was determined by comparison with the genomic sequence of cosmid Luca-22 and H3938P1.contig13 as obtained from the Washington University Genome Sequencing Center.

Table 2 Intron–exon boundaries of the *RBM6* gene

(1; n.d.)	cactgtctttgtttctacag	TGGGAGCCAA	(1; >177)	GACCTTTTCG	gtaagtctcaaattgaaat
(2; n.d.)	tttttcttcaatttaag	GATCAAGATT	(2; 1279)	GGACCTTCAG	gtatgttgatggggtgatt
(3; n.d.)	taaatgccatattttgacag	ATTCTTAATG	(3; 90)	AAAAGAAGAG	gtaagcagatgcttctctcc
(4; n.d.)	tatttaaactgctgcccag	GTTACGACTA	(4; 70)	TATAACACAG	gtgaagtttctgacttcca
(5; n.d.)	ttctgtctttgtctttacag	GGAAGCTTAA	(5; 74)	GGCCAACCAG	gttgctttatacttcggta
(6; 6013)	gatttcttcttcttctacag	TGTAAGGCAA	(6; 75)	CTGCAAACGA	gtaagtaccaagaatccctt
(7; 3334)	ctttgtttctgttccactag	TGACAGAGGC	(7; 61)	CCAAGAGAAG	gtgagtgccgaaagtgttag
(8; 398)	cccatttatggttgattcag	CTATCATGCT	(8; 276)	GAGAGCAAAA	gtaagtagttgtcagggca
(9; 1086)	aagtgtgtctggttctatag	GAAGCTCTTC	(9; 161)	CTCCCATGCG	gtgagtttctccaccttgg
(10; 1208)	tggtgtttatctgttgcag	AAATGATTCT	(10; 98)	GAAAACGAAG	gtaaggcagaagggtgagga
(11; 141)	aacttgccctttctgtgatag	GGTAAAAAAT	(11; 43)	TTACTATCAG	gtaggctgaacaggtgggg
(12; 240)	tctgatctgttacccttag	CATCATCTGA	(12; 82)	GGACAACAGT	gtaagtaacctttgtttat
(13; 416)	tgttccaactgtatgcag	CAAGAAGTCT	(13; 86)	CAATACCCAG	gtgagttggggctttttt
(14; 2921)	gttacctctggtttgatcag	TTCAGGAAA	(14; 147)	AGTGACGAGG	gtaagaggaaatgtaattt
(15; 1115)	ccctggccttctcctccag	CCTAAAGTGG	(15; 96)	TCCCCCTCCA	gtaagaccaacattgatccc
(16; 2197)	agaattgcctctcttctcag	CAAAACCTGG	(16; 261)	CCTGCACAAG	gtattaggggaaggagctat
(17; 1691)	gctcttctgcttaccacag	GGAAAGTTTA	(17; 75)	AGAACGAGAG	gtaaaccttggtagcctatt
(18; 4649)	ggtgtctgcttctttgcag	TGATCGTAAA	(18; 98)	AAACTGACAG	gtaagccaggaactcttcat
(19; 1677)	agttgtctttgttttacag	GCTGAAGGCC	(19; 130)	ATCAGAGGAG	gtaaaatggtttccatcttt
n.d. = not determined			(20; 241)	TCAATGTGA	ttatctttttgtatttttt

Intron sequences are indicated by lower case letters; exon sequences by capitals. Numbers of the introns (or exons) and their total length (including the sequences shown) are indicated in parentheses. The length of the introns was determined by comparison with the genomic sequence of H3938P1.contig13 as obtained from the Washington University Genome Sequencing Center.

appeared that these 74 bp coincide with exon 5. Using primers on each side of this exon, in exons 4 and 7, RT-PCR on cDNA from normal lung tissue specimens, SCLC cell lines and one non-SCLC cell line, gave in all cases a 260 bp product including exon 5 and a 186 bp product excluding this exon, as verified by sequence analysis. It also appeared, however, that there was a difference in the intensity ratio of the two PCR products between normal lung cDNA and SCLC cDNA. Whereas in normal lung tissue, the shorter product had half the amount of the longer one, 10 of the 12 SCLC cell lines and the sole non-SCLC cell line analysed expressed substantially less of the shorter product than of the longer product (Table 3, Figure 4).

