

Organization and chromosomal localization of the human and mouse genes coding for LanC-like protein 1 (LANCL1)

H. Mayer, H. Bauer, and R. Prohaska

Institute of Medical Biochemistry, University of Vienna, Vienna Biocenter, Vienna (Austria)

Abstract. We have determined the organization and chromosome location of the human LANCL1 and mouse *Lancl1* genes encoding LANCL1, the lanthionine synthetase component C (LanC)-like protein 1. LANCL1 is related to the bacterial LanC family which is involved in the biosynthesis of antimicrobial peptides. The human and mouse genes span 45 kb and 38 kb, respectively, each comprising ten exons. Within the potential promoter regions, several consensus sequences for

ubiquitous and tissue-specific transcription factors are present, reflecting the expression data. The nucleotide sequence of the previously unknown mouse full-length transcript is also reported here. Fluorescence in situ hybridization analyses assigned the LANCL1 gene to human chromosome 2q34 and the *Lancl1* gene to mouse chromosome 1, region C2–C5, in accordance with the known homology.

Copyright © 2001 S. Karger AG, Basel

The 40-kDa protein LANCL1 (alias p40/GPR69A) was originally isolated from human erythrocyte membranes (Mayer et al., 1998a) by using its affinity for the C-terminus of the membrane protein stomatin (band 7.2b). It is mainly expressed in the brain, testis, various epithelia, and diverse hematopoietic cells (Mayer et al., 1998a; 1998b). Sequence analyses showed that human and mouse LANCL1 contain seven hydrophobic stretches with predicted α -helical transmembrane structure (Mayer et al., 1998a; 1998b), thereby suggesting a G protein-coupled receptor structure and function. However, recent data revealed that LANCL1 is a peripheral membrane protein similar to the bacterial lanthionine synthetase C (LanC) components (Bauer et al., 2000), which are part of a multimeric membrane-associated complex involved in the modification and transport of peptides (Sieggers et al., 1996; Kiesau et al., 1997).

The produced peptides, termed lantibiotics (Schnell et al., 1988), have antimicrobial properties (Sahl and Bierbaum, 1998). Functionally similar defense peptides exist in mammals and other vertebrates, insects, and plants (Ganz and Lehrer, 1998; Borregaard et al., 2000). Homologues of LANCL1 are present in *Drosophila melanogaster* and *Arabidopsis thaliana*, indicating that this protein plays a fundamental role in animals and plants. In the present study, we have determined and compared the organization and chromosome location of the human LANCL1 and mouse *Lancl1* genes (symbols approved by the Human and Mouse Gene Nomenclature Committees).

Materials and methods

Genomic library screening and PCR

In order to obtain genomic clones harboring the human LANCL1 gene, P1 and PAC libraries have been screened by a PCR-based approach (Genome Systems, Inc.) using pairs of gene-specific primers which were designed according to the human LANCL1 cDNA (GenBank accession number Y11395). In each of the libraries, one positive clone was identified (P1, clone address: 11 38 B8; PAC, clone address: 223 G3). Sequence analyses revealed that neither of them comprised the 5'-part of the gene. The P1 clone contained exons 6–10, and the PAC clone exons 4–10 (Fig. 1A). Two different approaches elucidated the remaining 5'-region of the human LANCL1 gene. Several cDNA-derived primer pairs were tested in PCR reactions on total human genomic DNA (Clontech) and the products were sequenced. This method revealed the remaining exon-intron boundaries. Finally, the 5'-flank-

Supported by the Austrian Science Fund (FWF) Grant No. P12743.

Received 1 December 2000; manuscript accepted 2 March 2001.

