

ARTICLE

# Physical and transcriptional map of the critical region for keratolytic winter erythema (KWE) on chromosome 8p22-p23 between D8S550 and D8S1759

Silke Appel<sup>1,9</sup>, Matthias Filter<sup>4</sup>, André Reis<sup>1,7</sup>, Hans Christian Hennies<sup>\*,1</sup>, Anton Bergheim<sup>2,9</sup>, Emma Ogilvie<sup>2</sup>, Silke Arndt<sup>2</sup>, Andrew Simmons<sup>5</sup>, Michael Lovett<sup>5</sup>, Winston Hide<sup>6</sup>, Michèle Ramsay<sup>2</sup>, Kathrin Reichwald<sup>3,9</sup>, Wolfgang Zimmermann<sup>3</sup> and André Rosenthal<sup>\*,3,8</sup>

<sup>1</sup>Department of Molecular Genetics and Gene Mapping Center, Max-Delbrück Center for Molecular Medicine, Berlin, Germany; <sup>2</sup>Department of Human Genetics, School of Pathology, The South African Institute for Medical Research and University of the Witwatersrand, Johannesburg, South Africa; <sup>3</sup>Genome Sequencing Center, Institute of Molecular Biotechnology, Jena, Germany; <sup>4</sup>Department of Bioinformatics, Max-Delbrück Center for Molecular Medicine, Berlin, Germany; <sup>5</sup>Division of Human Genetics, Washington University, School of Medicine, St. Louis, USA; <sup>6</sup>South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa; <sup>7</sup>Institute of Human Genetics, Friedrich-Alexander University, Erlangen, Germany; <sup>8</sup>Department of Biology, Friedrich-Schiller University, Jena, Germany

Keratolytic winter erythema is an autosomal dominant skin disorder characterised by erythema, hyperkeratosis, and peeling of the skin of the palms and soles, especially during winter. The keratolytic winter erythema locus has been mapped to human chromosome 8p22-p23. This chromosomal region has also been associated with frequent loss of heterozygosity in different types of cancer. To identify positional candidate genes for keratolytic winter erythema, a BAC contig located between the markers at D8S550 and D8S1695 was constructed and sequenced. It could be extended to D8S1759 by a partially sequenced BAC clone identified by database searches. In the 634 404 bp contig 13 new polymorphic microsatellite loci and 46 single nucleotide and insertion/deletion polymorphisms were identified. Twelve transcripts were identified between D8S550 and D8S1759 by exon trapping, cDNA selection, and sequence analyses. They were localised on the genomic sequence, their exon/intron structure was determined, and their expression analysed by RT-PCR. Only one of the transcripts corresponds to a known gene, encoding B-lymphocyte specific tyrosine kinase, *BLK*. A putative novel myotubularin-related protein gene (*MTMR8*), a potential human homologue of the mouse acyl-malonyl condensing enzyme gene (*Amac1*), and two transcripts showing similarities to the mouse L-threonine 3-dehydrogenase gene and the human SEC oncogene, respectively, were identified. The remaining seven transcripts did not show similarities to known genes. There were no potentially pathogenic mutations identified in any of these transcripts in keratolytic winter erythema patients.

*European Journal of Human Genetics* (2002) 10, 17–25. DOI: 10.1038/sj/ejhg/5200750

\*Correspondence: Dr HC Hennies, Department of Molecular Genetics and Gene Mapping Center, Max-Delbrück-Centrum, Robert-Rössle-Str. 10, 13125 Berlin, Germany. Tel: +49 30 94063532; Fax: +49 30 94063147; E-mail: hennies@mdc-berlin.de or Prof Dr André Rosenthal, metaGen GmbH, Oudenarder Str. 16, 13347 Berlin, Germany. Tel: +49 30 45082000; Fax: +49 30 45082001; E-mail: andre.rosenthal@metagen.de

<sup>9</sup>These authors contributed equally to this publication

Received 5 June 2001; revised 25 October 2001; accepted 26 October 2001

**Keywords:** transcript map; physical map; 8p22-p23; keratolytic winter erythema; KWE; BLK

## Introduction

Keratolytic winter erythema (KWE, OMIM 148370) is an autosomal dominant skin disorder of unknown etiology. It is also known as 'Oudtshoorn skin disease' in reference to the South African district where it was first described.<sup>1</sup> KWE is characterised by cyclical erythema, hyperkeratosis, and peeling of the skin of the palms and soles, arresting at major skin creases. The symptoms are especially noticeable during winter. The prevalence of KWE in the South African Afrikaans-speaking Caucasoid population is 1/7200,<sup>2</sup> and a founder effect has been confirmed with haplotype analysis.<sup>3</sup> In Germany, several cases have been reported.<sup>3</sup> The disorder usually manifests within the first 5 years of life, and the age of onset seems to be related to the severity of the phenotype, with individuals who display an early age of onset being more severely affected.<sup>2</sup> The gene causing KWE was localised to chromosome 8p22-p23 in a 5-cM interval between markers at D8S550 and D8S552 by linkage analyses, with a proposed ancestral recombination event pointing to a 1-cM interval between markers at D8S550 and D8S265.<sup>3</sup> The identification of the gene causing KWE will contribute to a better understanding of the process of epidermal differentiation.

The chromosomal region 8p22-p23 has also been associated with frequent loss of heterozygosity (LOH) in different types of cancer. There are several reports of LOH of chromosome 8p in breast cancer,<sup>4-9</sup> lung cancer<sup>10,11</sup> head and neck cancer,<sup>12</sup> prostate cancer,<sup>13</sup> hepatocellular carcinoma,<sup>14</sup> intrahepatic cholangiocarcinoma,<sup>15</sup> and urinary bladder cancer.<sup>16,17</sup>

This study reports on the construction of a transcript map between the markers at D8S550 and D8S1759 in the KWE critical region using exon trapping, cDNA selection, genomic sequencing, and sequence analyses.

