European Journal of Human Genetics (2002) 10, 52–61

www.nature.com/ejhg

ARTICLE

The transcriptional map of the common eliminated region 1 (C3CER1) in 3p21.3

Hajnalka Kiss*[,1], Ying Yang[1], Csaba Kiss[1], Kenth Andersson[1], George Klein[1], Stephan Imreh[1] and Jan P Dumanski[2,3]

[1]*Microbiology and Tumor Biology Center (MTC), Karolinska Institutet, S-17177, Stockholm, Sweden;* [2]*Department of Molecular Medicine, CMM building L8:00, Karolinska Hospital, S-17176, Stockholm, Sweden;* [3]*Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University Hospital, S-75185, Uppsala, Sweden*

Occurrence of chromosome 3p deletions in a large number of human tumours suggests the existence of uncharted tumour suppressor gene(s). We previously applied a functional assay, named the Elimination test (Et), for the identification of regions containing tumour growth antagonising genes. This resulted in the definition of chromosome 3 common eliminated region 1 (C3CER1) on 3p21.3, which is regularly eliminated from SCID-derived tumours. Systematic genomic sequencing of 11 PAC clones, combined with comparisons of genomic sequence against EST databases and PCR-based cloning of cDNA sequences allowed us to assemble a comprehensive transcriptional map of 1.4 Mb that includes 19 active genes and three processed pseudogenes. We report four novel genes: FYVE and coiled-coil domain containing 1 (*FYCO1*), transmembrane protein 7 (*TMEM7*), leucine-rich repeat-containing 2 (*LRRC2*) and leucine zipper protein 3 (*LUZP3*). A striking feature of C3CER1 is a presence of a cluster of eight chemokine receptor genes. Based on a new analysis of the microcell hybrid-derived panel of SCID tumours we also redefined the centromeric border of the C3CER1. It is now located within *LRRC2* gene, which is a relative of *RSP-1* (Ras Suppressor Protein 1). The detailed knowledge of gene content in C3CER1 is a prerequisite for functional analysis of these genes and understanding of their possible role in tumorigenesis.
European Journal of Human Genetics (2002) **10,** 52–61. DOI: 10/1038/sj/ejhg/5200758

Keywords: tumour suppressor genes; elimination test; chromosome 3 deletions; FYVE domain; leucine rich repeat domain (LRR); chemokine receptor cluster

Introduction

Hemizygous deletions on 3p are very frequent and have been described in virtually all types of solid tumours. Interstitial deletions including 3p21 have been described in at least 21 different tumours.[1] Loss of heterozygozity (LOH) and comparative genomic hybridisation (CGH) studies are concordant in delineating several regions of regional losses over 3p: (a) 3p25–p26 around the VHL gene; (b) 3p21.3–p22; (c) 3p 21.2; and (d) 3p13–p14 around the FHIT gene.[2,3] Five homozygous deletion (HD) regions have been described on 3p: on p25, p22–p21.3, p21.3–p21.2, p14.2 and p13–p12, respectively.[4–10] The length of these HDs varies between few hundred bp and 6–8 Mb.[8] Best studied are the two HDs involving p21 where the shortest overlapping HD segment was 630 kb and 120 kb, respectively.[11,12]

We have developed an assay, named the Elimination test (Et), for identification and fine mapping of chromosomal regions containing tumour growth antagonising genes.[13] The putative tumour suppressor region (named common eliminated region 1, C3CER1) was initially restricted to ~7 cM[14] and later to ~1.6 cM[15] in 3p21.3 region. C3CER1 was covered by a PAC contig and its physical size was determined as approximately 1 Mb[16]. Construction of a detailed transcription map of C3CER1 is a requirement for the identification of a gene(s) with tumour inhibitory

*Correspondence: H Kiss, Microbiology and Tumor Biology Center (MTC), Karolinska Institutet, Nobels väg 16, S-171 77, Stockholm, Sweden. Tel: +46-8728-6771; Fax: +46-8330-498;
E-mail: Hajnalka.Kiss@mtc.ki.se
Received 14 August 2001; revised 13 November 2001; accepted 19 November 2001

properties. The C3CER1 lies between two previously characterised homozygously deleted regions: The 630 kb lung cancer homozygous deletion region on 3p21.3[11] and the 685 kb homozygously deleted region in a lung carcinoma cell line on 3p22–p21.3.[9,17] In order to systematically identify genes within C3CER1, we initiated a large-scale sequencing of PACs. We previously reported a novel human LIM domain containing gene 1 (*LIMD1*) and its mouse ortholog (*Limd1*)[18] and a partial physical and transcriptional map of 250 kb, which included eg the Leucine Zipper Transcription Factor-Like 1 (*LZTFL1*).[19] As the continuation of this project, we have sequenced and analysed for active gene content of eight additional PAC clones from C3CER1. During this work and as a result of the Human Genome Project, the sequences of additional fully or partially sequenced genomic clones were released in the public databases. We now assembled all the sequencing data for C3CER1 and report a physical map of 21 BAC/PAC clones as well as a comprehensive transcriptional map of 1.4 Mb. We also identified and characterised four novel genes and three pseudogenes within C3CER1.

Materials and methods

Genomic and cDNA sequencing was performed as described previously.[20,21] Repetitive sequences were filtered out from genomic sequence using the local RepeatMasker (repeatmasker.genome.washington.edu). The blast family of programs was used for database searches on the NCBI/NIH server (www.ncbi.nlm.nih.gov/BLAST). Trace files for the ESTs were imported via ftp from genome.wustl.edu and assembled using the Staden program package.[22] PSORT II program was used to find protein sorting signals and localisation sites in the studied proteins (psort.nibb.ac.jp/form2.html).[23,24] SMART (Simple Modular Architecture Research Tool, smart.embl-heidelberg.de)[25,26] and the ISREC-server (http://www.isrec.isb-sib.ch/software/PFSCAN_form.html) were used for the study of protein domains structure, to identify Prosite profiles and Pfam domains in the predicted proteins. We used the GENSCAN program for the prediction of gene coding sequences (bioweb.pasteur.fr/seqanal/interfaces/genscan.html).

The *FYCO1* cDNA was covered with primers (primer 3–17, Table 1), which were used in PCR amplifications from Marathon-ready thymus and skeletal muscle cDNA libraries (Clontech, nos.7415-1, 7413-1). Marathon RACE was performed to obtain the 5′ end of the FYCO1 gene from Marathon-ready skeletal muscle cDNA library. Primer 1 and AP1 primer were used in the primary reaction and primer 2 and AP2 primer were used in the nested reaction. During characterisation of *TMEM7* gene, Marathon RACE was performed from Marathon-ready liver cDNA library (Clontech, no.7407-1) with primers 18 or 20 together with AP1 primer and primers 19 or 21 together with AP2 primer. Primer 24 was used for the sequencing of the obtained Marathon *TMEM7* fragment. Primers 30–35 were used for the amplification of different parts of the *LRRC2* cDNA. Marathon RACE was carried out from Marathon-ready skeletal muscle cDNA library (Clontech, no.7413-1) to isolate the *LRRC2* gene. Primers 25 or 27 were used together with AP1 primer and primer 26 or 28 or 29 were used together with AP2 primer. For the cloning of the *LUZP3* gene, we performed Marathon RACE from Marathon-ready skeletal muscle cDNA library (Clontech, no.7413-1) using primer 36 and primer AP1 in the primary reaction and primer 37 and primer AP2 in the nested PCR reaction. Primers 38–40 were used for sequencing of the obtained Marathon *LUZP3* fragment. The conditions of Marathon RACE PCR were according to the recommendations of the supplier. PCR amplified cDNA fragments were isolated in low melting point agarose and sequenced as described previously.[27] Human 12-Lane Northern Blot (Clontech, no.7780-1) was hybridised with a human *FYCO1* (primer 12-13, 229 bp), *TMEM7* (primer 22-23, 369 bp), *LRRC2* (primer 25-29, 311 bp) and *LUZP3* (primer 41-42, 261 bp) cDNA probes, in separate experiments. Probe labelling, hybridisation and washing were performed according to standard protocols.[28,29]