Request reprints from Dr. Rainer Prohaska,

Institute of Medical Biochemistry, University of Vienna, Vienna Biocenter,
Dr. Bohr-Gasse 9/3, A-1030 Vienna (Austria);
telephone: +43-1-4277-61660; fax: +43-1-4277-9616;
email: prohaska@bch.univie.ac.at

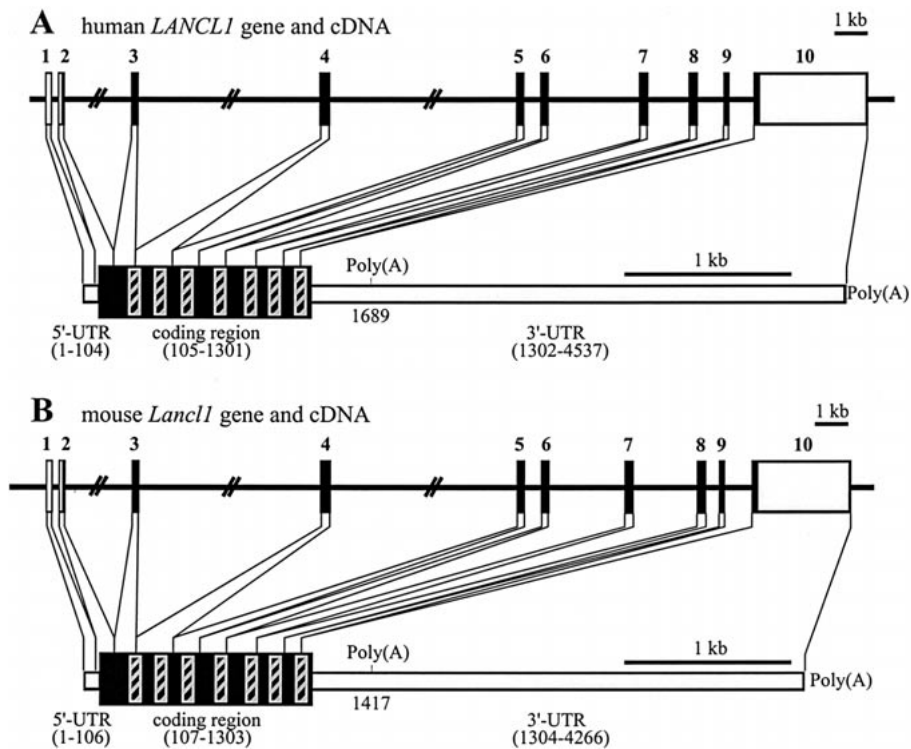


Fig. 1. Schematic model of the human LANCL1 (A) and mouse *Lancl1* (B) genes and respective cDNAs. Exons are numbered and represented by boxes. Filled boxes represent coding sequences and open boxes represent untranslated regions. Introns which are not drawn to scale are marked by bold double lines. In the cDNAs, the coding regions are depicted by filled boxes with the positions of the seven hydrophobic repeats shown as hatched rectangles. Two alternative polyadenylation sites are indicated. Numbering of nucleotides is from 5' to 3'.

ing region containing the potential promoter sequence was amplified from the *EcoRV* library of the human GenomeWalker™ Kit (Clontech) by the use of appropriate gene-specific primers and the adaptor primers provided. The organization of the murine *Lancl1* gene was elucidated solely by a series of PCR reactions using the libraries of the Mouse GenomeWalker™ Kit (Clontech). Pairs of gene-specific primers were designed according to the murine *Lancl1* cDNA (GenBank accession numbers Y11550 and Y16518) and employed to reveal the exon-intron boundaries and the sizes of the introns. PCR using gene-specific and adaptor primers yielded fragments containing the 5'- and 3'-flanking regions of the gene.

5'- and 3'-rapid amplification of cDNA ends (RACE)

Based on the cDNA sequence of murine LANCL1 (GenBank accession numbers Y11550 and Y16518), primers were designed to amplify the missing 5'-end and the long 3'-UTR which is transcribed due to the alternative usage of a second polyadenylation signal. In order to generate the 5'-sequence we applied CapSelect, a method which selectively enriches full-length cDNA, as described (Schmidt and Mueller, 1999). We used high-quality poly A⁺ RNA from mouse testis (Clontech) as the template for reverse transcription. Specific PCR bands were cloned into the pGEM®-T system (Promega) and sequenced using standard sequencing primers. The 3'-RACE was performed according to the manufacturer's protocol by using a Marathon-Ready™ cDNA library from murine brain (Clontech) and the Advantage™ cDNA PCR Kit (Clontech). Specific PCR bands were cloned into the pCR®II-TOPO vector (Invitrogen) and subjected to sequence analysis.