## Materials and methods

### Construction of a BAC contig

The CITB Human BAC Libraries B and C (Invitrogen, Inchinnan, Scotland, UK) were initially screened by PCR using the primers for D8S550, D8S265, and D8S1593, and then using STS primers derived from the BAC ends (164D9T7F: 5'-AAA TCA AGG TGT TGG CTG GG/164D9T7R: 5'-ACA TCT CAC TCT GTC ACG CA; 169O5SP6F: 5'-GTG GAA GGA TAC TGA TGT GG/169O5SP6R: 5'-CAG CAG CCT CTC GTG TCT TA; 169O5T7F: 5'-CCT GGT CAG CTA ATT TCT GG/169O5T7R: 5'-AAA CAG ACT CCT TGG CAG CA).

### Exon trapping

Exon trapping experiments<sup>18</sup> were performed following the recommendations of the manufacturer (Exon Trapping System (Life Technologies, Karlsruhe, Germany)) with the

BAC clones CTC-306G11, CTC-493P15, CTB-164D9, CTB-169O5, CTB-65D4, CTC-271O15, and CTC-367I24. In brief, BAC DNA was digested with either *Pst*I or *Bam*HI and *Bgl*II and subcloned into the exon trapping vector pSPL3B. The vector constructs were transfected into *E. coli*. The DNA pool of different plasmid clones was then isolated and transfected into COS-7 cells using LIPOFECTACE Reagent (Life Technologies). After 2 days of incubation at 37°C, total RNA was isolated with TRIzol Reagent (Life Technologies). Reverse transcription with the vector primer SA2 (5'-ATC TCA GTG GTA TTT GTC AGC) was followed by a PCR with vector primers SA2 and SD6 (5'-TCT GAG TCA CCT GGA CAA CC). To eliminate vector-derived products the reaction mix was then digested with *Bst*XI, and a second PCR with the primers dUSA4 (5'-CUA CUA CUA CUA CAC CTG AGG AGT GAA TTG GTC G) and dUSD2 (5'-CUA CUA CUA CUA GTG AAC TGC ACT GTG ACA AGC TGC) was performed. The dUMP residues in these primers enabled a further subcloning into the UDG cloning vector pAMP10 after incubation of the PCR products with Uracil DNA Glycosylase (UDG). The size of the inserts of all transformants was determined by PCR followed by agarose gel electrophoresis. Products longer than the spliced vector were further analysed. All putative exons were sequenced using standard protocols for an A.L.F. automated DNA sequencer (Amersham Pharmacia Biotech, Freiburg, Germany) and an ABI 377 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA). By BLAST searches<sup>19</sup> of the generated exon sequences, several ESTs were identified, and corresponding clones obtained from the Resource Center of the German Human Genome Project (<http://www.rzpd.de>) or the HGMP Resource Center (<http://www.hgmp.mrc.ak.uk>) and completely sequenced.

### cDNA direct selection

cDNA direct selection was performed essentially as described by Morgan *et al.*<sup>20</sup> Four overlapping YAC clones, 770E9, 915H4, 737E5 and 773G4, from contig WC8.1, were identified from the Whitehead Institute website (<http://www.genome.wi.mit.edu>). Pulsed field gel electrophoresis was used to separate the YAC clones from individual yeast chromosomes and the sizes were compared to the expected sizes (770E9 (1400 kb); 915H4 (740 kb); 737E5 (800 kb); 773G4 (1570 kb)). The YACs were excised, purified and biotinylated using both random priming and nick translation. Salivary gland (Sal) and foetal brain (Fb) cDNA pools were constructed using both oligo dT and random priming from total cytoplasmic RNA. Linkers (Fb: 5'-CGA GAA TTC TGG ATC CTC; Sal: 5'-CCA CTG AAT TCT CAG TGA) were annealed to each of the pools separately. Repeat suppressed cDNA pools were hybridised for 48 h in solution to the biotin labelled genomic DNA to an intermediate  $C_{0t_{1/2}}$  of between 100 and 200. Streptavidin beads were added to the solution

and three washes,  $2 \times$  SSC/0.1% SDS,  $1 \times$  SSC/0.1% SDS, and  $0.1 \times$  SSC/0.1% SDS, at  $65^\circ\text{C}$  were performed to elute any cDNAs not bound to the genomic DNA within the YACs. An alkali wash was used to elute the selected cDNAs. Two rounds of hybridisation selection were performed to ensure that quasi-normalisation occurred to increase the chance of identifying low abundance transcripts. Selected cDNAs were amplified using the library specific linker primers. The two tissue specific cDNA pools were separately subcloned non-directionally into pAMP10 (Life Technologies). Nine hundred and sixty clones from each of the libraries were picked into microtiter dishes and gridded on a Biomek 1000 station (Beckman, Fullerton, CA, USA) at a high density, in duplicate, onto nylon filters. To eliminate clones containing repetitive sequences, the filters were hybridised with total human DNA and colonies showing positive hybridisation were excluded from further analysis. Redundancy screening was done by hybridisation. DNA from the isolated non-redundant cDNA clones was cycle sequenced with the M13R primer (5'-CAG GAA ACA GCT ATG AC) using fluorescent DyeDeoxy Terminators (FSkit) on an ABI 377 automated DNA sequencer (Applied Biosystems).

#### Genomic sequencing and analyses

The BAC clones CTC-306G11, CTC-493P15, CTB-164D9, CTB-169O5, CTB-65D4, CTC-271O23, and CTC-367I24 (CITB Human BAC Libraries, Invitrogen) were completely sequenced using a combination of shotgun and directed methods as described.<sup>21</sup> Data were collected on ABI 377 automated sequencers (Applied Biosystems) and assembled and edited using the GAP package.<sup>22</sup> The generated contiguous genomic sequence of 507 616 bp (acc. nos. AF131215, AF131216) and the draft sequence of BAC clone RP11-148O21 between D8S1695 and D8S1759 (acc. no. AC022239) were analysed using RUMMAGE<sup>23</sup> and several other programmes (<http://www.hgmp.mrc.ac.uk/NIX>) including BLAST and FASTA.<sup>19,24</sup> For analysis, the translated proteins were submitted to the GeneQuiz system (<http://jura.ebi.ac.uk:8765>) and the ExPASy molecular biology server (<http://www.expasy.ch>).