PolyA mRNA was extracted from 10^7 IB-4 cells using Dynabeads (Dynal) according to the manufacturer's protocol. Oligo dT-primed, first-strand cDNA was synthesised from 0.1 μg polyA mRNA from IB-4 cell line and 5 μg total RNA from brain, heart, kidney, liver lung and trachea (Clontech, no.K4000-1) in a 20 μl volume using Superscript II (Life Technologies, Inc., Grand Island, NY, USA) according to instructions from the supplier. One μl of each synthesised cDNA and Marathon-ready thymus, skeletal muscle and liver cDNA (Clontech, nos 7415-1, 7413-1, 7413-1) were subjected to PCR in a volume of 20 μl using primers 41 and 42 and primer 43 and 44 as a control. The cell lines MCH 906.8 and their SCID derived tumours were described previously.[16] DNA was prepared by proteinase K digestion and phenol/chloroform extraction according to standard protocol (Sambrook *et al*., 1989). The PCR-markers used for the genomic analysis were primers 22-23, 30-31, 32-33, 34-35, 41-42 and primer 49-50. High-density filters with human PAC libraries were constructed at the Roswell Park Cancer Institute, Buffalo, USA.[30] We screened the RPCI-5 and RPCI-6 libraries as described previously.[21] Probe labelling, hybridisation and washing of the colony hybridisation filters were performed according to standard methods.[28,29]

Results

Sequencing of PAC clones within C3CER1

We have previously established a PAC contig that was assumed to fully cover C3CER1.[16] However, in the course of the RP4-787C23 sequencing project, we realised that this clone was not overlapping with the sequence of RP6-123I13. This false overlap was initially based on the PCR analysis of a single STS marker (123i13-S.endB). We therefore rescreened the PAC libraries using two probes to bridge the gap between

Table 1 Primers used for the identification and characterisation of the human *FYCO1, TMEM7, LRRC2* and *LUZP3* genes and for the generation of PAC library screening probes

No.	*Name*	*Primer sequence (5′→3′)*	*Ann. temp °C*
1	FYCO1.GSP1.R	TGGCCAGGCAGGCACAGAAGTAATC	68
2	FYCO1.GSP2.R	GGGGGTTTCCACAGGTAAGCAGAAGA	68
3	FYCO1.F2A	TCCAACTTTGTCCTGAACAGC	
4	FYCO1.F2	AGTTGGAGGTGCGGGAGA	
5	FYCO1.R4A	TTGGTATCGTCTAACTGTGCCT	55
6	FYCO1.F4	AGGCACAGTTAGACGATACCAA	
7	FYCO1.R5A	CTCCCTCTGCTGGGCCTCAT	55
8	FYCO1.F6	GATGAGGCCCAGCAGAGG	
9	FYCO1.F6	CTGGTAGTCCATGATTTCAGCA	57
10	FYCO1.F7	AGAGGCTGCACACCAGGA	
11	FYCO1.R7	TTGGTTCTGTTGAGGTCATCC	57
12	FYCO1.F9	CTTTGCCTGAAACACCCACT	
13	FYCO1.R9	GTGCTGGACCTCACAAACAG	57
14	FYCO1.F9A	CACTTCATCCACAGGAAACA	
15	FYCO1.R9A	GGGCTTTATTTGAGGACAGA	57
16	FYCO1.F10	TCCCTGTTCTCTGTGGTCCT	
17	FYCO1.R10	GAGTCCCTGACCCATTCTCA	60
18	TMEM7.GSP1.R	TGAGAATGACACAACAGCAGAAAATGCTT	68
19	TMEM7.GSP2.R	CCAGGGCTTAACTACCTCCTCCACCTTG	68
20	TMEM7.GSP3.F	TCAGGAGTTAATGCGGGAGGTGAAGC	68
21	TMEM7.GSP4.F	GCATTTTCTGCTGTTGTGTCATTCTCATTG	68
22	TMEM7.F	GATCCTGAAAAACCTGGTGT	
23	TMEM7.R	TCCAAGGCTCATATAGCAGT	58
24	TMEM7.R2	CTTGTCTGGTCTCAGGGTC	55
25	LRRC2.GSP1.F	CACAGGCACGGGAGGCTTCTTCC	68
26	LRRC2.GSP2.F	GCTTGGCAGAAGAAGGAGGTGGAAA	68
27	LRRC2.GSP3.R	CAAACCAATTTTTCAATTCTCCAGTCCA	68
28	LRRC2.GSP4.R	GGCCAAACCAGCCTCGAACTCC	68
29	LRRC2.GSP5.R	CGGCCACAAAGTTCCACTCCTCCT	68
30	LRRC2.F1	AGGAACTCCTTATCAGTGTGGA	
31	LRRC2.R1	GTTAAGTGCATTGTTTTCAGG	55
32	LRRC2.F2	TGAACATTCCTAGGCACTGA	
33	LRRC2.R2	TGGGTTTGACCTGATGGAAA	55
34	LRRC2.F3	ATTCTCCTTTGGAATGTTTCTG	
35	LRRC2.R3	GGTGTAAAACAGGCAGCGCATT	55
36	LUZP3.GSP1.F	GTTTCTTCACCACTGGCTGTGGACTTTG	68
37	LUZP3.GSP2.F	CACCCGTCACCCTTTACCCTCATCC	68
38	LUZP3.R1	ACCCTAACCAGCACTCCAAA	55
39	LUZP3.R2	CGCCCTGCTATGCAATAGA	55
40	LUZP3.R3	GACTTCCATCACATGCCCAT	55
41	LUZP3.F4	CATGGGCATGTGATGGAAGT	
42	LUZP3.R4	CTTTGCGACTTTGCATCCTT	56
43	GAPDH.F	TCCACCACCCTGTTGCTGTA	
44	GAPDH.R	ACCACAGTCCATGCCATCAC	55
45	165I16end.F	TCTAAGGTTCAGACACATGCC	
46	165I16end.R	AGACAAACCAGCCCTTGCT	55
47	188G11end.F	TCAGGTTTCTAGGGCTTGGTT	
48	188G11end.R	TTGAAATGCTAGTCTGACCCC	55
49	TDGF1.F	ACAGAACCTGCTGCCTGAAT	
50	TDGF1.R	AGGCAGATGCCAACTAGCAT	55

RP11-165I16 and RP6-188G11. We used PCR fragment from the centromeric end of RP11-165I16 (primer 45-46) and from the telomeric end of RP6-188G11 (primer 47-48). Primer 45-46 produced the following positive clones: RP6-21A17, RP6-58E17, RP6-153F14, RP6-159P24 and RP5-844B4. Screening with primer 47-48 resulted in clones RP6-146E1, RP6-188G11, RP5-897E4, RP5-1025F12 and RP5-1080O15. Clones RP6-153F14 and RP6-146E1 were selected for sequencing (Figure 1). Therefore the last gap in the C3CER1 PAC contig was closed. Substantial amount of new sequence data was recently deposited in the databases as a result of the Human Genome Project. Three fully sequenced clones (RP5-1053D16, Acc. AC006515; RP11-165I16, Acc. AC005669; and BAC110P12, Acc. U95626) and seven partially sequenced clones (RP11-111P21, Acc. AC063916; RP11-107O13, Acc. AC022951; RP11-697K23, Acc. AC062007; RP11-852E15, Acc. AC024150; RP11-91E8, Acc. AC026349; RP11-793E15, Acc. AC024739; and RP11-509I21, Acc. 068720) could be

