



CELLULAR & MOLECULAR BIOLOGY LETTERS

<http://www.cml.org.pl>

Received: 04 April 2008

Volume 14 (2009) pp 90-99

Final form accepted: 23 September 2008

DOI: 10.2478/s11658-008-0036-3

Published online: 06 October 2008

© 2008 by the University of Wroclaw, Poland

Short communication

**THE CARNITINE ACETYLTRANSFERASE GENE (CRAT):
A CHARACTERIZATION OF PORCINE TRANSCRIPTS WITH
INSIGHTS INTO THE 5'-END VARIANTS OF MAMMALIAN
TRANSCRIPTS AND THEIR POSSIBLE SUB-CELLULAR
LOCALIZATION**

ANNIE ROBIC*, THOMAS FARAUT, LAURENCE LIAUBET
and DENIS MILAN

Institut National de la Recherche Agronomique (INRA), Laboratoire de
Génétique Cellulaire, UMR 444, BP52627, 31326 Castanet Tolosan Cedex,
France

Abstract: Carnitine acetyltransferase (CRAT) is an important enzyme for energy homeostasis and fat metabolism. We characterized the predicted full length cDNA sequence of the porcine CRAT gene. Its structure is very similar to that in humans with respect to the size and organization of the 14 exons. We demonstrated the existence of a porcine alternative transcript resulting from a partial intron-retention at the 5' end of exon 2. To perform a comparison of the 5' end variants of the mammalian CRAT gene, we analyzed the Genbank data, and here we propose a new 5' variant for dog, rat and mouse. In contrast to other mammals where this variant encodes a shorter protein (-21 aa in human, mouse and rat, and -14 aa in dog), the pig variant encodes for a longer protein (+18 aa). In all mammalian species, variant 1 has a high probability of a preferential mitochondrial sub-cellular localization. Nevertheless, it is not evident, in particular in porcine and dog species, that the second variant is associated with a different sub-cellular specificity.

Key words: Pig, CRAT, mRNA, Alternative splicing, Mammals, Fat metabolism, Leader peptide

* Author for correspondence: e-mail: annie.robic@toulouse.inra.fr, tel.: +33 5 61 28 51 15, fax: +33 5 61 28 53 08

Abbreviations used: CRAT – carnitine acetyltransferase; HSA9 – human chromosome 9; EST – expression sequence tag; v.1 – variant n°1; aa – amino acid; RT-PCR – reverse transcription PCR

INTRODUCTION

Carnitine availability becomes a limiting factor for β -oxydation in certain physiological and pathological conditions [1, 2]. The carnitine acyltransferases are an important family of enzymes for energy homeostasis and fat metabolism through the modulation of the pools of acetyl-CoA and long-chain acyl-CoA in the sub-cellular compartments of animal cells [3]. This protein family is considered a promising target for the development of drugs against type II diabetes and other human diseases [1, 2]. Its members can be classified on the basis of their acyl-CoA selectivity. Five genes in the human genome are known to encode different active forms of related carnitine acyltransferase. CPT1A, CPT1B and CPT2 encode for palmitoyltransferases, CROT for octanoyltransferase, and CRAT for acetyltransferase [4]. Carnitine acetyltransferase is a key enzyme in the metabolic pathway in the mitochondria, peroxisomes and endoplasmic reticulum [1, 2]. The organelle membranes are impermeable to CoA, and CRAT operates as a buffering compartmental system by maintaining the appropriate levels of acetyl-CoA and CoA in the cellular compartments. Carnitine acetyl transferases are very active toward short chain acyl-CoAs but not toward medium- or long-chain acyl CoAs. In humans, the gene extends over 16 kb on HSA9q, and four transcripts have been described (<http://www.ensembl.org/index.html>). The best characterized transcript (RefSeq transcript NM_00755) results from the transcription of 14 exons and corresponds to variant 1, also depicted as the mitochondrial variant. Mitochondrial CRAT is expressed in the human liver, and high mRNA expression was detected in skeletal muscle. Physical exercise is known to upregulate mRNA synthesis for CRAT in muscle tissue [5].

