

Overlapping gene structure of human *VLCAD* and *DLG4*

Changcheng Zhou, Bruce Blumberg*

Department of Developmental and Cell Biology, University of California, Irvine, 5205 McLaugh Hall, Irvine, CA 92697-2300, USA

Received 19 August 2002; received in revised form 16 December 2002; accepted 20 December 2002

Received by A. Sippel

Abstract

Very-long-chain acyl-CoA dehydrogenase (*VLCAD*) is a major enzyme catalysing the first step in mitochondrial β -oxidation of long-chain fatty acids. During analysis of the *VLCAD* promoter, we discovered that another gene, discs-large-related 4 (*DLG4*), overlaps *VLCAD* and is transcribed in the opposite direction. *DLG4* encodes postsynaptic density-95 (PSD95) protein, which plays critical roles in the formation and maintenance of synaptic junctions. The transcription start site of the *VLCAD* gene was determined by primer extension analysis and the overlapping structure of *VLCAD* and *DLG4* was clarified. *VLCAD* and *DLG4* are arranged in a head-to-head orientation on chromosome 17p13, and share a 245 bp overlapping region that contains part of *DLG4* exon 1 and the entire exon 1 of *VLCAD* including 62 bp of protein coding sequence. Despite the overlap of their 5' ends, *DLG4* and *VLCAD* exhibit peak mRNA expression in different tissues, suggesting that they are independently regulated at the transcriptional level. Interestingly, *VLCAD* and *DLG4* genes do not overlap in the mouse or *Drosophila* genomes.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Overlapping gene; Quantitative real-time reverse transcription–polymerase chain reaction; Primer extension; Mitochondria β -oxidation

1. Introduction

Very-long-chain acyl-CoA dehydrogenase (*VLCAD*) is the major enzyme catalysing the first dehydrogenation step in the mitochondria β -oxidation of very long chain fatty acids (Strauss et al., 1995; Andresen et al., 1996). Four enzymes with different chain length specificities participate in the initial step in the mitochondria β -oxidation of straight-chain fatty acids. These are *VLCAD*, which requires straight-chain acyl-CoA substrates lengths of C₁₄–C₂₀; long-chain acyl-CoA dehydrogenase (*LCAD*); medium-chain acyl-CoA dehydrogenase (*MCAD*) and short-chain acyl-CoA dehydrogenase (*SCAD*). Four distinct dehydrogenase deficiencies have been described, with *VLCAD* deficiency being documented to cause hepatocellular dysfunction, severe lipid accumulation in organs, and

cardiac disease, particularly hypertrophic cardiomyopathy (Aoyama et al., 1995).

Human *VLCAD* cDNA and genomic sequences were reported (Orii et al., 1995; Strauss et al., 1995; Andresen et al., 1996). The *VLCAD* gene is about 5.4 kb long, contains 20 exons that encode a 655-amino-acid protein and is located on chromosome 17p13. *VLCAD* differs from the other three acyl-CoA dehydrogenases in that it is a homodimer of 70 kDa subunits bound to the inner membrane whereas the others are soluble enzymes found in the matrix. Since patients with *VLCAD* deficiency were first reported (Aoyama et al., 1993; Bertrand et al., 1993), more than 80 *VLCAD* mutations have been identified. These include nucleotide substitution, small and gross deletions and insertions, and small insertions/deletions. There is a distinct relationship between the severity of *VLCAD* deficiency and the nature of the mutations. Severely affected patients have mutations that result in complete loss-of-function, whereas patients with milder phenotypes have mutations that result in residual enzyme activity (Andresen et al., 1999). Little is currently known about the regulation of *VLCAD* transcription. In this report, we analysed the 5' region of *VLCAD* and discovered another gene, discs-large-related 4 (*DLG4*), overlapping *VLCAD*. *DLG4* encodes the postsynaptic density-95 (PSD95) protein

Abbreviations: *VLCAD*, very-long-chain acyl-CoA dehydrogenase; *DLG4*, discs-large-related 4; PSD95, postsynaptic density-95; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; AP-1, activator protein 1; AP-2, activation protein 2; Sp1, Specificity protein 1; NF- κ B, nuclear factor kappa-B; C/EBP, CCAAT/enhancer binding protein; kb, kilobase(s); bp, base pair(s); RT-PCR, reverse transcription–polymerase chain reaction; UTR, untranslated region.