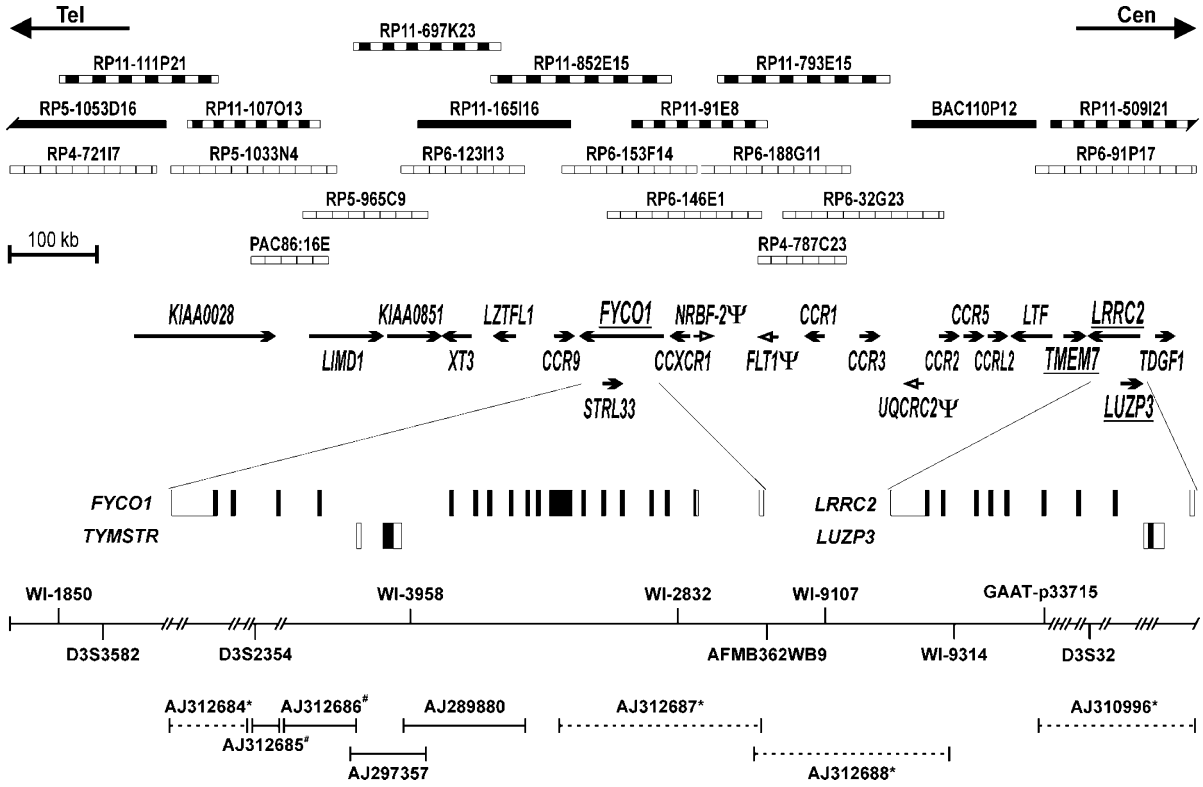

Figure 1 Transcriptional map of 1.4 Mb encompassing the C3CER1. On the top, the BAC and PAC clones are represented as boxes. Black boxes stand for clones that were fully sequenced: RP5-1053D16 (Acc. AC006515) by the Baylor College of Medicine (Baylor), RP11-165I16 (Acc. AC005669) by the Whitehead Institute for Biomedical Research (Whitehead) and BAC110P12 (Acc. U95626) by the Cold Spring Harbor Laboratory (CSHL). Boxes with zebra-pattern show clones that were partially sequenced: RP11-111P21 (Acc. AC063916, 11 contigs), RP11-852E15 (Acc. AC024150, 20 contigs) and RP11-91E8 (Acc. AC026349, 22 contigs) by Baylor, RP11-107O13 (Acc. AC022951, 29 contigs) and RP11-697K23 (Acc. AC062007, 26 contigs) by Whitehead and RP11-793E15 (Acc. AC024739, 22 contigs) and RP11-509I21 (Acc. AC068720, 20 contigs) by the Washington University Genome Sequencing Center. The boxes with thin vertical lines denote PAC clones that were sequenced in the course of our project. Filled and empty arrows represent the genes and pseudogenes, respectively. The genes, which names are underlined (*FYCO1*, *TMEM7*, *LRRC2* and *LUZP3*), are reported in this paper. Arrows, which represent genes that are smaller than 10 kb, are not drawn to scale. The gene abbreviations are as follows: *KIAA0028* (Acc. XM003255), mitochondrial leucyl-tRNA synthetase; *LIMD1* (Acc. AJ132408), LIM Domain containing 1; *KIAA0851* (also named *SAC1*) (Acc. NM014016), suppressor of actin 1; *XT3* (Acc. AJ276207), orphan transporter; *LZTFL-1* (Acc. AJ297351), Leucine zipper transcription factor like 1; *CCR9* (Acc. NM031200), CC chemokine receptor 9; *FYCO1* (Acc. AJ292348), FYVE and coiled-coil domain containing 1; *STRL33* (also named *TYMSTR* or *Bonzo*) (Acc. NM_006564), G protein-coupled receptor; *CCXCR1* (also named *GPR5*) (Acc. XM003249), chemokine (C motif) XC receptor 1; *CCR1* (Acc. XM003248), CC chemokine receptor 1; *CCR3* (Acc. XM003247), CC chemokine receptor 3; *CCR2* (Acc. XM002924), CC chemokine receptor 2; *CCR5* (Acc. NM000579), CC chemokine receptor 5; *CCRL2* (also named *CRAM-B*) (Acc. NM003965), CC chemokine receptor-like 2; *LTF* (Acc. NM002343), lactotransferrin; *TMEM7* (Acc. AJ312776), transmembrane protein 7; *LRRC2* (Acc. AJ308569), leucine-rich repeat-containing 2; *LUZP3* (Acc. AJ312775), leucine zipper protein 3; *TDGF1* (Acc. NM003212), teratocarcinoma-derived growth factor 1. Two genes (*FYCO1* and *LRRC2*) embrace fully two other genes (*STRL33* and *LUZP3*), which have an anti-parallel transcriptional orientation. The exon–intron organization of these two pair of genes is shown below. The protein coding regions are represented by black boxes and the non-coding regions by white boxes. The three pseudogenes are: nuclear receptor binding factor-2 pseudogene (*NRBF-2*Ψ), fms-related tyrosine kinase 1 pseudogene (*FLT1*Ψ) and ubiquinol-cytochrome *c* reductase core protein II pseudogene (*UQCRC2*Ψ). The combined sequence of the 1.4 Mb contains 11 gaps, represented by broken lines. The position of 10 selected chromosomal markers is displayed. In the bottom of the figure, the names and positions of accession files deposited by us in the EMBL/GenBank database are shown. Dotted lines represent not contiguous submissions and asterisks (*) indicate submissions made for the present paper. Two additional submissions (labelled by # sign) were recently updated.

recognised within C3CER1 (Figure 1). We have joined all these data together with our sequences, which are derived from 11 PAC clones. This resulted in 12 sequencing contigs that span at least 1380 kb.