Discussion

We isolated two genes, *RBM5* and *RBM6*, which both map to the 3p21.3 region. *RBM5* has a homolog on the X chromosome, *DXS8237E* at Xp11.23. For this gene a rat homolog, *SI-1*, is known. Functionality tests of *SI-1* revealed that its protein product, synthesised by *in vitro* translation, binds to RNA homopolymers, with a preference for G and U polyribonucleotides and little affinity to poly(A).¹⁹ The amino acid sequences derived from the genes belonging to the *RBM5* family (*RBM5*, *DXS8237E* and *SI-1*) show two RNA-binding motifs, two potential zinc finger domains and a bipartite nuclear signal (Figure 2), suggesting that the protein products are located in the nucleus and have a RNA/DNA binding function.

Sequence comparison also led to the conclusion that *RBM6* is another member of the *RBM5* family. At the

amino acid level both genes show 30% identity. They share some clusters of exons that have a high sequence similarity (Figure 3). Whereas the *RBM6* protein also contains the two zinc finger domains and the nuclear signal present in the *RBM5* protein, its RNA-binding motifs are much less conserved. The octamers at the corresponding positions in *RBM6* have only five amino acids that fit the RNP-1 pattern (Figure 2). The N-terminal part of *RBM6* (aa1-300), contains 20 repeats of six amino acids. The consensus sequence D-F/Y-R-G/D-R/G-D/E of the repeats, separated by four to six amino acid intervals, is not found in other known genes. A repeat sequence with some similarity to the consensus repeat sequence of the *RBM6* repeat is, however, found six times in the C-terminal amino acid sequence of Nopp44/46, a tyrosine-phosphorylated protein of *Trypanosoma brucei* (Figure 5). Nopp44/46 localises to the nucleolus and *in vitro* binding experiments have demonstrated that this protein is capable of binding nucleic acids with a preference for poly (U).²³ Therefore, the function of *RBM6* might still be comparable with the function of *RBM5*.

The sequence of *RBM6* is identical to the sequence of the *NY-LU-12* gene, cloned by screening a recombinant lung cancer cDNA expression library with autologous patient serum.²⁰ Patient sera showed anti-*NY-LU-12* seroreactivity in two of 21 allogenic lung cancer patients. A mutation search was performed on DNA from a sole anti-*NY-LU-12* seroreactive patient, but no *NY-LU-12* mutations could be detected.

Although *RBM5* is located in the region which we consider critical for the development of lung cancer, northern and Southern analysis of *RBM5* in a number of lung cancer cell lines did not reveal any aberrant