Although the first studies of the sub-cellular distribution of the CRAT enzyme were performed on pigs [6] in 1973, this gene has not generated interest and further study with this species. It could be a key gene in the energetic metabolism of the pig, and thus could be significant for meat quality. Here, we present a detailed analysis of porcine and mammalian transcripts of the CRAT gene combining experimental and *in silico* approaches. We sequenced a porcine cDNA and, taking advantage of data available from Genbank, we demonstrated the existence of an alternative 5' end variant in porcine species. A comparison of the 5' end variants of the CRAT gene in mammals highlights the putative preferential sub-cellular location of CRAT proteins.

MATERIALS AND METHODS

All of the experiments were performed on samples of pig organs collected from animals weighing c. 100 kg at the time of slaughter. Pig muscle and backfat samples were collected, frozen immediately in liquid N₂ and stored at -80°C. The frozen samples were ground into powder before RNA extraction. Total RNA was treated by DNase and reverse transcribed (RT-PCR) with oligodT

primers using the M-MULV reverse transcriptase enzyme (New England Biolabs).

To sequence the cDNA of the CRAT porcine gene, PCR was performed on cDNA with primers chosen in exon 2 (CTCAAGCCCTCCTTGAC) and 13 (AGGTGTCCATGAAGATGTCG). A primer located in a new alternative first exon (CTGGGAAGTCAGGGCAGA) or in exon 1 (GGAGGGAGCGGAC AGTGTGT) associated with a primer located in exon 3 (GCGGTACTGGAG GTAAGCTG) was used to detect the 5' variants of porcine transcripts of the CRAT gene.

To sequence the PCR products, an aliquot (1 to 12 µl) was purified by a single treatment (45 min at 37°C followed by 30 min at 80°C) using 0.5 U of Shrimp Alkaline Phosphatase (Promega) and 0.8 U of exonuclease I (New England Biolabs). Sequencing was done on an 3730 ABI capillary DNA sequencer using a Big Dye terminator V3.1 cycle sequencing kit.

The ensembl web site was used to obtain known mammalian transcripts (<http://www.ensembl.org/>) of the CRAT gene. To study the main transcript of this gene in the porcine species, the *Iccare* software [7] was used to identify pig ESTs corresponding to this gene among all the available ESTs (<http://bioinfo.genopole-toulouse.prd.fr/iccare/>). The BLAST program (<http://www.ncbi.nlm.nih.gov/BLAST>) was used for sequence homology searches in the public databases. Multiple sequence (nucleic acid or protein) alignments were performed using the multalin program (<http://bioinfo.genopole-toulouse.prd.fr/multalin/>). The Lalign web tool (http://www.ch.embnet.org/software/LALIGN_form.html) allowed the determination of the percentage of identity between the two sequences.

The web tool EMBL-InterProScan (<http://www.ebi.ac.uk/InterProScan/>) was used to identify putative proteins. To propose a sub-cellular localization, we used TargetP [8] (<http://www.cbs.dtu.dk/services/TargetP/>) and Predotar [9] (<http://urgi.versailles.inra.fr/predotar/predotar.html>).

RESULTS AND DISCUSSION

To study the CRAT gene and its associated transcripts in porcine species, we identified an EST containing the first 3 exons (Genbank accession no DM134317). The assembly of three additional ESTs (BF198209, AW429639 and BF190277) allowed the reconstitution of the two last exons. Primers located in exon 2 and 13 were used to target a corresponding CRAT muscle transcript. A unique amplification product was obtained and sequenced. The genomic organization was determined by the alignment of the genomic sequence (BACs CH242 68C14 & 13F23, CU076102 & CU041302) with the cDNA sequence and by comparative analysis with human, mouse, rat and dog CRAT. A high level of conservation between the coding region sequences was found (90.0% and 84.9% identity respectively for human/pig and mouse/pig comparisons). The structure with 14 exons described in the genbank submission