Identification and characterisation of the human *FYCO1* gene

After masking repetitive elements in the sequences of RP6-153F14 and RP6-146E1, blastn search against dbest database

Table 2 Genomic organisation of the *FYCO1* (Acc. AJ292348), *TMEM7* (Acc. AJ312776) and *LRRC2* (Acc. AJ308569) genes. The intronic and exonic sequences are shown in lower-case and upper-case characters, respectively. The first two and the last two bases of introns (**gt** for donor and **ag** for acceptor-splice sites) are shown in bold. Sequences for putative polyadenylation signals are also indicated in bold characters. The exact size of the intron 7 of *LRRC2* is not known because of the gap in the genomic sequence

	EXON		INTRON	
Intron	*number*	*size (bp)*		*size (bp)*
FYCO1				
	1	87	ACGAGCGCGTAA**gt**gccgcggaag	10698
tcctctgtgc**ag**GTGCCACCATGG	2	167	GAGACTTGCAAG**gt**atggtaggtt	3179
atctcctgca**ag**TAAGTGATGCCT	3	107	TATCTCCTGCAA**gt**aagtgatgcc	1738
ttccctaccc**ag**TTTGATCAGAAA	4	126	TCTATCTCAGAG**gt**aagctgagca	4316
tgctttttcc**ag**CTCCGAACATCC	5	107	CAAAGTGACCAG**gt**aaatacagga	2009
gtttttctgc**ag**TGACTGGTACTA	6	144	AACATTTGCCAG**gt**actttgaagg	3322
gtggccccac**ag**GAGGACGCTGAC	7	91	AGCTACCTGCAG**gt**aaggccagga	971
tttccctccc**ag**ACTCAGAGATG	8	2427	AGCAGACTTTAAG**gt**aatcctagct	1151
tgtcctctct**ag**AATGCTGGTGAA	9	93	GAGAAGCTGAAG**gt**aggccctggt	578
gtcctgtgcc**ag**GCCACCCAAGCA	10	119	AGACCTAGAAAG**gt**ttagtcacagt	1943
ctcctcccccc**ag**GACCCAGAAGGA	11	168	AGAGCTGCTCAG**gt**gagtgctgtg	2682
gttttgtccc**ag**GGACAAGGATGC	12	150	GCACCACTGCAG**gt**cagcagggag	773
ttttgttccc**ag**GATATGTGGCCG	13	212	CAGGAGGCCAAG**gt**cagtctcctg	3006
tctgctgccc**ag**GAGCAAATACAG	14	145	TGCGGCTGAACA**gt**gagtagaggt	18705
gtgacatttc**ag**GGATACTACATC	15	96	TGGAGAACTGAT**gt**aagtggcaac	5166
tttttatttt**ag**GATCAAAGTACC	16	211	GATCAGTGTAAG**gt**aaggctgggc	7303
tttccctccc**ag**GTCCTCATTCC	17	110	TACCTTCTCAAG**gt**agggccacct	1813
ctgtttcctc**ag**GTTTGTCTCTAA	18	3940	TTTA**AATAAA**GTAAACACTATGACATTT+polyA	
TMEM7				
	1	223	GACCTTCGCCAG**gt**gaggggaat	2140
ctgtctccac**ag**GTTCCAGTGCTC	2	594	AGAAAAAAAATCA**GATACA**AAAAGTC+polyA	
LRRC2				
	1	148	CCCTCGGGGCAG**gt**aagcgctggg	14566
ttttttaaat**ag**GTAATTTCATG	2	144	CGCCTTGGAGAA**gt**acgtgttttc	6213
tgcctgttttt**ag**GATAAAGGAGGA	3	208	GAGCACTGGACG**gt**gagtcgaggc	5879
tttctatttt**ag**GAGCTCCCAGAT	4	157	CAGCAGAAATCG**gt**atgtcaactt	6136
ttctgctgac**ag**GTTGTTTGAAGA	5	137	CTGCCCTTTGAA**gt**aagagataaa	2722
attcatttct**ag**TTAAGTAATTTG	6	146	AGATATAGACA**gt**agctacttgc	2322
acctcttctt**ag**GCTAGAGGAGCT	7	156	CACACCTTTAAA**gt**gagtagcccq	>6146
tttctcgaat**ag**ATTTGTAAGCCT	8	137	TTAAAGAAAGAG**gt**gggttactca	2454
tttctgttac**ag**AATCTGTTCCCA	9	3628	CAG**AATACA**TGTATGAATATTAAAGTAA+polyA	

revealed similarities with multiple ESTs. We could also detect a match with mouse partial cDNA sequence, which contains only the 3′ untranslated region of the mouse *Mem2* gene (Acc. X95350). The human ESTs were assembled and compared with the genomic sequence, which partially revealed the genomic structure of this new gene. Since the cDNA sequence of the *FYCO1* gene, which was assembled on the basis of ESTs, contained several gaps, we used the GENSCAN prediction program to detect the missing exons. Several primer pairs were designed to verify the already available cDNA sequence and to fill in the missing parts. The sequence of the 5′ end of the gene was verified by Marathon RACE (primer 1, 2). The resulting cDNA is 8500 bp long; it contains 199 bp of the 5′ untranslated region, with the predicted ATG initiation codon at position 200. The open reading frame is 4434 bp and is capable of encoding a protein of 1478 aa, with the predicted molecular mass of 167 kDa. The TAG stop codon starts at position 4634. The 3′ untranslated region of the *FYCO1* cDNA contains a regular polyadenylation signal (AATAAA) starting at position 8477 bp. Comparison of the *FYCO1* cDNA with the genomic sequences revealed that the gene consists of 18 exons (Figure 1, Table 2). The genomic size of *FYCO1* is 77.8 kb and its structure is typical of other human genes with a large first intron (10.7 kb). Northern hybridisation with the human cDNA probe (Figure 2A) revealed an 8.5 kb transcript, which is expressed mainly in heart and skeletal muscle. Strong overexposure of the X-ray film detected 8.5 kb transcript bands also in brain, kidney, liver, small intestine, placenta and lung (not shown).

The SMART and the ISREC ProfileScan servers have predicted a number of protein domains, which might shed light on the normal function of the FYCO1 protein. The FYVE zink finger domain was predicted between aa 1165 – 1232 by both servers. Also, the RUN domain (between aa 104 – 167) was detected, that is involved in Ras-like GTPase signalling. The ISREC ProfileScan server has further predicted a glutamine-rich region (496 – 897 aa), a spectrin repeat (468 – 550 aa), an ERM domain (377 – 680 aa) and a Granin domain (359 – 946 aa). Multiple FYCO1 protein regions were also predicted as containing coiled-coil structure. Using the