#### RT-PCR

About 150 ng of total RNA from skin, primary keratinocytes, HeLa cells, and lymphoblastoid cell lines, respectively, were used for cDNA synthesis with either oligo(dT)<sub>15</sub> or random hexamers as primers in a  $10 \mu\text{l}$  reaction. One microlitre was then used in a  $15 \mu\text{l}$  PCR. The PCR products were either sequenced directly or cloned into the TA vector pCR2.1 (Invitrogen) and sequenced using standard protocols for A.L.F. automated DNA sequencers (Amersham Pharmacia Biotech) or ABI 377 automated DNA sequencers (Applied Biosystems) with primers –21M13F (5'-TGT AAA ACG ACG GCC AGT), M13R primer (5'-CAG GAA ACA GCT ATG AC), or gene specific primers. The sequences were analysed using BLASTN,<sup>19</sup> and corresponding cDNA clones<sup>25</sup> were ordered

from either the German Resource Center (<http://www.rzpd.de>) or the HGMP Resource Center (<http://www.hgmp.mrc.ac.uk>) and completely sequenced.

#### Expression studies

Patterns of transcription of the isolated cDNAs were studied by RT-PCR experiments using commercially available cDNA pools from 12 different tissues (set 1 and set 2 (Origene, Rockville, MD, USA)), and cDNA from skin, primary keratinocytes, HeLa cells, and lymphoblastoid cell lines. The primers used were as follows: *BLK*: BLK4F (5'-TTG CTC CAA TCA ACA AGG CC) and BLK4R (5'-ACA TGG TTC CCT CCT TCA GC); *MTMR8*: S8aG4F (5'-GAT GAA GCT CTT CGG AAG GT) and S9A10R (5'-CTC TTT GAT GTG AGT CAG CC); *TDH*: vir17F (5'-GTT CAT TAG GAT GCT GAG GC) and vir17R (5'-ATA GAT GGT CCT GGG TCT CT); C8orf13: vir3F (5'-GAG ACC CAC TGG AGT AAC TT) and vir3R (5'-GAA GTG GGT GCA GAA GAG GG); *AMAC*: vir8F (5'-CTG AGT TGG AGT TGT GTG GG) and vir8R (5'-CGG GCT CCC AAG TTC TAT CT); C8orf12: vir33F (5'-AGG TCT CCC TGC TTC TTC AA) and vir33R (5'-TCC CTT GCC AAT GTA ATC GG); C8orf14: vir35F (5'-AAG CTC TCA TCC AAT GTC CC) and vir35R (5'-ATT AGT CCA GGG TGA GTC TG); C8orf7: vir11F (5'-TGT AGT CCT GCA CGA ACC AG) and vir11R (5'-TCG ATG CAC ATC TGC CAC TG); C8orf8: vir32F (5'-GCA GAC ATG ATG GCA TCT TA) and vir32R (5'-GAA GGT CCT TGT GCA TGA AG); C8orf5: vir15F (5'-TGA GCA CAA ATG AAA GCG AC) and vir15R (5'-TGA GCC ACA TAT CCA TTC AG); C8orf9: 205798\_5rev (5'-TGC TGG AGC CCA CAA CAA CT) and 205798pr (5'-CCT GGA GGA TTC CGT AAG GT); C8orf6: vir25F (5'-TTT CTA CAA TGA CCC ACC AC) and vir25R (5'-TGA CCA CAT ACA GCT CTT CT).

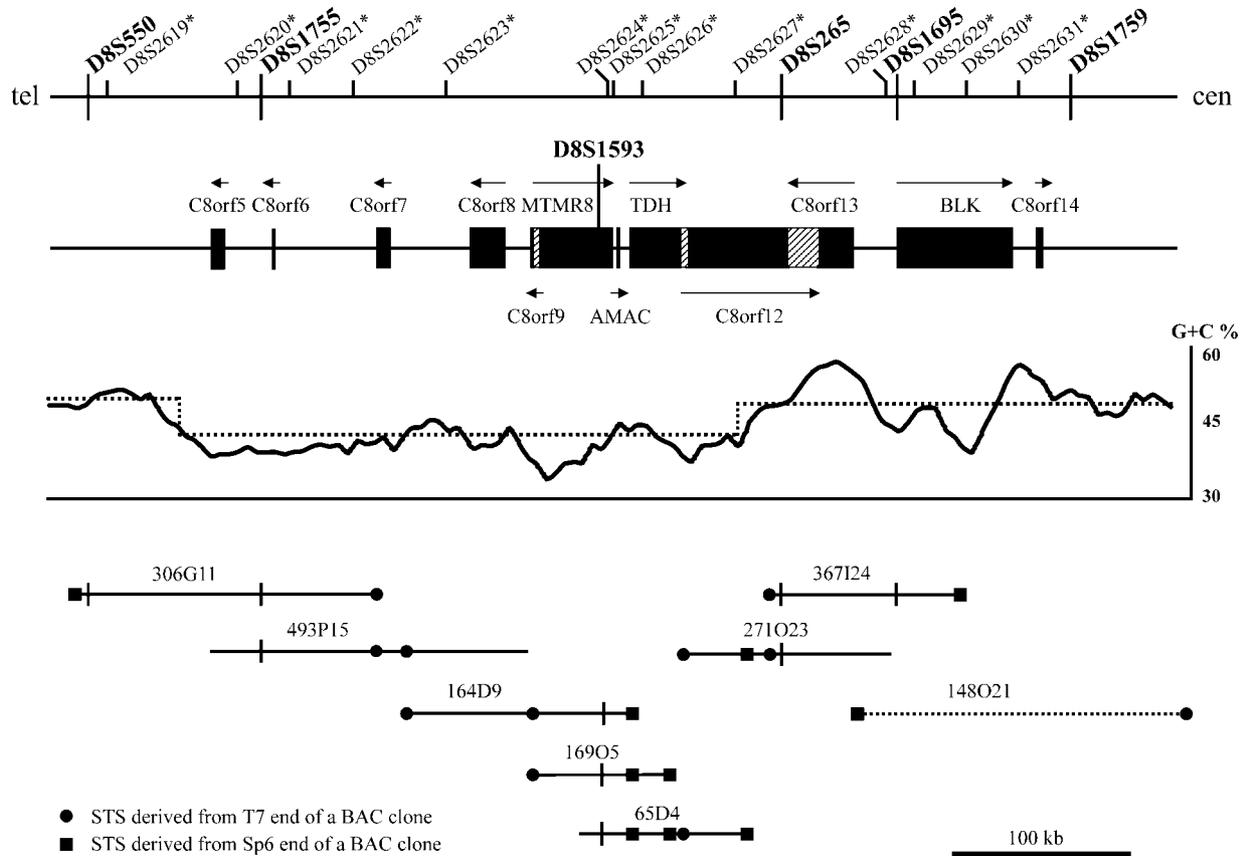
#### Mutation analyses

Mutational analyses were performed with genomic DNA of at least one healthy German control individual and one or more patients from the German KWE pedigree investigated in the original linkage study.<sup>3</sup> Each exon was amplified by PCR, directly sequenced using the BigDye terminator cycle sequencing kit (Applied Biosystems), and electrophoresed on an ABI 377 automated DNA sequencer (Applied Biosystems).