* Corresponding author. Tel.: +1-949-824-8573; fax: +1-949-824-4709.

E-mail address: blumberg@uci.edu (B. Blumberg).

that is one of four known human proteins belonging to the discs-large-related (DLG) subfamily of the member of membrane-associated guanylate kinase (MAGUK) proteins (Gomperts, 1996). PSD95 interacts with both *N*-methyl-D-aspartate (NMDA) receptors and Shaker-type potassium channels and plays an important role in the formation and maintenance of synaptic junction (Stathakis et al., 1999). *DLG4* is positioned in a head-to-head orientation with the *VLCAD*, such that the 5'-untranslated region of *DLG4* overlaps with the 5'-UTR and a part of the protein-coding region of *VLCAD*. *VLCAD* and *DLG4* are both widely expressed in adult tissues but each has peak expression in different tissues. Thus, despite their shared 5' regions, *DLG4*

and *VLCAD* are likely to be differentially regulated at the level of transcription. Consistent with this model, the *VLCAD* and *DLG4* homologs are not overlapping in the mouse and *Drosophila* genomes.

2. Materials and methods

2.1. Primer extension analysis

Primer extension analysis was performed using the *VLCAD* specific oligonucleotide primer VL-66R (5'-TCCTGCACGCCACGTCCT-3'), located 66 bp upstream