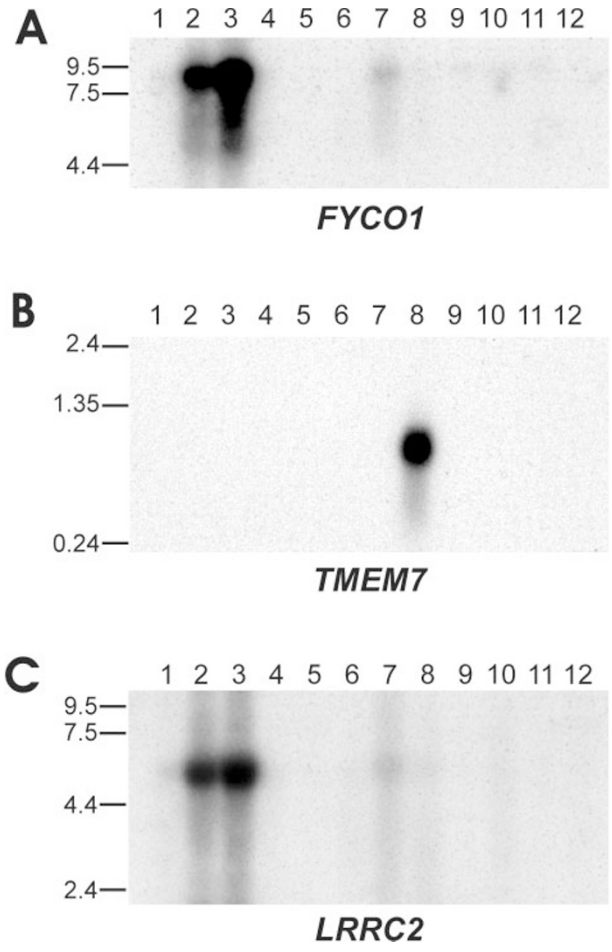

Figure 2 Northern blot analysis of the human *FYCO1* (**A**), *TMEM7* (**B**) and *LRRC2* (**C**) genes. The same human 12-Lane Multiple Tissue Northern Blot (Clontech no. 7780-1) was hybridised with cDNA probes. Numbers 1–12 correspond to brain, heart, skeletal muscle, colon, tymus, spleen, kidney, liver, small intestine, placenta, lung and peripheral blood lymphocytes, respectively. Molecular size markers are indicated on the left of the autoradiograms. The same blot was also tested with β-actin probe (Clontech), which established that approximately equal amount of mRNA has been loaded in each lane of the Northern blots (not shown).

coils program we could detect coiled-coil domain between positions 4–31, 224–281, 394–453, 464–558, 595–669, 675–754, 760–912, 914–1065, 1069–1113 and 1117–1151 (with score of at least 0.9, using 28-analysis window, MTIDK matrix and 'no weights' option). The multicoil program produced a score over 0.9 between positions 484–524, 616–632, 688–715, 924–975 and 1018–1060, using a dimer probe. Low complexity sequence regions were detected in 11 short instances: 196–207 aa, 427–439 aa, 484–507 aa, 639–651 aa, 686–700 aa, 763–779 aa, 854–877 aa, 885–899 aa, 953–996 aa, 1231–1247 aa and 1249–1262 aa. The blastp analysis, using the FYVE domain sequence of the FYCO1 protein as query, recognised a strong match with human EIP1 (Acc. AF361055) and EEA1 (endosome-assotiated protein) (47% identity, 57% similarity and 45% identity, 53% similarity, respectively). Interestingly, the similarity between FYCO1 and EIP1 is not restricted to the FYVE domain regions of both proteins. Thus, the proposed name of this gene, *FY*VE and *CO*iled-coil domain containing 1 (*FYCO1*), reflects the presence of a FYVE domain and coiled-coil regions within the predicted FYCO1 protein.

The *STRL33* (also named *TYMSTR* or *Bonzo*) consist of two exons, localised in the 14th intron of the *FYCO1* gene (Figure 1). The *STRL33* is a chemokine receptor for CXCL16 chemokine[31] and it is also a coreceptor for HIV/SIV.[32]

Characterisation of the human *TMEM7*, *LRRC2* and *LUZP3* genes

We identified several EST clusters by using the repeat masked sequence of RP6-91P17 as query, against the dbest database. We designed PCR primer pairs for several assembled EST contigs, in order to confirm the expression profile of the predicted genes in a panel of human tissues (heart, brain, liver, kidney, trachea, lung, skeletal muscle and thymus). When positive evidence was obtained, we continued with designing gene specific primers for Marathon RACE system, to uncover the whole cDNA sequence.

In the case of *TMEM7*, we used primer 22-23 to test the expression by RT–PCR, and the transcript was detected exclusively in liver. We designed primary and nested gene specific primers for this EST cluster, both in 5′ and 3′ direction. Two bands were amplified and sequenced, which resulted in an 817 bp cDNA sequence with the predicted initiation codon start at position 69 bp. The ORF is composed of 696 bp and it is capable of encoding a protein of 232 aa, with a predicted molecular mass of 27 kDa. The stop codon starts at position 765 bp. The *TMEM7* gene does not apparently contain any regular polyadenylation signal. The probable polyadenylation signal might be G*ATACA* (Table 2).[33] The *TMEM7* gene consists of two exons, with genomic size of approximately 3 kb. Northern hybridisation with the *TMEM7* cDNA (Figure 2B) revealed a transcript exclusively in liver, among the twelve tested tissues. The SMART programme predicted a single transmembrane domain near the C-terminus (211–228 aa) of the TMEM7 protein. For this reason, this gene was named ''*T*rans*MEM*brane protein *7*'' gene (*TMEM7*). Using the PSORT II program we could also find an endoplasmic reticulum membrane retention signal (VKTA), located in the immediate vicinity of the C-terminal end.

During the cloning of the *LRRC2* gene, we identified 18 ESTs and these were assembled into five contigs. This, together with Marathon–RACE PCR, resulted in a 4860 bp cDNA sequence, and it consists of a 167 bp 5′UTR, 1116 bp ORF and 3577 bp 3′ UTR region. As was the case for *TMEM7*, the *LRRC2* gene does not apparently contain any regular polyadenylation signal, the probable signal being AA*TACA*.[33]

Comparison of the cDNA to the genomic sequence uncovered nine exons (Table 2). The genomic size of *LRRC2* is at least 51 kb and its structure is typical, with a large first intron (14.5 kb). Northern hybridisation with the *LRRC2* cDNA produced a strong signal in heart and skeletal muscle and weak signal in kidney (Figure 2C). The predicted LRRC2 protein consists of 371 amino acids, with molecular mass of 43 kDa. The SMART programme predicted seven Leucine-rich repeats, of which four were typical (143–165 aa, 166–189 aa, 236–258 aa and 259–282 aa) and three were unusual (189–212 aa, 213–235 aa and 282–301 aa). This gene was therefore named as *L*eucine-*R*ich *R*epeat-*C*ontaining *2* (*LRRC2*). The PSORT II program detected two putative nuclear localisation signals; KKHK at position 22 and PKDRGKR at position 91. Using *blastp*, we found that the human RAS suppressor protein (RSP-1) shows the highest similarity to LRRC2 protein. The RSP-1 contains 7 typical leucine-rich repeats and the similarity between the LRRC2 and RSP-1 proteins extends outside of the leucine-rich repeat regions (aa 69–190 of LRRC2 shows 33% identity and 53% similarity between aa 78–203 of the RSP-1 protein).

Cloning of the *LUZP3* was initiated by a match between genomic sequence and a cluster of 6 ESTs, corresponding to a part of the 3′ UTR region of *LUZP3*. The RACE PCR allowed us to obtain 2291 bp cDNA sequence, with 415 bp of 5 UTR. The ORF is 516 bp and it is capable of encoding a protein with 171 aa, with a predicted molecular mass of 18.8 kDa. No regular polyadenylation signal was recognised and the putative signal is TC*TAAA*.[33] The only domain/motif which could be detected in the protein sequence is a leucine zipper pattern (between aa 132–153) and this gene was therefore named as '*Leu*cine *Z*ipper *P*rotein *3*' gene (*LUZP3*). It consists of one exon that is located on the opposite strand and within the first intron of *LRRC2*. Northern hybridisation did not reveal any visible bands (not shown). We tested therefore the expression profile of this gene by RT–PCR, using 10 cDNA samples (primer 41-42, Figure 3) and detected PCR products in IB-4 cell line, kidney, trachea and skeletal muscle.