## Results and discussion

### Construction of a BAC contig and genomic sequencing

As an essential step towards identifying the KWE gene, a complete physical map has been generated spanning the KWE critical region between the markers at D8S550 and D8S265.<sup>3</sup> These two markers were initially used to screen a BAC library. Additional STSs generated by end-sequencing of BACs were used in a subsequent screening to close remaining gaps (Figure 1). Seven BAC clones covering the region were completely sequenced and a contiguous genomic sequence of 507 616 bp was generated (acc. nos. AF131215,



**Figure 1** Physical and transcript map of the KWE gene critical region on human chromosome 8p22-p23. The order of previously identified microsatellite markers (bold) is given as determined by the genomic sequence, and the positions of 13 new microsatellite markers (marked by asterisks) are shown at the top. The transcript map is shown underneath. Twelve transcripts, depicted as filled boxes, were identified using various approaches. The EST locus D8S1593 is part of the novel myotubularin-related protein gene *MTMR8*. Partial overlaps of genes are shown as hatched boxes. The direction of transcription is presented by arrows. Local analysis of the genomic sequence suggested three different G+C content domains. The average G+C content of these regions is shown by the dotted line. BAC clones were identified by PCR-screening of a human BAC library. The BAC clones that cover the whole region between D8S550 and D8S1695 were completely sequenced. The contig was extended to D8S1759 by the partially sequenced BAC clone RP11-148O21, which was identified by database searches.

AF131216). Within this sequence, markers at D8S550, D8S1755, D8S265, and D8S1695 were identified. In addition, the contig could be extended to D8S1759 by the overlapping BAC clone RP11-148O21 (acc. no. AC022239) reported to consist of six unordered contigs. Three of these were ordered based on overlaps with BAC CTC-367I24 (acc. no. AF131216), another two contigs were ordered by alignments to human *BLK* mRNA (acc. no. S76617) which also defined their orientation relative to the latter contig, and the orientation of the remaining contig is given by the clone end. Thus, a genomic sequence of 634 404 bp comprised of four contigs of 518 835, 18 713, 77 179, and 19 677 bp was generated and analysed. In contrast to the Génethon linkage map,<sup>26</sup> in which the order of markers is tel-D8S1755-(D8S1695-D8S550)-D8S265-D8S1759-cen, we found the order of

markers as tel-D8S550-D8S1755-D8S265-D8S1695-D8S1759-cen. In addition, we identified 13 new polymorphic microsatellite loci (D8S2619 through D8S2631) between D8S550 and D8S1759 (Figure 1).

The average G+C content of the 634 kb region is 45% (Figure 1). Local analysis revealed compositionally distinct regions having average G+C contents of 49% (50 kb in the distal portion, positions 1–50 000), 42% (300 kb representing the central portion, positions 50 001–350 000) and 48% (284 kb representing the proximal part of contig). These differences suggest that the analysed region spans different G+C content domains: tel-H2-L-H2-cen. Seven CpG islands were identified in the genomic region (length  $\geq 500$  bp, G+C  $\geq 50\%$ , ratio of observed to expected CpGs  $\geq 0.6$ <sup>27</sup>). Five of these are associated with genes or transcription units.

**Table 1** Transcripts isolated from the region between D8S550 and D8S1759

Transcript	Accession number	Direction of transcription	Position in genomic sequence	Number of SNPs/indels <sup>a</sup>	cDNA (bp)	Number of exons	ORF (codons)	Exon trapping	Exons isolated by cDNA selection	Genomic sequencing	BLASTX similarities
C8orf5	AJ305312	Reverse	85026–86961 <sup>b</sup>	–	1936	1	97	–	1/1	1/1	–
C8orf6	AJ307469	Reverse	114137–115842 <sup>b</sup>	3	1706	1	95	–	1/1	1/1	–
C8orf7	AJ301560	Reverse	171701–178573 <sup>b</sup>	2	1519	2	288	1/2	2/2 <sup>f</sup>	2/2	–
C8orf8	AJ301561	Reverse	38569–58818 <sup>c</sup>	2	241	2	31	2/2	–	–	–
C8orf9	AJ291676	Reverse	74458–76026 <sup>c</sup>	–	1569	1	121	–	–	1/1	–
MTMR8	AJ297823	Forward	75765–119095 <sup>c</sup>	3	7081	10	550	6/10	9/10 <sup>f</sup>	10/10	myotubularin acyl-malonyl condensing enzyme (mouse)
AMAC	AJ291677	Forward	121944–123144 <sup>c</sup>	–	1201	1	339	–	1/1	1/1	L-threonin 3-dehydrogenase (mouse)
TDH	AJ301562	Forward	130595–159410 <sup>c</sup>	11	1357	9	250	2/9	–	4/9	–
C8orf12	AJ301563	Forward	159360–229615 <sup>c</sup>	6	1222	7	105	–	–	5/7	–
C8orf13	AJ301564	Reverse	212422–257634 <sup>c</sup>	9	4002	3	215	1/3	1/3	3/3	SEC protein
BLK	S76617	Forward	284958–>321259 <sup>d</sup>	3	2608	13	1518	–	2/13 <sup>f</sup>	13/13	B-lymphocyte specific tyrosine kinase
C8orf14	AJ291678	Forward	16017–20823 <sup>e</sup>	–	1927	3	93	–	–	1/3	–

<sup>a</sup>Insertion/deletion polymorphisms. <sup>b</sup>Position in acc. no. AF131215. <sup>c</sup>Position in acc. no. AF131216. <sup>d</sup>BLK goes beyond position 321 259 of AF131216. The part not included there is present in AC022239. <sup>e</sup>Position in acc. no. AC022239. <sup>f</sup>Various cDNA clones.