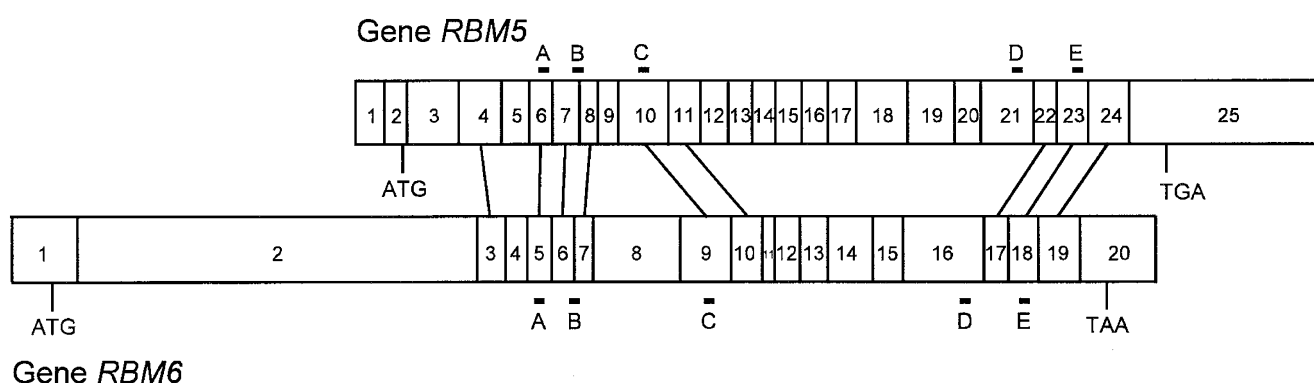


Figure 3 Schematic representation of the exon structures of genes *RBM5* and *RBM6*. Numbers indicate exon numbers. ATG indicates the start methionine, TGA and TAA indicate the stop codons. The bars above gene *RBM5* and below gene *RBM6* indicate the different protein motifs. A and C: RNA binding motif RNP1; B: Potential zinc finger domain of type C4; D: Potential zinc finger domain of type C2H2; E: Bipartite nuclear signal. The lines connect exons in *RBM5* and *RBM6* that share high (over 25%) amino acid similarity.

patterns. The 7.4 kb band detected on northern blots in addition to the 3.5 kb band most likely represents a premessenger RNA. SSCP analysis of *RBM5* revealed a single mutation in an adenocarcinoma cell line substituting valine for glycine at amino acid position 398. Because both amino acids belong to the same non-polar group, this change will most likely not affect the function of the protein. In blood lymphocytes from 70 control individuals no such mutation was found.

The 3' end of *RBM6* (exons 6–20) also maps to this critical region homozygously deleted in GLC20, but the

Table 3 Ratio between product intensities of the different splice variants of *RBM6* in cDNAs from different tissues and cell lines

Tissue/cells	Ratio of fluorescent intensities of 186 bp and 260 bp PCR product	Lung cancer cell lines	Ratio of fluorescent intensities of 186 bp and 260 bp PCR product
Normal lung 1	0.46	GLC 1	0.13
Normal lung 2	0.62	GLC 2	0.35
Normal lung 3	0.68	GLC 3	0.08
Normal lung 4	0.63	GLC 4	0.09
Normal lung 5	0.23	GLC 7	0.17
Normal lung 6	0.61	GLC 8	0.36
Normal kidney	0.24	GLC 16	<0.05
Cultured fibroblasts	<0.05	GLC 34	0.05
Foetal brain	0.58	GLC 35	0.10
Spinal cord	0.41	GLC 36	<0.05
Skeletal muscle	0.45	GLC 42	<0.05
Blood lymphocytes	0.46	GLC 45	<0.05
		GLC P1 (non-SCLC)	<0.05

Gels were scanned and peak areas were measured and normalised for bp length. The ratios are the average of three independent experiments (except in the normal lung 5 specimen where the average of two experiments is given).

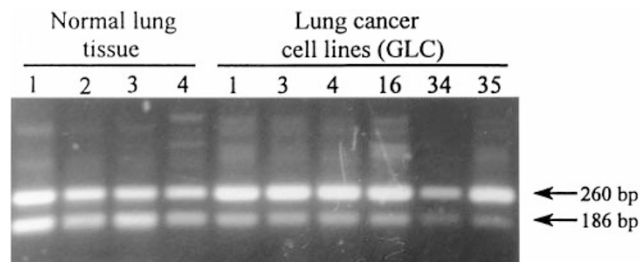


Figure 4 RT-PCR products from gene *RBM6* obtained by amplification of cDNA isolated from normal lung tissue samples and SCLC cell lines. The primer set consisted of one primer in exon 4 and one in exon 7. The longer PCR product includes exon 5, the shorter misses it. Note the differences in the ratio of fluorescence intensities of the two products, between normal lung tissue samples and lung cancer cell lines.