Three processed pseudogenes are located in C3CER1

We identified three processed pseudogenes (*NRBF-2*Ψ, *UQCRC2*Ψ and *FLT1*Ψ) within 290 kb of the centromeric part of C3CER1. The genomic sequence located 1.3 kb to the centromere from *CCXCR1*, showed a high similarity (89–95% within a stretch of 1.8 kb) to the nuclear receptor binding factor-2 gene (*NRBF-2*, Acc. NM_030759). The *NRBF-2* gene is located on chromosome 10 and contains 4 exons. Using blastn search in the entire human genome with *NRBF-2* cDNA sequence as query, we could identify several other, highly similar (89–96%) regions on chromosomes 1, 3, 8, 15 and 18. In all instances the region of similarity was contiguous and not interrupted by introns, which represent multiple processed pseudogenes of the *NRBF-2* gene (*NRBF-2*Ψ). In C3CER1, the similarity starts at the very beginning of the NRBF-2 cDNA and continues through the entire *NRBF-2* cDNA, with the exception of three short stretches of 9–68 bp. Interestingly, we noticed a LINE/L1 repeat immediately ahead and after the *NRBF-2*Ψ, which might be a trace of the distant retrotransposition event. The ORF of *NRBF-2*Ψ is interrupted by four stop codons (details not shown). No ESTs or cDNA sequences perfectly matching the *NRBF-2*Ψ could be recognised.

Approximately 1.5 kb around the telomeric end of BAC110P12 shows 82–85% similarity to the human ubiquinol-cytochrome C reductase core protein II (*UQCRC2*, Acc. BC000484). The *UQCRC2* is located on chromosome 16 and contains 14 exons. The similarity between *UQCRC2* cDNA and C3CER1 starts at 125 bp (start of the second fully translated exon) of *UQCRC2* cDNA and continues, without intronic interruptions, through the entire *UQCRC2* cDNA, with the exception of three short 18–60 bp stretches of sequence. Also in the case of this pseudogene, we noticed LINE/L1 repeat elements, which are on both sides flanking the sequence similar to the *UQCRC2* cDNA. Furthermore, all three ORFs contain multiple stop codons and no ESTs or cDNAs that show high similarity to the *UQCRC2*Ψ could be identified (details not shown).

Yet another pseudogene is located between the *CCR1* gene and the above-described *NRBF-2* pseudogene (*NRBF-2*Ψ) (Figure 1). Two regions of 2.7 and 1.4 kb, interrupted by insertion of 1.1 kb LINE/L1 repeat, show 88–96% similarity to the fms-related tyrosine kinase 1 cDNA (vascular endothelial growth factor/vascular permeability factor receptor, *FLT1*, Acc. NM_002019). The *FLT1* gene is located on chromosome 13 and contains 35 exons. The cDNA of the *FLT1* gene is 7680 bp and the ORF is located between 250 and 4266 bp. The similarity starts at position 2955 bp of the *FLT1* cDNA and continues with no intronic interruptions until the end of *FLT1* cDNA, except for five short no-similarity regions (between 27–49 bp). The ORF of the *FLT1*Ψ is interrupted by four stop codons. There are no detectable EST or cDNA sequences that show high similarity to *FLT1*Ψ.

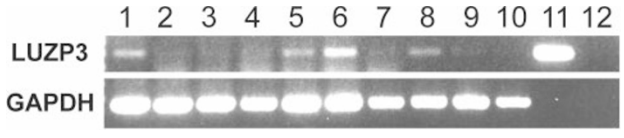

Figure 3 RT–PCR analysis of the *LUZP3* gene. The lane number 1–10 correspond to cDNA from IB-4 cell line, heart, brain, liver, kidney, trachea, lung, Marathon-ready cDNA from skeletal muscle, Marathon-ready cDNA from thymus and Marathon-ready cDNA from liver, respectively. As positive control (lane 11), DNA from clone RP6-91P17 was used. Lane 12 is a negative control.

Improved definition of the centromeric border of C3CER1

As summarised in Figure 1, the centromeric part of C3CER1 is very dense in active genes, which has prompted us to

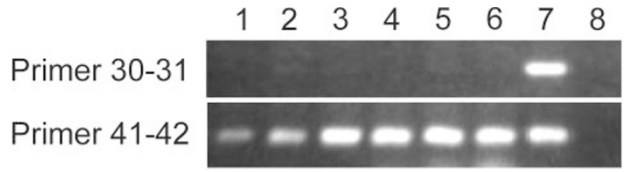

Figure 4 PCR analysis of human/mouse MCH (microcell hybrid) 908.6 line and derived SCID tumors.[16] Lanes 1 – 8 correspond to T5, T51, T52, T53, T54, T55, MCH 908.6 and negative control, respectively. Two new STS-es were used; primer 30-31 and primer 41-42 are located within the last exon of the *LRRC2* gene and within the first intron of the *LRRC2* gene, respectively.

redefine its centromeric border. We therefore tested the MHC906.8 microcell hybrid-derived panel of SCID tumours that was used in our previous study.[16] We have used six new STS-es that are located in RP6-91P17 (primer 22-23 within *TMEM7*, primer 30-31, 32-33, 34-35 within *LRRC2*, primer 41-42 within *LUZP3* and primer 49-50 within *TDGF1*). The last eliminated marker on the centromeric side of C3CER1 is primer 30-31, located in the 3′UTR region of *LRRC2* and the first retained marker is primer 41-42 located in the *LUZP3* gene, in the first intron of *LRRC2* (Figure 4). The distance between these two primers is approximately 40 kb. We can therefore conclude that the centromeric border of C3CER1 is positioned within the *LRRC2* gene.

Discussion

The C3CER1 chromosomal segment was identified based on the regular elimination of approximately 1 Mb from SCID-derived tumours.[15,16] This implies that C3CER1 contains one or several tumour growth antagonising genes. Our work fully defines the gene content of C3CER1, which is a prerequisite for understanding of its role in tumorigenesis. We characterised four novel C3CER1-located active genes and three processed pseudogenes. The assembling of our sequencing data, derived from shotgun sequencing of 11 PAC clones, together with the data available from the public databases, resulted in a comprehensive view of 1.4 Mb, containing 19 active genes. We can not exclude that additional genes are present within C3CER1. If as yet unknown transcriptional units exist in C3CER1, it is, however, unlikely that those will be identified based on further comparisons of C3CER1-derived genomic sequence with the content of EST databases. A large-scale comparison of genomic sequences between species is emerging as a powerful tool for exploring unknown complexity of genome anatomy, also with respect to not yet characterised genes. This approach also allows characterising important gene- or locus-specific regulatory elements.[34] The 19 active C3CER1-located genes (from the beginning of the telomeric end of *KIAA0028* to the centromeric end of *TDGF1* gene, Figure 1) occupy at least 1180 kb of genomic sequence. The combined length of the transcribed sequences (60.46 kb) constitute approximately 5.1% of all sequences, which is a higher number than the 3% of expressed sequences in the human genome commonly mentioned in the literature. Detailed Repeat_Masker-assisted analysis of the above mentioned 1180 kb region showed 46.9% total content of repeats and overall 43.6% G+C nucleotides. The contribution of major classes of repeats was as follows: SINEs, 12% (Alus 9.75% and MIRs 2.3%); LINE1, 15%; LINE2, 2.6%; LTR elements, 12.3%; and DNA transposons, 3.7%. A surprising finding was that the LINE1 elements were more predominant than the SINEs, in the region which can be considered as moderately high in its G+C content and which is certainly rich in active genes. The total number of individual repeat elements identified by Repeat_Masker was 262 and the three most predominant elements were MIR, LINE2 and AluSx, which occurred 158, 152 and 115 times, respectively.