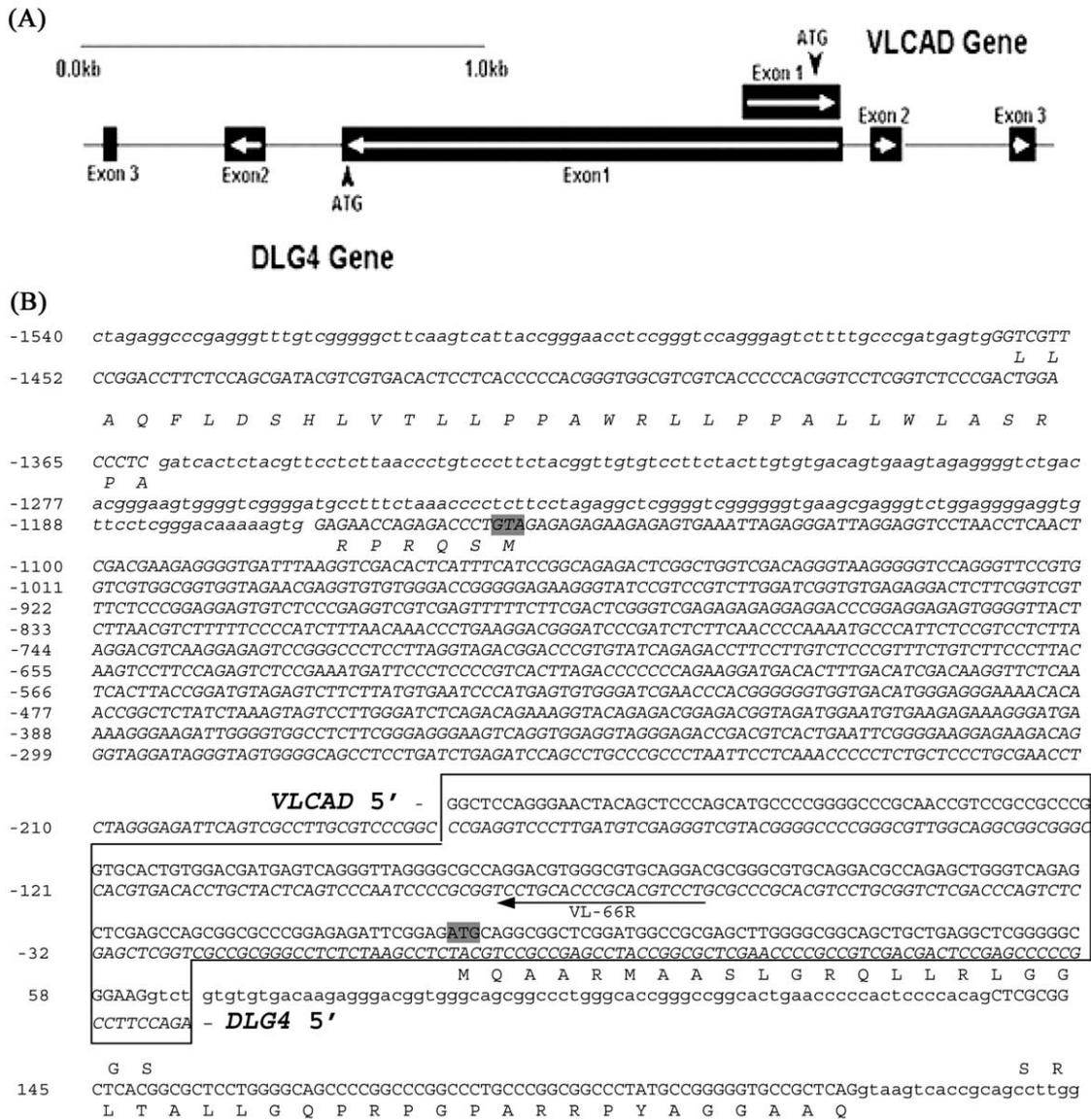


Fig. 1. (A) Schematic diagram of the genomic structure of human *VLCAD* and *DLG4*. The first three exons of *VLCAD* and *DLG4* are shown in black boxes and arrows indicate the direction of transcription. Inferred ATG translation start codons are also indicated. (B) Nucleotide sequence of 1.7 kb of human genomic DNA comprising the first two exons of *VLCAD* and *DLG4*. The sequence is numbered beginning with the first ATG, which is the translation start of *VLCAD*. ATG translation start codons are highlighted in gray. Nucleotides in the overlapping region are boxed. The two exons of *VLCAD* and *DLG4* are shown in uppercase and the amino acid sequence is shown below the exon nucleotides in one-letter code; introns are shown in lowercase. The oligonucleotide used for primer extension analysis is indicated by arrows positioned below.

of the inferred *VLCAD* ATG translation start codon. Total RNA was extracted from HeLa cells using the RNeasy kit (Qiagen) and primer extension performed using AMV reverse transcriptase (Primer Extension System; AMV Reverse Transcriptase Kit, Promega) according to the manufacturer's recommended protocols. Briefly, 5' end-labeling reactions were conducted in a 10- μ l volume containing 10 pmol VL-66R, 1 \times T4 polynucleotide kinase buffer, 30 μ Ci [γ - 32 P]ATP and 10 units T₄ polynucleotide kinase. Primer extension reactions were conducted in 20 μ l volume containing 10 μ g total RNA, 1 pmol 32 P-labeled primer, 2.8 mM sodium pyrophosphate, 1 \times AMV buffer and 2 units AMV reverse transcriptase. The transcription start site was mapped by electrophoresis on a 6% denaturing polyacrylamide gel and comparison with a DNA sequencing reaction from the *VLCAD* promoter using the same primer.

2.2. Quantitative real-time RT-PCR analysis

Human multiple tissue cDNA panels, which contain sets of first-strand cDNA, generated from different human tissues were commercially obtained (BD Clontech). Quantitative real time PCR was performed using the following primer sets: *VLCAD* (5'-GTCCTTTGCTGTGGGAATGT-3') (5'-CAAGTGGTCTCCTCCACCAT-3'), *DLG4* (5'-ACCAGATCCTGTTCGGTCAAC-3') (5'-TGCTGTTTCATGAGCTGTTCC-3'), or *GAPDH* (5'-GGCCTCCAAGGAGTAAGACC-3') (5'-AGGGGAGATTCAAGTGTGGTG-3'), with the SYBR green PCR kit (Applied Biosystems) in a DNA Engine Opticon – Continuous Fluorescence Detection System (MJ Research). All samples were analysed in duplicate and quantitated by the comparative Ct method for relative quantitation of gene expression, normalized to *GAPDH* (Livak and Schmittgen, 2001).