5' end of this gene (exons 1–5) crosses the telomeric breakpoint of the deletion. Exons 6–20 are present on H3938P1.contig13 (Washington University Genome Sequencing Center). Although this contig contains 34 kb upstream to exon 6, it does not contain exons 1–5. This means that an intron of at least 34 kb is located between exons 5 and 6. We were able to determine, in agreement with results published by Wei *et al.*,¹⁶ that *RBM6* also crosses the telomeric breakpoint of the homozygous 3p deletion of the NCI-H740 SCLC cell line. That breakpoint appeared to be located between exons 6 and 7. Thus, in two different SCLC cell lines, *RBM6* is interrupted by homozygous deletions. Still, as in the case of *RBM5*, neither northern, nor Southern, nor SSCP analysis revealed any aberrant pattern or mutation in 15 other lung cancer cell lines. The 4.7 kb band detected on northern blots in addition to the 3.9 kb band may well represent a premessenger RNA. Since also a 2.1 kb is detected on northern blots, *RBM6* may also show splice variants. By RT-PCR, we could demonstrate another alternative splicing, namely of exon 5. One variant includes exon 5, the other variant misses the 74 bp that represent exon 5. This deletion of 74 bp will result in a truncated protein of 520 amino acids instead of 1123 amino acids. Because of the frameshift caused by the 74 bp deletion, the last 26 amino acids at the C-terminus differ from those at the same position in the non-deleted form (Figure 6). Both variants occur in normal lung tissue as well as in SCLC

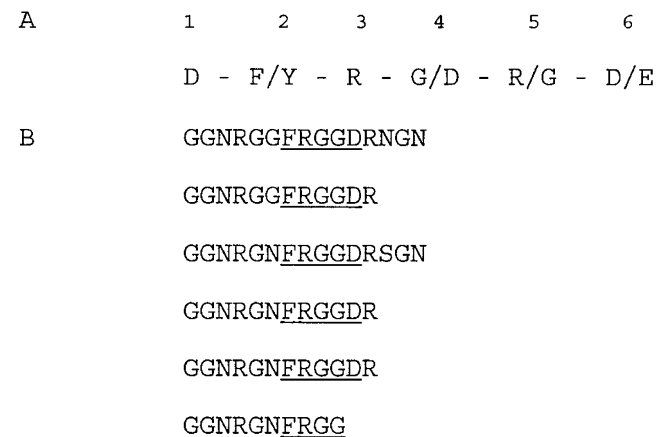


Figure 5 Comparison of repeats in the N-terminal part of the derived protein product of *RBM6* and in the C-terminal part of the RNA binding tyrosine-phosphorylated protein of *Trypanosoma brucei*. A: Consensus sequence of the 20 repeats in *RBM6*; B: C-terminal part of the RNA binding tyrosine-phosphorylated protein of *Trypanosoma brucei* printed in six consecutive sequences to emphasise the repetitive character. The underlined amino acids correspond with the consensus sequence of the *RBM6* repeat.