(AM919499) (Fig. 1) is very similar to those described for other mammalian CRAT genes: ensembl gene IDs ENSG00000095232 (human), ENSCAFG00000019996 (dog), ENSMUSG00000028207 (mouse), ENSRNOG00000018145 (rat). This standard porcine transcript encodes for a protein containing 626 residues, which is similar to the other known mammalian transcripts with the exception of that for rat (634 aa). The porcine CRAT protein exhibits 91.4% identity with the human protein. The EMBL-InterProScan web tool confirmed the identity of the porcine protein as CRAT (data not shown).

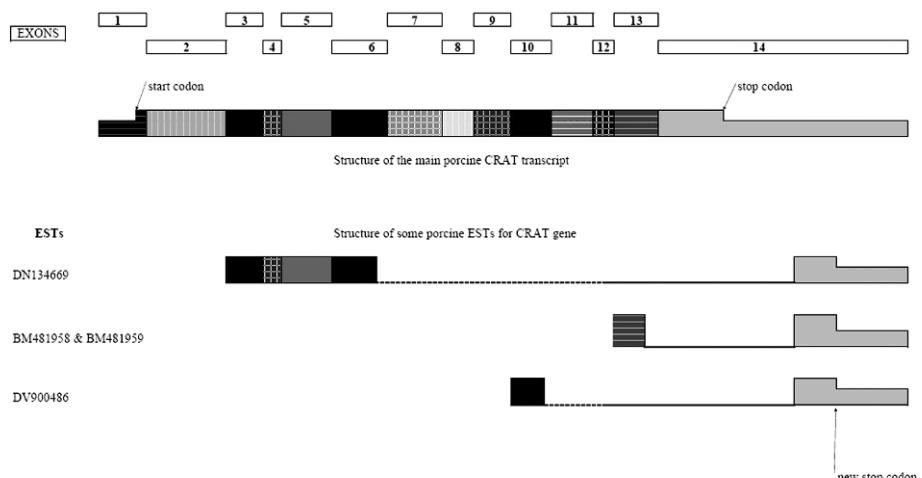


Fig. 1. The structure of the main transcript of the porcine CRAT gene and predicted alternative transcripts. 14 exons were included in the full cDNA sequence. Each exon was drawn by a box. Non-coding regions are shown with a lesser thickness. Four particular ESTs were found in the Genbank and their structures were schematized. Skipped sequences are shown with a dotted line.

A complete genome sequencing is in progress for the pig [10], and the sequence of this genomic porcine region is now available (BACs CH242 68C14 & 13F23, Genbank CU076102 & CU041302). The CRAT gene extends over 13.7 kb, and on average, the intronic sequences are smaller in the pig gene than in human. Examining these genomic sequences showed that all the sequences at the exon-intron junction are in consensus with the AG-GT consensus sequence (data not shown).

Alternative splicing is a very frequent phenomenon in the mammalian transcriptome, and it is now considered one of the main mechanisms responsible for increasing the diversity and complexity of the mammalian proteome [11]. ESTs are currently the most important resources to infer gene structure and alternative splicing in humans [12] or in porcine species [13]. A comparative analysis of the porcine ESTs found in the Genbank dbEST division (*Sus scrofa*,

1.33×10^6 ESTs) was performed. Some sequences of this division, reported as mRNA, were suspicious and therefore not considered in this study. For example, BE233424 contained only a part of exon 11 and some additional intronic sequences, and BF079421 contained sequences from exons 12, 13 and 14 but the first 272 bp were not included in the genomic DNA of this region (BACs CH242 68C14 & 13F23, Genbank CU076102 & CU041302). No EST is available to confirm the sequences of exons 6, 7 and 8, and this study allows the first characterization of porcine CRAT sequences corresponding to these exons. Among the porcine ESTs found for the CRAT gene, four could have resulted from alternative splicing and may encode a truncated protein. These porcine ESTs are schematized in Fig. 1 to highlight the involved sequences and characterize the predicted alternative transcripts. These three predicted transcripts lead to a frameshift and the apparition of a new termination codon in exon 14. This new coding phase and associated termination codon is shared by all four ESTs. A conserved alternative 3' splicing site in exon 14 (-446 bp) was also characterized. Similar minor alternative variants were reported for human and mouse (<http://fast-db.com/fastdb2/frame.html>) but the biological function of these minor transcripts is difficult to understand [14].