3. Results and discussion

3.1. Characterization of the 5' region and transcription start site of *VLCAD* gene

The *VLCAD* gene, located on chromosome 17p13, is about 5.4 kb long and consists of 20 exons and 19 introns (Orii et al., 1995). A BLAST search of the draft human genome using the *VLCAD* gene sequence (GenBank: SEG_D78279S) identified a chromosome 17-derived clone HRPC4G17 (GenBank: AC003688) that contains the entire *VLCAD* gene on its plus strand. This 98-kb PAC clone was sequenced by the Whitehead Institute/MIT center for Genome Research for the human chromosome 17 genome sequencing project. Further analysis of the upstream region of *VLCAD* gene identified another previously described gene, *DLG4* (Stathakis et al., 1997). The *DLG4* gene encompasses approximate 30 kb and is composed of 22 exons. The HRPC4G17 clone contains the first 10 exons of *DLG4* gene on its minus strand. Of particular interest is the

fact that the 5' end of *DLG4* gene overlaps with the *VLCAD* gene and its transcription start site extends to the coding region of *VLCAD* exon 1 (Fig. 1A).

The genomic structure and transcription start site of *DLG4* were reported (Stathakis et al., 1999). However, the transcription start site of *VLCAD* was unknown. We therefore conducted primer extension analysis of *VLCAD*. Total RNA from HeLa cells, which express the *VLCAD* mRNA (data not shown), was used to define the *VLCAD* transcription start site. Several groups have characterized the *VLCAD* cDNA. Only one major transcript was detected in a variety of human tissues and the mRNA size corresponds closely to the size of the *VLCAD* cDNA sequence (Aoyama et al., 1995; Orii et al., 1995; Strauss et al., 1995; Andresen et al., 1996). The primer used for primer extension analysis is located 66 bp upstream of *VLCAD* translation start site, within the *VLCAD* exon 1 and transcription region. Primer extension revealed a single, distinct 113 bp band (Fig. 2). This indicates that the major *VLCAD* transcription start site is located 179 nucleotides upstream from the inferred ATG-translation initiation codon

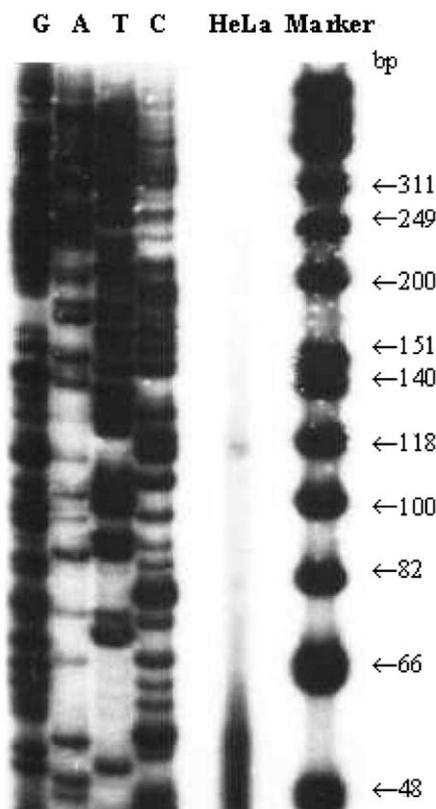


Fig. 2. Primer extension analysis of *VLCAD*. *VLCAD* specific primer VL-66R was hybridized to 10 μ g of total RNA from HeLa cells and extended with reverse transcriptase. The first four lanes represent a DNA sequence ladder generated by dideoxy chain termination reactions using VL-66R as primer and ddGTP(G), ddATP(A), ddTTP(T) or ddCTP(C). The fifth lane is the primer extension product. A 113 bp band represents the major primer extension product of *VLCAD* gene. The last lane is a DNA size marker (lane Marker).

of *VLCAD*. This position is within *DLG4* exon 1, 245 nucleotides downstream from its transcription start site (Fig. 1B).

After clarifying the transcription start site of the *VLCAD* gene, we found that *VLCAD* and *DLG4* share 245 nucleotides at their 5' ends, and are transcribed in opposite directions. The sequence of the overlapping region of *VLCAD* and *DLG4* genes was confirmed by sequencing of a 1 kb genomic PCR fragment obtained directly from human genomic DNA using a primer set that corresponds to both *VLCAD* and *DLG4* (data not shown).