Fluorescence in situ hybridization (FISH)

The LANCL1-positive human P1 clone described above was used for FISH analysis (Genome Systems, Inc.). DNA from this clone was labeled with digoxigenin-dUTP and hybridized to normal metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes. Specific hybridization signals were detected by incubating the hybridized slides with fluoresceinated anti-digoxigenin antibody followed by counterstaining with DAPI. In a second experiment, a biotin-labeled probe specific for the centromere of chromosome 2 was cohybridized with the P1 clone. Probe detection was accomplished by incubating the slides with fluoresceinated anti-digoxigenin antibody and Texas red-avidin followed by counterstaining with

DAPI. A 5-kb genomic PCR product from murine *Lancl1* was used as a probe for FISH analysis (SeeDNA, Inc.). DNA from the respective clone was labeled with biotinylated dATP and FISH analysis was performed as described (Heng et al., 1992; Heng and Tsui, 1993).

Results and discussion

Genomic structure of human LANCL1 and mouse *Lancl1*

Both the genes for human LANCL1 (GenBank accession numbers AJ289236 to AJ289239) and mouse *Lancl1* (GenBank accession numbers AJ289603 to AJ289608) comprise ten individual exons (Fig. 1A, B). The sizes of exons 1–9 vary between 73 bp (exon 9) and 208 bp (exon 4) (Table 1). In contrast, exon 10 is considerably larger (3,313 bp in man; 3,037 bp in mouse), and contains the long 3'-UTRs of the corresponding cDNAs. The overall length of the two genes is about 45 kb (human) and 38 kb (mouse). In the human LANCL1 gene, most introns have been sequenced except introns 2, 3, and 4 which are larger than 4 kb each. From these introns, only the 5'- and 3'- junctions with the flanking exons have been analyzed. The introns 2, 3, 4, 6, and 7 of the murine gene have been analyzed in a similar manner. All exon-intron boundaries are absolutely conserved between the two species. All of the splice acceptor and donor sequences agree with the general GT/AG rule (Table 1) and all introns show pyrimidine-rich stretches preceding the 3' splice sites. Only three of the introns interrupt codons (I3, I4, I9). There is one intron (I1) within the 5'-UTR of the human and mouse genes. This intron is unspliced in the majority of the mouse cDNAs analyzed (Mayer et al., 1998b), but not in the human cDNAs. Supposing that this intron may have a func-