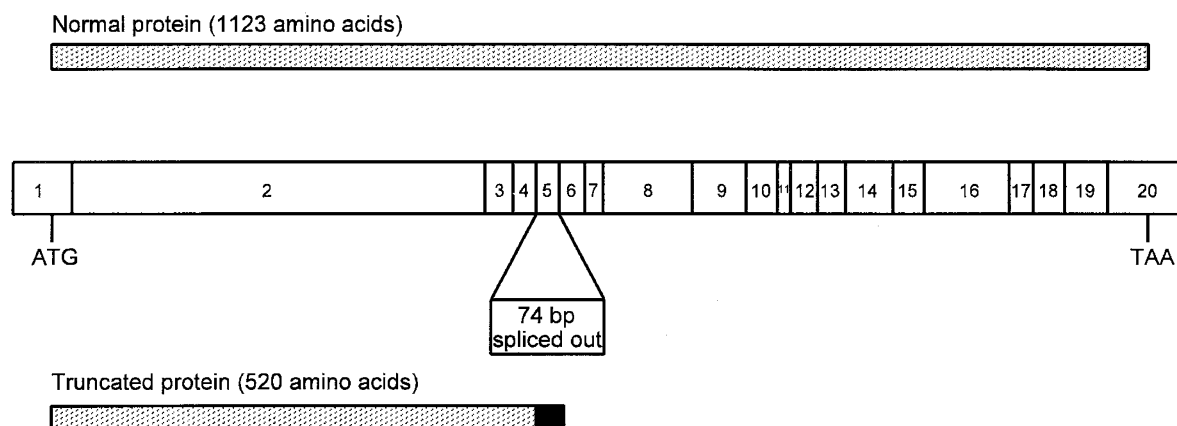


Figure 6 Representation of gene *RBM6* and the protein products derived from different splice variants. In the truncated protein which will be formed when exon 5 is spliced out, the amino acid residues represented by the dark part differ from those in the normal protein product, due to the frameshift caused by this splicing.

cell lines. In normal lung tissue, the amount of the shorter product is in general about half the amount of the longer product (Figure 4). In the lung cancer cell lines, however, the amount of the shorter product is in many cases lower than 5% of the amount of the longer product. On average, the relative amount of the shorter product in normal lung tissue was 4.5 times that in the lung cancer cell lines. Thus, in a relative sense, normal lung tissue contains much more truncated protein than lung tumour cell lines. If *RBM6* were to represent a tumour suppressor gene, one would expect the opposite result, unless some tumour suppressor function may be attributed to the shorter protein. The question remains as to which regulatory mechanism is involved in reversing the proportions of the shorter and longer variants when cells become malignant. Alternative splicing is often regulated in a tissue, sex, and/or developmental stage-specific fashion^{24,25} and regulated by factors that are expressed in a tissue-specific manner and are necessary for the splicing events to occur.²⁶ It may thus well be possible that proliferating cells have different regulation of splicing. This might explain why cultured fibroblasts show a transcript pattern similar to that shown by the lung cancer cell lines.

Acknowledgements

We thank Professor Dr W Timens, Department of Pathology, for making available normal lung tissue specimens, and E Verlind from our Department for sequencing part of the cDNA clones. This work was supported by grants RUG 94-831 from the Dutch Cancer Society and BMH1-CT-94-6269 from the European Commission. Database accession number *RBM5*: AF091263; Database accession number *RBM6*: AF091264. Parts of this study were presented at the

29th annual meeting of the European Society of Human Genetics, Genoa, 17–20 May 1997.²⁷

References

- 1 Kok K, Naylor SL, Buys CHCM: Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. *Adv Cancer Res* 1997; **71**: 27–92.
- 2 Kok K, Van den Berg A, Veldhuis PMJF *et al*: A homozygous deletion in a small cell lung cancer cell line involving a 3p21 region with a marked instability in yeast artificial chromosomes. *Cancer Res* 1994; **54**: 4183–4187.
- 3 Daly MC, Xiang RH, Buchhagen D *et al*: A homozygous deletion on chromosome 3 in a small cell lung cancer cell line correlates with a region of tumour suppressor activity. *Oncogene* 1993; **8**: 1721–1729.
- 4 Roche J, Boldog F, Robinson M *et al*: Distinct 3p21.3 deletions in lung cancer and identification of a new human semaphorin. *Oncogene* 1996; **12**: 1289–1297.
- 5 Killary AM, Wolf ME, Giambrenardi TA, Naylor SL: Definition of a tumour suppressor locus within human chromosome 3p21–p22. *Proc Natl Acad Sci USA* 1992; **89**: 10877–10881.
- 6 Kok K, Van den berg A, Buchhagen DL, Carritt B, Buys CHCM: A PCR-aided transcript titration assay revealing very low expression of a gene at band 3p21 in 33 cell lines derived from all types of lung cancer. *Eur J Hum Genet* 1993; **1**: 156–163.
- 7 Green ED, Olson MV: Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc Natl Acad Sci USA* 1990; **87**: 1213–1217.
- 8 Schwartz DC, Cantor CR: Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984; **37**: 67–75.
- 9 Van der Steege G, Draaijers TG, Grootsholten PM *et al*: A provisional transcript map of the spinal muscular atrophy (SMA) critical region. *Eur J Hum Genet* 1995; **3**: 87–95.