The alternative 5' end of transcripts for this gene were found for human and mouse, and are reported on the Ensembl web site. The transcript ENST00000318080 corresponding to variant 1 is the result of the transcription described in Fig. 2. The size of the intron separating exons 1 and 2 is 2405 bp. In the human transcript ENST00000372536 (variant 2), the alternative exon 1 is found in normal intron 1 and the beginning of the translation is reported to take place in exon 2 (Fig. 2). The resulting human protein includes only 605 residues. In the mouse alternative transcript ENSMUST00000102855, the first exon is split by a small intron without modification of the translation (Fig. 2, variant 1bis). The structure of the porcine transcript characterized in this study (Genbank AM919499) is described in Fig. 2. With a 2478-bp first intron (BAC CH242 13F23, Genbank CU041302) and a protein of 626 residues, this porcine transcript corresponds to the first variant. To characterize the second variant, porcine EST databases were screened in order to identify ESTs exhibiting an alternative structure upstream from exon 2. We found two ESTs in Genbank (BI347008 and BI343328), and EST cluster 1001437886 Ss1.1-Pig2-113N23.5 from the pig EST database of the Sino-Danish consortium [15] (10^6 ESTs are not included in Genbank, <http://pigest.ku.dk/cgi-bin/public.exe>). These correspond to this new variant, including a longer exon 2 and probably skipping exon 1 as shown in Fig. 2. To preserve the coding phase, a new starting codon is probably used, and the new predicted protein is longer, with 644 residues. Some experiments were performed to confirm the biological existence of these two transcript variants in porcine species. This new variant (v.2) and variant 1 were successfully screened by PCR (primers defined in the Materials and Methods section) in muscle and backfat cDNA (data not shown) from pigs. The expression