Overall, about 41% of the 634 kb segment is composed of repetitive elements, such as SINEs (12%), LINEs (16%), and LTR elements (7%). Locally, the repeat content is low in the distal region of high G+C content with an average of 36%, and rises to 42% in the central and proximal portions of the region analysed.

#### Identification of transcripts by exon trapping, cDNA selection, and sequence analysis

Various approaches were taken to identify positional candidate genes in the KWE critical region. Exon trapping was used to generate novel ESTs from the seven BAC clones of the sequenced genomic contig. Five different cDNAs (corresponding to parts of C8orf7, C8orf8, *MTMR8*, *TDH*, C8orf13) were initially derived from 12 trapped exons by RT-PCR (Table 1). BLAST searches of their sequences identified several ESTs extending the cDNAs, which were then confirmed by RT-PCR. Moreover, the transcription of 21 predicted exons contained in the transcripts C8orf7, *TDH*, *MTMR8*, and C8orf13 could be verified by RT-PCR. The cDNA selection using four YAC clones from the contig WC8.1 of the Whitehead Institute led to the initial isolation of 790 non-redundant cDNA clones (acc. nos. AW408850 to AW409398, AW441142 to AW441186, and AW737018 to AW737061).<sup>28</sup> The resulting EST sequences were compared to the genomic sequence, and of 128 clones with at least 90% sequence identity, 26 contained exonic sequences. Four additional transcripts (corresponding to C8orf5, C8orf6, *AMAC*, *BLK*) were detected using selected EST sequences. Furthermore, they confirmed C8orf7, *MTMR8*, and C8orf13. Another three transcripts (C8orf9, C8orf12, C8orf14) were identified by BLAST searches of the genomic sequence between D8S550 and D8S1759 (Table 1).

By far the most exons of the transcripts described could be identified by direct analysis of the genomic sequence using various techniques (79%; Table 1). The total number of exons identified by exon trapping (91) is relatively small, however, 61 of these were confirmed as transcribed sequences and mapped to the original region (data not shown). In contrast, 378 exons were predicted in the genomic sequence by at least one algorithm but only 23% of the exons analysed could be confirmed experimentally. At least in the tissues tested, the number of false positive sequences obtained by exon trapping is thus rather small (33%) as compared to cDNA selection (80%) and genomic sequence analysis by exon prediction (77%). In particular, it became obvious that a combination of different tools to predicting exonic sequences<sup>23</sup> was opportune to diminish false positive results. Although there is a considerable overlap between the exons found by the different approaches described, a number of transcripts would not have been isolated with one method alone. These findings revealed that it is still rather useful to combine the current availability of human genomic sequences with at least one other method of searching transcribed sequences.

**Table 2** Expression patterns of the transcripts based on RT-PCR

Tissue	C8orf5	C8orf6	C8orf7	C8orf8	C8orf9	MTMR8	AMAC	TDH	C8orf12	C8orf13	BLK	C8orf14
HeLa	—	+	+	—	+	+	+	—	—	+	—	+
Lymphoblastoid cell line	n.d.	+	+	+	+	+	n.d.	+	+	n.d.	+	+
Primary keratinocytes	+	+	+	—	+	+	+	—	+	+	—	+
Skin	+	+	+	—	+	+	+	—	+	+	—	+
Heart	+	+	—	—	+	+	+	—	+	—	—	—
Spleen	+	+	—	—	+	+	+	—	—	+	+	—
Brain	+	—	—	—	+	+	+	—	+	—	—	—
Liver	—	—	—	—	—	+	—	—	—	—	+	—
Leukocytes	+	+	—	—	—	+	+	+	—	+	+	—
Kidney	+	+	—	—	+	+	+	—	+	+	—	—
Ovary	+	+	—	—	+	+	+	—	+	—	+	—
Muscle	+	+	—	—	+	+	+	+	+	—	+	—
Testis	+	+	—	+	+	+	+	+	+	+	+	+
Lung	+	+	—	—	+	+	+	+	+	+	—	—
Small intestine	+	+	+	—	+	+	+	+	+	+	—	—
Prostate	—	+	—	—	+	+	+	—	—	+	—	—

### Transcript map between D8S550 and D8S1759

Altogether, 12 transcripts were identified in the region between D8S550 and D8S1759 (Figure 1; Table 1). Transcription units as well as corresponding genes were localised on the genomic sequence and their exon/intron structures determined. The transcription was analysed by RT-PCR with cDNAs from different tissues. The expression patterns are shown in Table 2.

One known gene, *BLK*, encoding B-lymphocyte specific tyrosine kinase, was identified and localised between D8S1695 and D8S1759. The gene is comprised of 13 exons spanning about 70 kb. It harbours an open reading frame of 1518 codons starting in exon 2 and terminating in exon 13. In contrast to Drebin *et al.*,<sup>29</sup> the transcription of *BLK* could be detected in lymphoblastoid cell lines, spleen, liver, leukocytes, ovary, muscle, and testis (Table 2). As has been described for genes having tissue specific expression,<sup>27</sup> the 3' end of *BLK* is associated with a CpG island of 1097 bp in length and a G+C content of 70%. *BLK* is a member of the SRC family, which is thought to play an important role in the signalling pathways controlling cell proliferation and differentiation.<sup>30</sup> *BLK* may also be involved in membrane attachment and thymopoeisis.<sup>29,30</sup> This gene is therefore a good positional candidate for the cancers mapping to this region.