- 10 Tabor S, Richardson CC: A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc Natl Acad Sci USA* 1995; **92**: 6339–6343.
- 11 Murray V: Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res* 1989; **17**: 8889.
- 12 Zimmermann J, Voss V, Schwager C *et al*: A simplified protocol for fast plasmid DNA sequencing. *Nucleic Acids Res* 1990; **18**: 1067.
- 13 Riley J, Butler R, Ogilvie D *et al*: A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* 1990; **18**: 2887–2890.
- 14 Carritt B, Kok K, Van den Berg A *et al*: A gene from human chromosome region 3p21 with reduced expression in small cell lung cancer. *Cancer Res* 1992; **52**: 1536–1541.
- 15 Staden R: Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* 1984; **12**: 505–519.
- 16 Wei MH, Latif F, Bader S *et al*: Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: Progress toward the isolation of a lung cancer TSG. *Cancer Res* 1996; **56**: 1487–1492.
- 17 Coleman MP, Ambrose HJ, Carrel L, Németh AH, Willard HF, Davies KE: A novel gene, *DXS8237E*, lies within 20 kb upstream of *UBE1* in Xp11.23 and has a different X inactivation status. *Genomics* 1996; **31**: 135–138.
- 18 Nagase T, Seki N, Ishikawa K *et al*: Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201–KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res* 1996; **3**: 321–329.
- 19 Inoue A, Takahashi KP, Kimura M, Watanabe T, Morisawa S: Molecular cloning of a RNA binding protein, S1-1. *Nucleic Acids Res* 1996; **24**: 2990–2997.
- 20 Güre AO, Altorki NK, Stockert E, Scanlan MJ, Old LJ, Chen YT: Human lung cancer antigens recognised by autologous antibodies: definition of a novel cDNA derived from the tumour suppressor gene locus on chromosome 3p21.3. *Cancer Res* 1998; **58**: 1034–1041.
- 21 Nagai K, Oubridge C, Ito N, Avis J, Evans P: The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem Sci* 1995; **20**: 235–240.
- 22 Robbins J, Dilworth SM, Laskey RA, Dingwall C: Two interdependent basic domains in nucleoplasmic nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* 1991; **64**: 615–623.
- 23 Das A, Peterson GC, Kanner SB, Frevert U, Parsons M: A major tyrosine-phosphorylated protein of *Trypanosoma brucei* is a nucleolar RNA-binding protein. *J Biol Chem* 1996; **271**: 15675–15681.
- 24 Smith CW, Patton JG, Nadal Ginard B: Alternative splicing in the control of gene expression. *Annu Rev Genet* 1989; **23**: 527–577.
- 25 Hodges D, Bernstein SI: Genetic and biochemical analysis of alternative RNA splicing. *Adv Genet* 1994; **31**: 207–281.
- 26 Latchman DS: Cell-type-specific splicing factors and the regulation of alternative RNA splicing. *New Biol* 1990; **2**: 297–303.
- 27 Timmer T, Van den Berg A, Terpstra P *et al*: Homology in the human genome between a 3p21.3 region and a region of Xp11.23 [Abstract]. *Med Genet* 1997; **9**: 94–95.
- 28 Horton P, Nakai K: A probabilistic classification system for predicting the cellular localisation sites of proteins. *ISMB* 1996; **4**: 109–115.
- 29 Bairoch A, Bucher P: PROSITE: recent developments. *Nucleic Acids Res* 1994; **22**: 3583–3589.