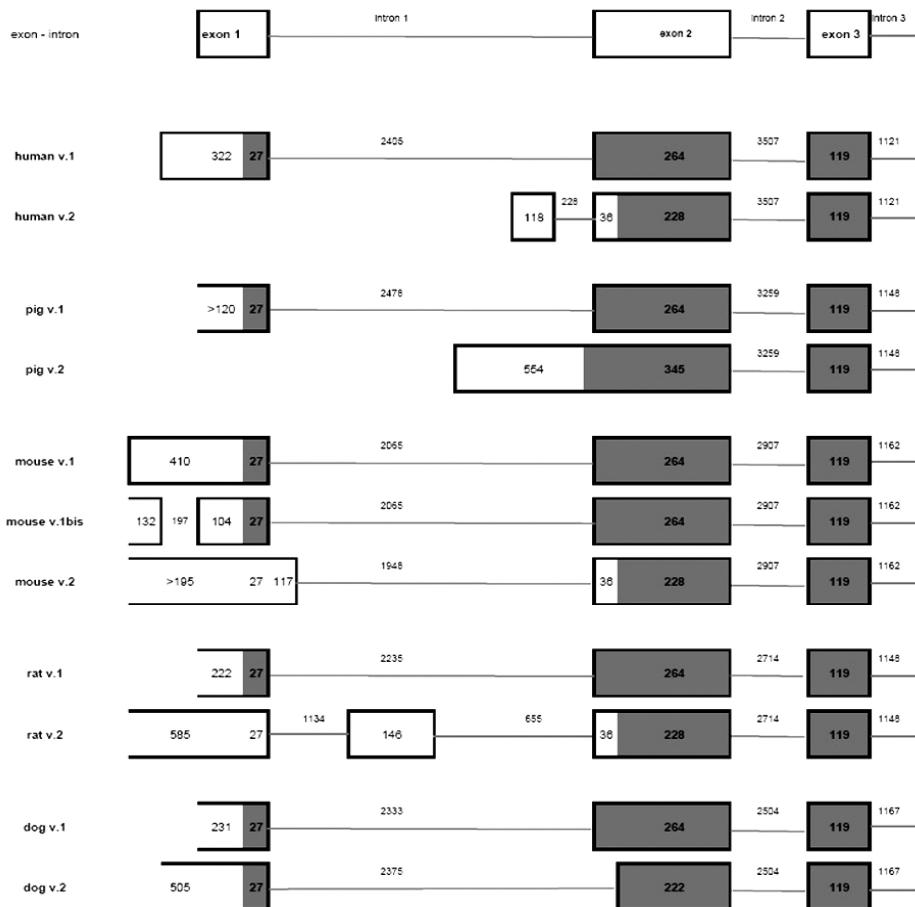


Fig. 2. A comparison of the structure and transcription of 5' end of the CRAT gene in mammals. Each transcription structure was represented by a line. Only the first three exons were drawn. The size of each exon is indicated in the box representing the exons. The grey boxes show encoding regions. 5' UTR are represented in white. The sequences used to characterize these variants are given in Tab. 1.

of this gene was detected more easily in backfat than in muscle. These two porcine variant transcripts originated from the single gene characterized here (screening of BACs library, data not shown), and we expected two CRAT proteins of 626 or 644 amino acids.

To perform a comparison of 5' end variants of the mammalian CRAT gene, we analyzed the Genbank data for mouse (4.85×10^6 ESTs), rat (0.84×10^6 ESTs), dog (0.36×10^6 ESTs) and cow (1.46×10^6 ESTs). We propose a new 5' variant

Tab. 1. The sequences used to characterize the 5' end variants of the CRAT gene in mammals.

	Variant 1	Variant 1bis	Variant 2
Human	ENST00000318080		ENST00000372536
Pig	AM919499		EST BI347008 & EST BI343328
Mouse	ENSMUST0000028207	ENSMUST00000102855	EST CB195405
Rat	ENSRNOT0000024774		EST CK469925 & EST EX489602
Dog	ENSCAFT0000031823 cDNA DN426391		cDNA DN427260

for mouse, rat and dog (Fig. 2 and Tab. 1). Unlike in other mammals, where this variant encodes a shorter protein (-21 aa in human, mouse, rat and -12 aa in dog), the pig variant encodes a longer protein.

Corti *et al.* [16] suggested that the different subcellular localizations of the CRAT mRNAs result from alternative splicing of the CRAT gene indicated by the divergent sequences in the 5' region. In human, the longer 626-aa form of the protein corresponding to the translation of the 14 exons was attributed to the mitochondrial form, whereas translation from a different start codon results in the shorter peroxisomal form (605 aa). This 626-aa human variant contained an amino acid sequence with the characteristics of a mitochondrial leader peptide [16]. In all the studied mammals, the form encoded by the 14-exon transcript harbored a leader peptide region which is very similar to the human one (Fig. 3).

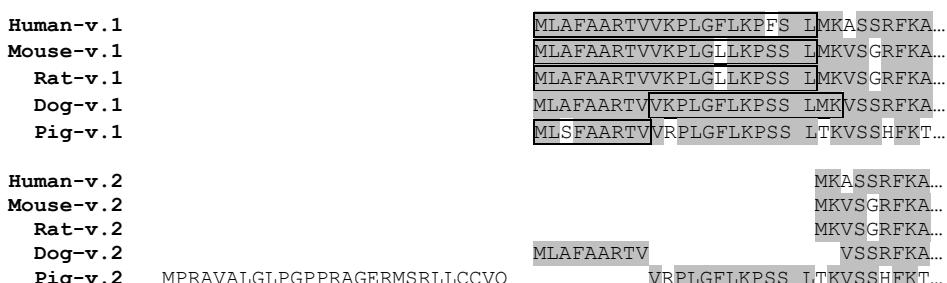


Fig. 3. A sequence comparison of the NH2-end of mammalian CRAT proteins. Homologies between these sequences have a grey background. The amino acids of variant 1 which were absent in the corresponding variant 2 are framed.