The overlapping gene structure of *VLCAD* and *DLG4* is unusual in the human genome. The N-terminus of the *DLG4* gene product, the PSD95 protein, is modified by thioester-linked palmitate, a 16-carbon saturated fatty acid that targets PSD95 to cell membranes (Topinka and Bredt, 1998). PSD95 is a major palmitoylated protein in intact cells and palmitoylation is a critical regulatory mechanism for receptor interactions with PSD95. On the other hand, *VLCAD* is the major enzyme for the degradation of palmitate during its first step of β -oxidation in vivo. Therefore, the unusual structure of these two genes suggests a possible link between the use of palmitate as an energy substrate and the signal transduction pathway mediated by PSD95.

3.2. Tissue specific distribution of *VLCAD* and *DLG4*

Human multiple tissue cDNA panels, which contain sets of first-strand cDNA generated from different human tissues were used to investigate the tissue specific distribution of *VLCAD* and *DLG4* by quantitative real-time RT-PCR. Although the overlapping region of *VLCAD* and *DLG4* covers the entire *VLCAD* exon 1 including 62 bp of coding region, transcription of these two genes is not impaired by their overlap. Both genes can be co-expressed in the same tissues, but exhibit different patterns of maximal expression (Fig. 3). *DLG4* mRNA is most abundant in brain with prominent expression detected in heart, placenta, lung, pancreas, spleen, thymus, testis, ovary and small intestine. In contrast, *VLCAD* is highly expressed in heart and skeletal muscle with lower levels detectable in placenta and pancreas. Both genes were expressed at lower, but detectable levels in the other tissues examined. We infer from the expression of their mRNAs that these two genes share common regulatory elements but also have distinct tissue-specific elements required for high-level expression.

Computer analysis of the upstream regions of *VLCAD* and *DLG4* genes including their overlapping region was performed using promoter scan II (<http://biosci.cbs.umn.edu/software/software.html>) and MatInspector 2.2 ([| Tissue | <i>VLCAD</i> RNA expression | <i>DLG4</i> RNA expression |
|----------------------------|-----------------------------|----------------------------|
| Heart | ~140 | ~40 |
| Brain | ~5 | ~200 |
| Placenta | ~30 | ~35 |
| Lung | ~15 | ~65 |
| Liver | ~10 | ~5 |
| Skeletal Muscle | ~125 | ~40 |
| Kidney | ~10 | ~15 |
| Pancreas | ~50 | ~35 |
| Spleen | ~15 | ~85 |
| Thymus | ~5 | ~45 |
| Prostate | ~5 | ~15 |
| Testis | ~10 | ~100 |
| Ovary | ~5 | ~60 |
| Small Intestine | ~10 | ~35 |
| Colon | ~5 | ~20 |
| Peripheral Blood Leukocyte | ~5 | ~30 |](http://</p>
</div>
<div data-bbox=)

Fig. 3. Tissue-specific expression of *VLCAD* and *DLG4*. Real time RT-PCR was used to determine which human adult tissues expressed *VLCAD* and *DLG4*. Data are shown in the form of RNA expression levels normalized to *GAPDH* controls. Values represent the average of duplicates \pm range.

transfac.gbf.de/TRANSFAC). This analysis identified two potential TATA-less promoters and potential binding sites for several transcription factors including Sp1, AP-1, AP-2, C/EBP and NF- κ B. The computer prediction of *VLCAD* gene's promoter region and AP-2 binding site are consistent with a published report (Zhou et al., 1999). Further studies to characterize transcription factors essential for the correct regulation of both these promoters will provide insight into the mechanism governing the regulation of both genes at the transcriptional level.

3.3. Evolutionary analysis of *VLCAD* and *DLG4* gene loci

Overlapping genes occur frequently in prokaryotic, viral and mitochondrial genomes but only rarely in most eukaryotic genomes, including the human genome (Shintani et al., 1999). Since the first two overlapping genes reported by Emi et al. (Emi et al., 1988), a small number of human overlapping genes have been identified (Ubeda et al., 1999; Faurholm et al., 2001; Dan et al., 2002). Although a few cases have been reported, anti-parallel overlapping of exons of functional protein coding genes is quite rare (Nicolaidis et al., 1995; Herzog et al., 1997; Cooper et al., 1998).