A putative novel myotubularin-related protein gene<sup>31</sup> (*MTMR8*, acc. no. AJ297823), containing D8S1593 in its 3' UTR, was identified between D8S1755 and D8S265. It is comprised of 10 exons spanning 43 kb. The 5' end of the gene is associated with a CpG island of 792 bp and a G+C content of 70%. The island also shows transcriptional activity in the opposite direction. The cDNA corresponding to *MTMR8* is 7081 bp in length as confirmed by Northern blot analysis. The open reading frame predicts a primary structure of 549 residues. It starts in exon 1 and terminates in the last exon. Based on the similarity of the polypeptide to members of the

myotubularin protein family (32–40%), the gene was named myotubularin-related protein gene 8. The gene is expressed in all 16 tissues investigated by RT-PCR. So far, eight other proteins belonging to the myotubularin family have been identified, and two of these have been associated with diseases. Myotubularin (*MTM1*) is mutated in X-linked myotubular myopathy, a severe recessive congenital muscle disorder<sup>32</sup> (XLMTM, OMIM 310400), and mutations in myotubularin-related protein 2 (*MTMR2*) have recently been reported to cause Charcot-Marie-Tooth disease type 4B, an autosomal recessive demyelinating neuropathy with myelin outfoldings.<sup>33</sup>

The gene product of *TDH* (acc. no. AJ301562) shows similarities to murine L-threonine 3-dehydrogenase (79% identity in 200 residues, acc. no. AF134346), to the CG5955 gene product of *D. melanogaster* (55% identity in 194 residues, acc. no. AAF51607) and the hypothetical protein F08F3.4 of *C. elegans* (51% identity in 187 residues, acc. no. T29433). The human gene consists of nine exons spanning 30 kb, and its 5' end is associated with a 70% GC-rich CpG island of 1195 bp in length. The cDNA encoding *TDH* is 1357 bp in length and harbours an open reading frame of 250 codons, which starts in the first exon and ends in exon 7. *TDH* is expressed in only six of the investigated tissues (Table 2). There are at least three different splice forms of *TDH* in lymphoblastoid cells. We detected an alternatively spliced exon of 111 bp between exons 1 and 2. The resulting transcript contains an open reading frame of 231 codons that starts in exon 2. A further alternatively spliced exon of 59 bp was detected between exons 6 and 7. This splice variant contains an open reading frame of more than 280 codons.

C8orf13 (acc. no. AJ301564) contains D8S265 in its 3' UTR. The gene is comprised of three exons spanning 45 kb. Its 5' end is associated with a CpG island of 1097 bp in length and a G+C content of 70%. The transcript was detected in 10 of the 15 tissues tested (Table 2). The open reading frame

predicts a primary structure of 214 amino acids and shows similarities to the human SEC oncogene (37% identities in 83 residues, acc. no. X52259), which is involved in human breast, colon, and prostate carcinomas.<sup>34</sup> This is noteworthy with respect to the frequent LOH reported for this region. In particular, Wang *et al.*<sup>5</sup> and Richard *et al.*<sup>8</sup> observed LOH in 8p22-pter in invasive ductal breast carcinoma, and Sigbjörnsdóttir *et al.*<sup>9</sup> noticed significantly more frequent chromosomal deletions between D8S550 and D8S1752 in patients with *BRCA2* 999del5 linked breast cancer than in patients with sporadic breast cancer. Kawaki *et al.*<sup>15</sup> reported LOH in intrahepatic cholangiocarcinoma between markers D8S265 and D8S258. In bladder cancer, Takle and Knowles<sup>16</sup> observed deletions between markers D8S264 and D8S133. Also, the region between markers D8S277 and D8S258 is reported as a 'hot spot' of allelic loss in small cell and non-small cell lung cancer,<sup>11</sup> which was observed early in the pathogenesis of lung cancer.<sup>10</sup> Therefore, C8orf13 is a good candidate for a tumour suppressor gene thought to be located in this chromosomal region.

AMAC (acc. no. AJ291677) exhibits high similarity to murine acyl-malonyl condensing enzyme. In contrast to the mouse gene (*Amac1*, acc. no. NM\_019871), however, the human transcript AMAC consists of a single exon. The longest open reading frame starts at position 122 and is 1017 bp in length. Remarkably, all three possible forward open reading frames, which were verified by RT-PCR experiments and genomic sequencing, show similarities to different parts of the mouse protein (in detail, positions 845–1135, reading frame +2, 82% identity in 97 residues; positions 237–530, reading frame +3, 56% identity in 99 residues; positions 574–732, reading frame +1, 64% identity in 53 residues). This indicates that AMAC might be a non-functional copy of the gene and hence a transcribed pseudogene. Moreover, a 3'-polyadenylation tract, which is present in the genomic DNA adjacent to AMAC, and the absence of introns are characteristic of pseudogenes that have arisen by retrotransposition.<sup>35</sup> In the human genome, three additional sequences with high similarities are located on 17p13.1 (acc. no. AC007732.3), 18p11.2 (acc. no. AC022030.2), and 17q11.2 (acc. no. AC022706.4). Functional analyses, however, are still necessary to investigate whether AMAC is the homologue of murine *Amac1* or a transcribed pseudogene.

The remaining seven transcripts do not show similarities to known genes (Table 1). The gene encoding C8orf12 (acc. no. AJ301563) spans 70 kb and consists of seven exons. The longest open reading frame of 105 codons starts in exon five and ends in the last exon. This transcript was detected in 11 of the 16 tissues tested by RT-PCR. C8orf14 (acc. no. AJ291678) is only transcribed in HeLa cells, lymphoblastoid cell lines, primary keratinocytes, skin, and testis. It is 1927 bp in length and consists of three exons. The last exon harbours the longest open reading frame of 93 codons. C8orf7 (acc. no. AJ301560) is comprised of two exons with an open reading

frame of 288 codons and spans about 7 kb. The transcript was detected in five of the 16 tissues investigated. C8orf8 (acc. no. AJ301561) consists of two exons which are separated by an intron of 20 kb in length. It was isolated by exon trapping and shown to be transcribed in lymphoblasts and testis. In contrast, C8orf5 (acc. no. AJ305312) could be detected in almost every tissue tested. This transcript is generated by only one exon. The same is true for C8orf9 (acc. no. AJ291676) and C8orf6 (acc. no. AJ307469).