In human and rat, a new exon is defined in the first intron, and in mouse, the first exon is extended. When these new exonic sequences are involved in the transcript, translation starts in exon 2, and the first 21 amino acids are excluded from variant 2. In dog, there is no additional exon, and exon 2 is shorter (-42 bp). The second protein preserves the 9 amino acids issued from exon 1, but in total, is 14 aa shorter than variant 1. It is very surprising to observe that the NH2-end of variant 1 is identical in human and dog, as these species adopted divergent

ways to produce the second variant. It would be tempting to draw a conclusion on the importance of the exclusion of amino acids n°10 to n°21 (Fig. 3) in the second variant, but the examination of porcine transcripts is not concordant with this idea. Indeed, in porcine species, the transcription of the second variant is initiated upstream of exon 2, and the translation excludes the 9 first amino acids (encoded by exon 1) to replace them with 27 new residues. The amino acids located downstream of amino acid n°10 are preserved in variant 2 in porcine species, contrary to the situation in other mammals (Fig. 3).

Determining the sub-cellular localization of a protein is an important step toward understanding its function. As proposed by Emanuelsson *et al.* [17], we used two internet-accessible tools (TargetP and Predotar) to propose a sub-cellular prediction for mammalian CRAT proteins. As shown in Tab. 2, the results are very concordant for variant 1. This variant in all mammalian species has a high probability to have a preferential mitochondrial sub-cellular localization, and contains an amino acid sequence bearing the characteristics of a mitochondrial leader peptide. As shown in this study, the exclusion of 14 aa in the dog v.2 is not sufficient to lose the mitochondrial peptide. In porcine species, the situation is not as clear, but it is difficult to affirm that only variant 1 is mitochondrial. In pig and dog, while two protein variants are suspected to exist, it is not clear that they are associated with a different sub-cellular specificity.

Tab. 2. Analyses of the 5' end of the CRAT protein to predict its sub-cellular localization.

Sequence	Predotar v.1.03			TargetP v. 1.1		
	Mitoch.	Elsewhere	Prediction	Mitoch.	Elsewhere	Prediction
Mouse v.1	0.56	0.44	mitochondrial	0.767	0.199	mitochondrial
Human v.1	0.52	0.47	mitochondrial	0.816	0.173	mitochondrial
Pig v.1	0.58	0.41	mitochondrial	0.842	0.153	mitochondrial
Rat v.1	0.56	0.44	mitochondrial	0.773	0.210	mitochondrial
Dog v.1	0.57	0.43	mitochondrial	0.841	0.204	mitochondrial
Mouse v.2	0.16	0.78	none	0.463	0.476	
Human v.2	0.15	0.79	none	0.568	0.449	mitochondrial
Pig v.1	0.21	0.77	possibly mitochondrial	0.749	0.129	mitochondrial
Rat v.2	0.16	0.78	none	0.523	0.452	mitochondrial
Dog v.2	0.63	0.37	mitochondrial	0.855	0.114	mitochondrial

This study once again emphasises that the regions related to alternative splicing frequently code for a protein region involved in the function of the protein [18]. It is interesting to observe that various alternative splicing is used in five mammals to obtain a secondary CRAT protein. While it seems clear that the first variant is preferentially expressed in the mitochondria, the potential preferential sub-cellular expression of the second variant remains to be elucidated.