To investigate the evolution of this overlapping organization, we examined the structure of *VLCAD* and *DLG4* in the draft *Drosophila* and mouse genomes. The *Drosophila*-gene *CG7461* (FlyBase ID: FBgn0034432) that encodes *VLCAD* is located on chromosome 2R 56C9-56C10. *Discs large 1* (FlyBase ID: FBgn0001624), the fly ortholog of the human *DLG* family is located on the X chromosome at locus 10B11–10B12. A BLAST search of the mouse genome with *VLCAD* (XM_126408) and *DLG4* (NM_007864) revealed that these genes are both located on chromosome 6 at 11B3–B4 in a head-to-head orientation. In contrast to the human genes, the mouse genes do not overlap and contain a gap of 3.48 kb between them. This suggests that the *VLCAD* and *DLG4* genes were originally separated but have become co-localized during the course of evolution. Further analysis of the genomic structures of these genes in other organisms (as the genomic sequences become available) will allow one to deduce when this co-localization event occurred during the course of vertebrate evolution.

3.4. Conclusion

We identified two overlapping genes *VLCAD* and *DLG4* in the human genome. The transcription start site of *VLCAD* was determined and it was shown that *VLCAD* and *DLG4* are organized in a head-to-head orientation sharing a 245 bp overlapping region that encompasses part of *DLG4* exon 1 and the entire first exon of *VLCAD* including 62 bp of protein-coding region. Both genes were expressed in all adult tissues examined although sites of peak expression were different. *VLCAD* and *DLG4* exonic sequences do not overlap in the *Drosophila* or mouse genomes, suggesting

that the loss of intergenic sequences happened late in mammalian evolution.