The *FYCO1* gene contains a FYVE zink finger domain that can bind two Zn^{2+} ions and was named after four proteins containing this motif: Fab1, YOTB/ZK632.12, Vac1 and EEA1. The FYVE finger can bind with high specificity to the membrane lipid such as phosphatidylinositol-3-phosphate (PtdIns(3)P).[35] PtdIns(3)P is crucial regulator of a variety of biological processes in yeast and higher eukaryotes, for instance membrane trafficking, apoptosis, and cytoskeletal regulation.[36,37] According to the SMART, 42 human proteins contain one or two FYVE domains. The EIP1 appears to be the closest paralog of FYCO1, as it displays highest similarity almost throughout the entire EIP1 protein (aa 27 – 587) and has the same domain structure (RUN domain followed by coiled-coil- and FYVE-domains). The 3′UTR region of *FYCO1* cDNA shows high similarity to a reported mouse cDNA clone, named *Mem2*. The *Mem2* cDNA clone had been identified by differential display analysis of cDNA libraries prepared from unfertilized eggs and preimplantation embryos and named as *M*aternal *E*mbryonic *M*essage *2*.[38] Northern blot analysis using mouse cDNA probe revealed a predominant 8.2 kb band and a minor 4.5 kb transcript. The mouse *Mem2* gene was also mapped to the distal part of mouse chromosome 9. The *Mem2* mapping results are in agreement with our previous findings of synteny between C3CER1 and mouse chromosome 9F.[18,19] In conclusion, *Mem2* and *FYCO1* are orthologous genes.

The *TMEM7* gene encodes a liver-specific transcript with a predicted single transmembrane domain located near the C-terminus. We also noticed at the C-terminus a KKXX-like motif (VKTA) predicted to function as endoplasmic reticulum (ER) membrane retention signal. Retention of proteins in the ER is accomplished by a variety of mechanisms. One is employing specific signals to distinguish proteins to be maintained in the ER. Two kinds of retrieval signals for ER membrane proteins are known; one is the di-lysine motif (the KKXX motif) always present near the C-terminus of type I proteins. The other is the di-arginine motif (the XXRR motif), present near the N-terminus of type II proteins.[39] Since the TMEM7 protein has a di-lysine-like near its C-terminus, it is accordingly a type I ER retention protein.

The *LRRC2* gene has also a distinct tissue-specific expression pattern. Its transcript was detected only in heart, skeletal muscle and kidney. The LRRC2 protein contains seven Leucine-rich repeats (LRRs) which are relatively short motifs (22–28 aa) found in a variety of cytoplasmic, membrane and extracellular proteins.[40] LLRs proteins are associated with widely different functions, with a common characteristic involving protein–protein interactions. The closest relative of *LRRC2* is *RSP-1* (Ras Suppressor Protein 1) that plays a role in the *ras* signal transduction pathway. *RSP-1* is capable of suppressing v-*ras* transformation *in vitro*.[41] RSP-1 contains seven leucine-rich repeats, like LRRC2. Thus, there are both functional and positional reasons to study this gene further with regard to its possible role in tumorigenesis.

A striking feature of C3CER1 is the presence of a large cluster of chemokine receptors (Figure 1), which include eight genes; *CCR9*, *STRL33* (also named *TYMSTR* or *Bonzo*), *CCXCR1*, *CCR1*, *CCR3*, *CCR2*, *CCR5* and a chemokine receptors like *CCRL2* (also named *CRAM-B*). Families of chemokine genes (over 40 members) and chemokine receptor genes (16 members) occur in clusters on chromosome 4 and 17 and chromosome 2 and 3, respectively.[42] The *CCR5* gene has been identified as the major co-receptor for macrophage-tropic strains of HIV-1 virus. However, other chemokine and orphan receptors, such as *CCR2B*, *CCR3* and *STRL33*, have also been identified as potential co-receptors for HIV-1 virus.[43] CCRs all have a common structure of seven transmembrane domains, which is similar to the structure of G-protein-coupled receptors. Chemokines and their receptors mediate signals that are critical for the recruitment of effector immune cells to the site of inflammation. It might be hypothesised that regional elimination of a whole chemokine receptor cluster provides a selective advantage to the tumour cell, by escaping from the response to the inflammatory signals mediated by the chemokines.

In summary, we report a physical and transcriptional map of 1.4 Mb region on chromosome 3p21.3 containing 19 active genes. We have characterised four novel genes, ie FYVE and coiled-coil domain containing 1 (*FYCO1*), transmembrane protein 7 (*TMEM7*), leucine-rich repeat-containing 2 (*LRRC2*) and the leucine zipper protein 3 (*LUZP3*), and identified three processed pseudogenes (*NRBF-2*Ψ, *UQCRC2*Ψ, *FLT1*Ψ). This knowledge of gene content in C3CER1 provides a solid basis for further functional analysis of these genes and understanding of their role in tumour development.

Accession numbers for the sequences described in the paper:

AJ292348 Homo sapiens mRNA for FYVE and coiled-coil domain containing 1 gene (*FYCO1*)

AJ312776 Homo sapiens mRNA for transmembrane protein 7 gene (*TMEM7*)

AJ308569 Homo sapiens mRNA for leucine-rich repeat-containing 2 gene (*LRRC2*)

AJ312775 Homo sapiens mRNA for leucine zipper protein 3 gene (*LUZP3*)

AJ312684 Homo sapiens genomic sequence from 3p21.3 in 32 ordered contigs, clone RP5-1033N4

AJ312685 Homo sapiens genomic sequence partially covering the *KIAA0028* gene for mitochondrial leucyl-tRNA synthetase, exons 21-22

AJ312686 Homo sapiens genomic sequence partially covering the *LIMD1* gene, exons 1-2

AJ312687 Homo sapiens genomic sequence from 3p21.3 in 23 ordered contigs, clones RP6-153F14, RP6-146E1, RP6-188G11 and RP4-787C23

AJ312688 Homo sapiens genomic sequence from 3p21.3 in 26 ordered contigs, clones RP4-787C23, RP6-32G23, RP6-146E1, clone RP6-188G11

AJ310996 Homo sapiens genomic sequence from 3p21.3 in 42 ordered contigs, clone RP6-91P17

FYCO1, LRRC2, TMEM7 and LUZP3 gene symbols and ***C3CER1*** symbol for ''chromosome 3 common eliminated region 1'' have been approved by the HUGO Gene Nomenclature Committee, www.gene.ucl.ac.uk/nomenclature/

Acknowledgements

This work was supported by grants from the Swedish Cancer Foundation, the Swedish Medical Research Council, the Cancer Society in Stockholm, the Cancer Research Institute/Concern Foundation NY-LA, USA, Karolinska Hospital and Karolinska Institutet.