REFERENCES

1. Anderson, R.C. Carnitine palmitoyltransferase: a viable target for the treatment of NIDDM? **Curr. Pharm. Des.** 4 (1998) 1-16.
2. Wagman, A.S. and Nuss, J.M. Current therapies and emerging targets for the treatment of diabetes. **Curr. Pharm. Des.** 7 (2001) 417-450.
3. Jogl, G., Hsiao, Y.S. and Tong, L. Structure and function of carnitine acyltransferases. **Ann. NY Acad. Sci.** 1033 (2004) 17-29.
4. Van der Leij, F.R., Huijkman, N.C., Boomsma, C., Kuipers, J.R. and Bartelds, B. Genomics of the human carnitine acyltransferase genes. **Mol. Genet. Metab.** 71 (2000) 139-153.
5. Zeibig, J., Karlic, H., Lohninger, A., Damsgaard, R. and Smekal, G. Do blood cells mimic gene expression profile alterations known to occur in muscular adaptation to endurance training? **Eur. J. Appl. Physiol.** 95 (2005) 96-104.
6. Markwell, M.A., McGroarty, E.J., Bieber, L.L. and Tolbert, N.E. The subcellular distribution of carnitine acyltransferases in mammalian liver and kidney. A new peroxisomal enzyme. **J. Biol. Chem.** 248 (1973) 3426-3432.
7. Muller, C., Denis, M., Gentzbittel, L. and Faraut, T. The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species. **Nucleic Acids Res.** 32 (2004) W429-434.
8. Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. **J. Mol. Biol.** 300 (2000) 1005-1016.
9. Small, I., Peeters, N., Legeai, F. and Lurin, C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. **Proteomics** 4 (2004) 1581-1590.
10. Humphray, S.J., Scott, C.E., Clark, R., Marron, B., Bender, C., Camm, N., Davis, J., Jenks, A., Noon, A., Patel, M., Sehra, H., Yang, F., Rogatcheva, M.B., Milan, D., Chardon, P., Rohrer, G., Nonneman, D., de Jong, P., Meyers, S.N., Archibald, A., Beever, J.E., Schook, L.B. and Rogers, J. A high utility integrated map of the pig genome. **Genome Biol.** 8 (2007) R139.
11. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. Function of alternative splicing. **Gene** 344 (2005) 1-20.
12. Leipzig, J., Pevzner, P. and Heber, S. The Alternative Splicing Gallery (ASG): bridging the gap between genome and transcriptome. **Nucleic Acids Res.** 32 (2004) 3977-3983.
13. Zhang, L., Tao, L., Ye, L., He, L., Zhu, Y.Z., Zhu, Y.D. and Zhou, Y. Alternative splicing and expression profile analysis of expressed sequence tags in domestic pig. **Genomics Proteomics Bioinformatics** 5 (2007) 25-34.
14. Modrek, B. and Lee, C.J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. **Nat. Genet.** 34 (2003) 177-180.

15. Gorodkin, J., Cirera, S., Hedegaard, J., Gilchrist, M.J., Panitz, F., Jorgensen, C., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., Green, T., Nielsen, B.J., Havgaard, J.H., Rosenkilde, C., Wang, J., Li, H., Li, R., Liu, B., Hu, S., Dong, W., Li, W., Yu, J., Staefeldt, H.H., Wernersson, R., Madsen, L.B., Thomsen, B., Hornshoj, H., Bujie, Z., Wang, X., Bolund, L., Brunak, S., Yang, H., Bendixen, C. and Fredholm, M. Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. **Genome Biol.** 8 (2007) R45.
16. Corti, O., DiDonato, S. and Finocchiaro, G. Divergent sequences in the 5' region of cDNA suggest alternative splicing as a mechanism for the generation of carnitine acetyltransferases with different subcellular localizations. **Biochem. J.** 303 (Pt 1) (1994) 37-41.
17. Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. **Nat. Protoc.** 2 (2007) 953-971.
18. Yura, K., Shionyu, M., Hagino, K., Hijikata, A., Hirashima, Y., Nakahara, T., Eguchi, T., Shinoda, K., Yamaguchi, A., Takahashi, K., Itoh, T., Imanishi, T., Gojobori, T., and Go, M. Alternative splicing in human transcriptome: functional and structural influence on proteins. **Gene** 380 (2006) 63-71.