References

- Andresen, B.S., Bross, P., Vianey-Saban, C., Divry, P., Zabet, M.T., Roe, C.R., Nada, M.A., Byskov, A., Kruse, T.A., Neve, S., Kristiansen, K., Knudsen, I., Corydon, M.J., Gregersen, N., 1996. Cloning and characterization of human very-long-chain acyl-CoA dehydrogenase cDNA, chromosomal assignment of the gene and identification in four patients of nine different mutations within the *VLCAD* gene. *Hum. Mol. Genet.* 5, 461–472.
- Andresen, B.S., Olpin, S., Poorthuis, B.J., Scholte, H.R., Vianey-Saban, C., Wanders, R., Ijlst, L., Morris, A., Pourfarzam, M., Bartlett, K., Baumgartner, E.R., deKlerk, J.B., Schroeder, L.D., Corydon, T.J., Lund, H., Winter, V., Bross, P., Bolund, L., Gregersen, N., 1999. Clear correlation of genotype with disease phenotype in very-long-chain acyl-CoA dehydrogenase deficiency. *Am. J. Hum. Genet.* 64, 479–494.
- Aoyama, T., Uchida, Y., Kelley, R.I., Marble, M., Hofman, K., Tongsgard, J.H., Rhead, W.J., Hashimoto, T., 1993. A novel disease with deficiency of mitochondrial very-long-chain acyl-CoA dehydrogenase. *Biochem. Biophys. Res. Commun.* 191, 1369–1372.
- Aoyama, T., Souri, M., Ushikubo, S., Kamijo, T., Yamaguchi, S., Kelley, R.I., Rhead, W.J., Uetake, K., Tanaka, K., Hashimoto, T., 1995. Purification of human very-long-chain acyl-coenzyme A dehydrogenase and characterization of its deficiency in seven patients. *J. Clin. Invest.* 95, 2465–2473.
- Bertrand, C., Largilliere, C., Zabet, M.T., Mathieu, M., Vianey-Saban, C., 1993. Very long chain acyl-CoA dehydrogenase deficiency: identification of a new inborn error of mitochondrial fatty acid oxidation in fibroblasts. *Biochim. Biophys. Acta* 1180, 327–329.
- Cooper, P.R., Smilnich, N.J., Day, C.D., Nowak, N.J., Reid, L.H., Pearsall, R.S., Reece, M., Prawitt, D., Landers, J., Housman, D.E., Winterpacht, A., Zabel, B.U., Pelletier, J., Weissman, B.E., Shows, T.B., Higgins, M.J., 1998. Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics* 49, 38–51.
- Dan, I., Watanabe, N.M., Kajikawa, E., Ishida, T., Pandey, A., Kusumi, A., 2002. Overlapping of *MINK* and *CHRNE* gene loci in the course of mammalian evolution. *Nucleic Acids Res.* 30, 2906–2910.
- Emi, M., Horii, A., Tomita, N., Nishide, T., Ogawa, M., Mori, T., Matsubara, K., 1988. Overlapping two genes in human DNA: a salivary amylase gene overlaps with a gamma-actin pseudogene that carries an integrated human endogenous retroviral DNA. *Gene* 62, 229–235.
- Faurholm, B., Millar, R.P., Katz, A.A., 2001. The genes encoding the type II gonadotropin-releasing hormone receptor and the ribonucleoprotein RBM8A in humans overlap in two genomic loci. *Genomics* 78, 15–18.
- Gomperts, S.N., 1996. Clustering membrane proteins: It's all coming together with the PSD-95/SAP90 protein family. *Cell* 84, 659–662.
- Herzog, H., Darby, K., Ball, H., Hort, Y., Beck-Sickinger, A., Shine, J., 1997. Overlapping gene structure of the human neuropeptide Y receptor subtypes Y1 and Y5 suggests coordinate transcriptional regulation. *Genomics* 41, 315–319.
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408.
- Nicolaidis, N.C., Kinzler, K.W., Vogelstein, B., 1995. Analysis of the 5' region of *PMS2* reveals heterogeneous transcripts and a novel overlapping gene. *Genomics* 29, 329–334.
- Orii, K.O., Aoyama, T., Souri, M., Orii, K.E., Kondo, N., Orii, T., Hashimoto, T., 1995. Genomic DNA organization of human mitochondrial very-long-chain acyl-CoA dehydrogenase and mutation analysis. *Biochem. Biophys. Res. Commun.* 217, 987–992.
- Shintani, S., O'Huigin, C., Toyosawa, S., Michalova, V., Klein, J., 1999.

- Origin of gene overlap: the case of TCP1 and ACAT2. *Genetics* 152, 743–754.
- Stathakis, D.G., Hoover, K.B., You, Z., Bryant, P.J., 1997. Human postsynaptic density-95 (PSD95): location of the gene (DLG4) and possible function in nonneural as well as in neural tissues. *Genomics* 44, 71–82.
- Stathakis, D.G., Udar, N., Sandgren, O., Andreasson, S., Bryant, P.J., Small, K., Forsman-Semb, K., 1999. Genomic organization of human DLG4, the gene encoding postsynaptic density 95. *J. Neurochem.* 73, 2250–2265.
- Strauss, A.W., Powell, C.K., Hale, D.E., Anderson, M.M., Ahuja, A., Brackett, J.C., Sims, H.F., 1995. Molecular basis of human mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency causing cardiomyopathy and sudden death in childhood. *Proc. Natl. Acad. Sci. USA* 92, 10496–10500.
- Topinka, J.R., Bredt, D.S., 1998. N-terminal palmitoylation of PSD-95 regulates association with cell membranes and interaction with K⁺ channel Kv1.4. *Neuron* 20, 125–134.
- Ubeda, M., Schmitt-Ney, M., Ferrer, J., Habener, J.F., 1999. CHOP/GADD153 and methionyl-tRNA synthetase (MetRS) genes overlap in a conserved region that controls mRNA stability. *Biochem. Biophys. Res. Commun.* 262, 31–38.
- Zhou, Y., Kelly, D.P., Strauss, A.W., Sims, H., Zhang, Z., 1999. Characterization of the human very-long-chain acyl-CoA dehydrogenase gene promoter region: a role for activator protein 2. *Mol. Genet. Metab.* 68, 481–487.