Interestingly, the genes corresponding to transcripts *TDH* and C8orf12 partly overlap in their 3' and 5' UTRs, respectively. Since they are transcribed in the same direction, one can assume they may be controlled by the same mechanisms. However, this is apparently not a cell-type specific regulation as they have different expression patterns (Table 2), and there is no clue to a possible functional relationship between both genes so far. The overlapping genes C8orf12 and C8orf13 are transcribed from different DNA strands. The last four exons of C8orf12 harbouring the open reading frame are positioned in the second intron of C8orf13. C8orf9 and the gene encoding *MTMR8* also overlap and are transcribed from opposing DNA strands resulting in mRNA molecules sharing a complementary 5' UTR sequence. As C8orf9 is an intronless gene with a short open reading frame, it might function as an antisense control element of *MTMR8*. Although overlapping genes occur frequently in viral genomes as well as in genomes of cellular prokaryotes and prokaryote-derived organelles such as mitochondria,<sup>36</sup> they are less frequent in eukaryotes. Generally, the overlap of two genes transcribed from different DNA strands is more common<sup>37–44</sup> than the overlap of genes transcribed from the same DNA strand.<sup>45,46</sup> The functional role of the arrangement of such antisense genes has not been determined in most cases. Several functions have been proposed including transcriptional inhibition by steric hindrance of the two genes transcribed at the same time, interference of splicing and processing, inhibition of translation by formation of RNA duplexes, and effects on mRNA stability.

### Mutation analyses

In order to determine whether one of these transcripts was indeed the KWE gene, each exon belonging to a transcript was screened for mutations by direct sequencing of genomic DNA from unaffected individuals and KWE patients of the German pedigree linked to this region.<sup>3</sup> A total of 46 single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms were identified (for details, see acc. nos. AF131215 and AF131216). Thirty-nine of these are located within the genes described here (Table 1), including seven coding SNPs. In C8orf6, the exchange of T to C at position 1058 leads to the putative substitution of tryptophan by leucine at position 38 (W38L), and in *MTMR8* an exchange of A to T leads to the alteration of methionine to leucine (M200L). A further coding SNP, I104R, caused by a T to G exchange at position 352, was discovered in *TDH*, and in

C8orf12 the substitution of T to A at position 1105 affecting the stop codon results in an elongation of the predicted peptide by five residues, KCLSP. In C8orf13, two further coding SNPs were identified, H56Q, caused by an exchange of T to G at position 616, and T107S, caused by a C to G substitution at position 768. The exchange of A to C at position 1654 in C8orf14 leads to a potential exchange of threonine to proline (T67P). All the polymorphisms were validated by the analysis of 10 healthy individuals from five independent German pedigrees. The KWE causing mutation, though, was not identified in any of the transcripts, making it very unlikely that one of these is the KWE gene. As promoters and introns were not analysed exhaustively, it cannot be excluded that one of these harbours the pathogenic mutation. However, it seems now necessary to confirm the refined localisation of the KWE gene. As the proximal boundary of the region at D8S265 is only based on one family with a proposed ancestral recombination,<sup>3</sup> further members of this family should be recruited and new families analysed to ensure the exact localisation of the gene. However, the identification of the novel genes described here and the physical mapping of new microsatellite markers and SNPs will facilitate the further analysis of disease loci that map to chromosome 8p22-p23.

#### Acknowledgments

We wish to thank Susanne Lützkendorf, Uta Petz and Claudia Wenderoth for excellent technical assistance. We are grateful to Dr Matthias Platzer for his critical reading of the manuscript and many helpful suggestions. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the German Federal Ministry of Education, Research and Technology through Projektträger DLR in the framework of the German Human Genome Project, the Ministry of Science, Research and Art of the Free State of Thuringia, the South African Institute for Medical Research, and the South African Medical Research Council. S. Appel was supported by the Max-Delbrück Center for Molecular Medicine (Berlin). A. Bergheim and E. Ogilvie received bursaries from the National Research Foundation and the University of the Witwatersrand (Johannesburg).

#### References

- Findlay GH, Nurse GT, Heyl T *et al*: Keratolytic winter erythema or 'Oudtshoorn skin': a newly recognized inherited dermatosis prevalent in South Africa. *S Afr Med J* 1977; **52**: 871–874.
- Hull PR: Keratolytic winter erythema (Oudtshoorn disease): clinical, genetic and ultrastructural aspects. PhD thesis, University of the Witwatersrand, Johannesburg, 1986.
- Starfield M, Hennies HC, Jung M *et al*: Localization of the gene causing keratolytic winter erythema to chromosome 8p22-p23, and evidence for a founder effect in South African Afrikaans-speakers. *Am J Hum Genet* 1997; **61**: 370–378.
- Kerangueven F, Noguchi T, Coulier F *et al*: Genome-wide search for loss of heterozygosity shows extensive genetic diversity of human breast carcinomas. *Cancer Res* 1997; **57**: 5469–5474.
- Wang JC, Radford DM, Holt MS *et al*: Sequence-ready contig for the 1.4-cM ductal carcinoma in situ loss of heterozygosity region on chromosome 8p22-p23. *Genomics* 1999; **60**: 1–11.
- Yokota T, Yoshimoto M, Akiyama F *et al*: Localization of a tumor suppressor gene associated with the progression of human breast carcinoma within a 1-cM interval of 8p22-p23.1. *Cancer* 1999; **85**: 447–452.
- Forozan F, Mahlamaki EH, Monni O *et al*: Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Res* 2000; **60**: 4519–4525.
- Richard F, Pacyna-Gengelbach M, Schlüns K *et al*: Patterns of chromosomal imbalances in invasive breast cancer. *Int J Cancer* 2000; **89**: 305–310.
- Sigbjörnsdóttir BI, Ragnarsson G, Agnarsson BA *et al*: Chromosome 8p alterations in sporadic and BRCA2 999del5 linked breast cancer. *J Med Genet* 2000; **37**: 342–347.
- Wistuba II, Behrens C, Virmani AK *et al*: Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. *Cancer Res* 1999; **59**: 1973–1979.
- Girard L, Zochbauer-Muller S, Virmani AK, Gazdar AF, Minna JD: Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. *Cancer Res* 2000; **60**: 4894–4906.
- Li X, Lee NK, Ye YW *et al*: Allelic loss at chromosome 3p, 8p, 13q, and 17p associated with poor prognosis in head and neck cancer. *J Natl Cancer Inst* 1994; **86**: 1527–1529.
- MacGrogan D, Levy A, Bostwick D, Wagner M, Wells D, Bookstein R: Loss of chromosome arm 8p loci in prostate cancer: mapping by quantitative allelic imbalance. *Genes Chromosomes Cancer* 1994; **10**: 151–159.
- Qin LX, Tang ZY, Sham JS *et al*: The association of chromosome 8p deletion and tumor metastasis in human hepatocellular carcinoma. *Cancer Res* 1999; **59**: 5662–5665.
- Kawaki J, Miyazaki M, Ito H *et al*: Allelic loss in human intrahepatic cholangiocarcinoma: correlation between chromosome 8p22 and tumor progression. *Int J Cancer* 2000; **88**: 228–331.
- Takle LA, Knowles MA: Deletion mapping implicates two tumor suppressor genes on chromosome 8p in the development of bladder cancer. *Oncogene* 1996; **12**: 1083–1087.
- Ohgaki K, Iida A, Ogawa O, Kubota Y, Akimoto M, Emi M: Localization of tumor suppressor gene associated with distant metastasis of urinary bladder cancer to a 1-Mb interval on 8p22. *Genes Chromosomes Cancer* 1999; **25**: 1–5.
- Buckler AJ, Chang DD, Graw SL *et al*: Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc Natl Acad Sci USA* 1991; **88**: 4005–4009.
- Altschul SF, Madden TL, Schaffer AA *et al*: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
- Morgan JG, Dolganov GM, Robbins SE, Hinton LM, Lovett M: The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 1 and 5 genes. *Nucleic Acids Res* 1992; **20**: 5173–5179.
- Platzer M, Rotman G, Bauer D *et al*: Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. *Genome Res* 1997; **7**: 592–605.
- Dear S, Staden R: A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res* 1991; **19**: 3907–3911.
- Taudien S, Rump A, Platzer M *et al*: RUMMAGE – a high-throughput sequence annotation system. *Trends Genet* 2000; **16**: 519–520.
- Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988; **85**: 2444–2448.
- Lennon G, Auffray C, Polymeropoulos M, Soares MB: The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 1996; **33**: 151–152.