References

1 Mitelman F, Mertens F, Johansson B: A breakpoint map of recurrent chromosomal rearrangements in human neoplasia [see comments]. *Nat Genet* 1997; **15** Spec No: 417–474.
2 Kok K, Naylor SL, Buys CH: Deletions of the short arm of chromosome 3 in solid tumours and the search for suppressor genes. *Adv Cancer Res* 1997; **71**: 27–92.
3 Alimov A, Kost-Alimova M, Liu J *et al*: Combined LOH/CGH analysis proves the existence of interstitial 3p deletions in renal cell carcinoma. *Oncogene* 2000; **19**: 1392–1399.
4 Deng L, Jing N, Tan G *et al*: A common region of allelic loss on chromosome region 3p25.3-26.3 in nasopharyngeal carcinoma. *Genes Chromosomes Cancer* 1998; **23**: 21–25.
5 Murata Y, Tamari M, Takahashi T *et al*: Characterization of an 800 kb region at 3p22-p21.3 that was homozygously deleted in a lung cancer cell line. *Hum Mol Genet* 1994; **3**: 1341–1344.
6 Kok K, van den Berg A, Veldhuis PM *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.
7 Ohta M, Inoue H, Cotticelli MG *et al*: The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell* 1996; **84**: 587–597.
8 Rabbitts P, Bergh J, Douglas J, Collins F, Waters J: A submicroscopic homozygous deletion at the D3S3 locus in a cell line isolated from a small cell lung carcinoma. *Genes Chromosomes Cancer* 1990; **2**: 231–238.
9 Daigo Y, Isomura M, Nishiwaki T *et al*: Characterization of a 1200-kb genomic segment of chromosome 3p22-p21.3. *DNA Res* 1999; **6**: 37–44.

10 Kholodnyuk ID, Kost-Alimova M, Yang Y *et al*: The microcell hybrid based ''Elimination Test'' identifies a 1-Mb putative tumour suppressor region at 3p22.2-p22.1 centromeric to the homozygous deletion region detected in lung cancer. *Genes, Chromosomes and Cancer* 2001; in press.
11 Lerman MI, Minna JD: The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumour suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumour Suppressor Gene Consortium. *Cancer Res* 2000; **60**: 6116–6133.
12 Sekido Y, Ahmadian M, Wistuba, II *et al*: Cloning of a breast cancer homozygous deletion junction narrows the region of search for a 3p21.3 tumour suppressor gene. *Oncogene* 1998; **16**: 3151–3157.
13 Imreh S, Kholodnyuk I, Allikmetts R, Stanbridge EJ, Zabarovsky ER, Klein G: Nonrandom loss of human chromosome 3 fragments from mouse-human microcell hybrids following progressive growth in SCID mice. *Genes Chromosomes Cancer* 1994; **11**: 237–245.
14 Kholodnyuk I, Kost-Alimova M, Kashuba V *et al*: A 3p21.3 region is preferentially eliminated from human chromosome 3/mouse microcell hybrids during tumour growth in SCID mice. *Genes Chromosomes Cancer* 1997; **18**: 200–211.
15 Szeles A, Yang Y, Sandlund AM *et al*: Human/mouse microcell hybrid based elimination test reduces the putative tumor suppressor region at 3p21.3 to 1.6 cM. *Genes Chromosomes Cancer* 1997; **20**: 329–336.
16 Yang Y, Kiss H, Kost-Alimova M *et al*: A 1-Mb PAC contig spanning the common eliminated region 1 (CER1) in microcell hybrid-derived SCID tumours. *Genomics* 1999; **62**: 147–155.
17 Ishikawa S, Kai M, Tamari M *et al*: Sequence analysis of a 685-kb genomic region on chromosome 3p22-p21.3 that is homozygously deleted in a lung carcinoma cell line. *DNA Res* 1997; **4**: 35–43.
18 Kiss H, Kedra D, Yang Y *et al*: A novel gene containing LIM domains (LIMD1) is located within the common eliminated region 1 (C3CER1) in 3p21.3. *Hum Genet* 1999; **105**: 552–559.
19 Kiss H, Kedra D, Kiss C *et al*: The LZTFL1 Gene Is a Part of a Transcriptional Map Covering 250 kb within the Common Eliminated Region 1 (C3CER1) in 3p21.3. *Genomics* 2001; **73**: 10–19.
20 Peyrard M, Pan HQ, Kedra D *et al*: Structure of the promoter and genomic organization of the human beta'- adaptin gene (BAM22) from chromosome 22q12. *Genomics* 1996; **36**: 112–117.
21 Kedra D, Pan HQ, Seroussi E *et al*: Characterization of the human synaptogyrin gene family. *Hum Genet* 1998; **103**: 131–141.
22 Staden RE. *The Staden package*. Totawa: NJ, Humana Press Inc., 1994.
23 Nakai K, Kanehisa M: A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 1992; **14**: 897–911.
24 Nakai K, Horton P: PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* 1999; **24**: 34–36.
25 Schultz J, Milpetz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 1998; **95**: 5857–5864.
26 Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* 2000; **28**: 231–234.
27 Seroussi E, Pan HQ, Kedra D, Roe BA, Dumanski JP: Characterization of the human NIPSNAP1 gene from 22q12: a member of a novel gene family. *Gene* 1998; **212**: 13–20.
28 Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal Biochem* 1984; **137**: 266–267.
29 Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning; a laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989.
30 Ioannou PA, Amemiya CT, Garnes J *et al*: A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics* 1994; **6**: 84–89.
31 Wilbanks A, Zondlo SC, Murphy K *et al*: Expression cloning of the STRL33/BONZO/TYMSTRligand reveals elements of CC, CXC, and CX3C chemokines. *J Immunol* 2001; **166**: 5145–5154.
32 Zhang YJ, Zhang L, Ketas T, Korber BT, Moore JP: HIV type 1 molecular clones able to use the Bonzo/STRL-33 coreceptor for virus entry. *AIDS Res Hum Retroviruses* 2001; **17**: 217–227.
33 Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D: Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 2000; **10**: 1001–1010.
34 Onyango P, Miller W, Lehoczky J *et al*: Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res* 2000; **10**: 1697–1710.
35 Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Stenmark H, Aasland R: FYVE fingers bind PtdIns(3)P [letter; comment]. *Nature* 1998; **394**: 432–433.
36 Leevers SJ, Vanhaesebroeck B, Waterfield MD: Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr Opin Cell Biol* 1999; **11**: 219–225.
37 Rameh LE, Cantley LC: The role of phosphoinositide 3-kinase lipid products in cell function. *J Biol Chem* 1999; **274**: 8347–8350.
38 Heyer BS, Warsowe J, Solter D, Knowles BB, Ackerman SL: New member of the Snf1/AMPK kinase family, Melk, is expressed in the mouse egg and preimplantation embryo. *Mol Reprod Dev* 1997; **47**: 148–156.
39 Teasdale RD, Jackson MR: Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. *Annu Rev Cell Dev Biol* 1996; **12**: 27–54.
40 Rothberg JM, Jacobs JR, Goodman CS, Artavanis-Tsakonas S: Slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes Dev* 1990; **4**: 2169–2187.
41 Cutler ML, Bassin RH, Zanoni L, Talbot N: Isolation of rsp-1, a novel cDNA capable of suppressing v-Ras transformation. *Mol Cell Biol* 1992; **12**: 3750–3756.
42 Zlotnik A, Yoshie O: Chemokines: a new classification system and their role in immunity. *Immunity* 2000; **12**: 121–127.
43 Simmons G, Reeves JD, Hibbitts S *et al*: Co-receptor use by HIV and inhibition of HIV infection by chemokine receptor ligands. *Immunol Rev* 2000; **177**: 112–126.