- 26 Dib C, Faure S, Fizames C *et al*: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; **380**: 152–154.
- 27 Larsen F, Gundersen G, Lopez R, Prydz H: CpG islands as gene markers in the human genome. *Genomics* 1992; **13**: 1095–1107.
- 28 Bergheim A: A high density transcript map of 8p22: Positional candidate genes for keratolytic winter erythema. PhD thesis, University of the Witwatersrand, Johannesburg, 2001.
- 29 Drebin JA, Hartzell SW, Griffin C, Campbell MJ, Niederhuber JE: Molecular cloning and chromosomal localization of the human homologue of a B-lymphocyte specific protein tyrosine kinase (blk). *Oncogene* 1995; **10**: 477–486.
- 30 Islam KB, Rabbani H, Larsson C, Sanders R, Smith CI: Molecular cloning, characterization, and chromosomal localization of a human lymphoid tyrosine kinase related to murine Blk. *J Immunol* 1995; **154**: 1265–1272.
- 31 Appel S, Reichwald K, Zimmermann W, Reis A, Rosenthal A, Hennies HC: Identification and localization of a novel human myotubularin-related protein gene, MTMR8, in 8p22-p23. *Genomics* 2001; **75**: 6–8.
- 32 Laporte J, Hu LJ, Kretz C *et al*: A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nat Genet* 1996; **13**: 175–182.
- 33 Bolino A, Muglia M, Conforti FL *et al*: Charcot-Marie-Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nat Genet* 2000; **25**: 17–19.
- 34 Lane MA, Wong SK, Daugherty K *et al*: Nucleotide sequence of a human oncogene active in tumors of secretory epithelium. *Nucleic Acids Res* 1990; **18**: 3068.
- 35 Mighell AJ, Smith NR, Robinson PA, Markham AF: Vertebrate pseudogenes. *FEBS Lett* 2000; **468**: 109–144.
- 36 Normark S, Bergström S, Edlund T *et al*: Overlapping genes. *Annu Rev Genet* 1983; **17**: 499–525.
- 37 Cooper PR, Smilnich NJ, Day CD *et al*: Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics* 1998; **49**: 38–51.
- 38 Jong MTC, Gray TA, Ji Y *et al*: A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. *Hum Mol Genet* 1999; **8**: 783–793.
- 39 Shintani S, O'hUigin C, Toyosawa S, Michalová V, Klein J: Origin of gene overlap: the case of *TCPI* and *ACAT2*. *Genetics* 1999; **152**: 743–754.
- 40 Tvrdik P, Asadi A, Kozak LP *et al*: *Cig30* and *Pitx3* genes are arranged in a partially overlapping tail-to-tail array resulting in complementary transcripts. *J Biol Chem* 1999; **274**: 26387–26392.
- 41 Ubeda M, Schmitt-Ney M, Ferrer J, Habener JF: CHOP/GADD153 and methionyl-tRNA synthase (MetRS) genes overlap in a conserved region that controls mRNA stability. *Biochem Biophys Res Commun* 1999; **262**: 31–38.
- 42 Campbell HD, Fountain S, Young IG, Weitz S, Lichter P, Hoheisel JD: Fliih, the murine homologue of the *Drosophila melanogaster* flightless I gene: nucleotide sequence, chromosomal mapping and overlap with *Llglh*. *DNA Seq* 2000; **11**: 29–40.
- 43 Dear TN, Meier NT, Hunn M, Boehm T: Gene structure, chromosomal localization, and expression pattern of *Capn12*, a new member of the calpain large subunit gene family. *Genomics* 2000; **68**: 152–160.
- 44 Morelli C, Magnanini C, Mungall AJ, Negrini M, Barbanti-Brodano G: Cloning and characterization of two overlapping genes in a subregion at 6q21 involved in replicative senescence and schizophrenia. *Gene* 2000; **252**: 217–225.
- 45 Yamazaki N, Yamanaka Y, Hashimoto Y, Shinohara Y, Shima A, Terada H: Structural features of the gene encoding human muscle type carnitine palmitoyltransferase I. *FEBS Lett* 1997; **409**: 401–406.
- 46 Yamazaki N, Shinohara Y, Kajimoto K, Shindo M, Terada H: Novel expression of equivocal messages containing both regions of choline/ethanolamine kinase and muscle type carnitine palmitoyltransferase I. *J Biol Chem* 2000; **275**: 31739–31746.