Table 1. Exon/intron organization of the human LANCL1 and mouse *Lancl1* genes^a

Human LANCL1				Mouse <i>Lancl1</i>					
E	5' splice site	I	3' splice site	E	E	5' splice site	I	3' splice site	E
1	CTTTATCCGGgtggggttaag	1	tgttctgcagGCTTGCTTCC	2	1	GGCGTGGGACgtgggtgaac	1	tgttctgcagGTTTCGCTTGC	2
2	TGCCGGGAGGgtgaggttag A G R	2	taatatctagCTGACTCCTG L T P E	3	2	TACTGGGAGGgtgaggttag T G R	2	atttctgcagCTGACTCCTG L T P E	3
3	GGCTGGGCAGgtatgaagtg G W A G	3	tttttttcagGTATTGCTGT I A V	4	3	GGCTGGGCAGgtacagtgc G W A G	3	tacattccagGCATCGCTGT I A V	4
4	GCATCACACGgtaatcgtgt I T R	4	ttatttttagGCTAATTCAC L I H	5	4	GCATCACACGgtaaccatgg I T R	4	ctattcctagGCTCATTAC L I H	5
5	TATTCAGCAGgtaccacgtt I Q Q	5	ctccctgtagATTTGTGAAA I C E T	6	5	TATTCAGCAGgtaatacagt I Q Q	5	ctccctgtagATTTGTGAAA I C E N	6
6	CCTGATGCAGgtaaggagtg L M Q	6	tcatttttagCCCAGCCTTC P S L Q	7	6	CCTGATGCAGgtaaggagtg L M Q	6	tcatttttagCCCAGCCTTC P S L Q	7
7	GGCCTATAAGgtactgtgtg A Y K	7	gcctctgcagGTATTTCAGAG V F R E	8	7	AGCATACAAGgtacctcatg A Y K	7	gactctgcagGTGTTCAAAG V F K E	8
8	GGCCTGTAAGgtaggagtca A C K	8	tcatttttagTTTGCTGAAT F A E W	9	8	GGCCTGCAAGgtaagagcca A C K	8	tcatttttagTTTGCTGAGT F A E W	9
9	CTCTTTGAAGgtatttgtac L F E G	9	ctttcttcagGAATGGCTGG M A G	10	9	CTCTTTGAAGgtatttctca L F E G	9	gttcctctagGGATGGCTGG M A G	10
E	Exon length (bp)	I	Intron length (bp) ^b	E	E	Exon length (bp)	I	Intron length (bp) ^b	
1	88	1	208	1	90	1	184		
2	97	2	~4300	2	97	2	~4400		
3	118	3	~16000	3	118	3	~12000		
4	208	4	~14000	4	208	4	~11000		
5	136	5	563	5	136	5	533		
6	147	6	2724	6	147	6	~2300		
7	183	7	1296	7	183	7	~1900		
8	177	8	755	8	177	8	378		
9	73	9	822	9	73	9	781		
10	3313			10	3037				

^a Depicted are exon-intron junctions (caps: exon; small caps: intron), amino acids (aligned at the first position of the codons), exon and intron lengths. E, exon number; I, intron number.

^b "~" denotes approximate intron sizes determined by PCR.

tional relevance for the regulation of LANCL1 expression (Mayer et al., 1998b), we may speculate about different respective mechanisms in the two species. Interestingly, most of the seven hydrophobic stretches containing the motif G-X-X-G (Bauer et al., 2000) are encoded by different exons except for the first one, which is split in exons 3 and 4 and the last one, split in exons 9 and 10 (Fig. 1A, B). Recently, a human BAC clone (GenBank accession number AC007970) was deposited in the databases covering the region of the human LANCL1 gene (annotated by corresponding EST clones). Comparison of our sequences to this clone revealed several single nucleotide exchanges, restricted to intron sequences.

The 5' flanking regions of the human and murine LANCL1 genes have been analyzed for putative transcription factor binding sites by the use of the program MatInspector V2.2 (Quandt et al., 1995). Although TATA boxes are found in both species their functional relevance is not clear, because they are not located at appropriate distances. Nevertheless, there are classical CCAAT boxes in both genes which are flanked by two Sp1-binding sites. Further ubiquitous transcriptional activators found in both genes include USF, Oct-1 and AP-1. Six GATA-1 binding sites (Fujiwara et al., 1996) may explain the strong expression of LANCL1 in erythrocytes and megakaryocytes (Mayer et al., 1998a). A number of transcription factor binding

sites hint at an immune response-related function of LANCL1, as hypothesized (Mayer et al., 1998a). These include the interferon regulatory factor IRF-1 (Harroch et al., 1994), C/EBPbeta (Akira et al., 1990), Ik-1 and Ik-2 (Georgopoulos et al., 1994), and NF-kappaB (Perkins, 1997). The latter is found only in the human LANCL1 gene. In both species, there are multiple sites for the transcription factor SRY, which is the master regulator of male sex determination (Sinclair et al., 1990). This factor may be involved in the strong expression of LANCL1 in human and mouse testis (Mayer et al., 1998a; 1998b). In accordance with the observed expression of LANCL1 throughout the central nervous system (Mayer et al., 1998a), we found respective binding sites for Tst-1 (Verrijzer et al., 1991) and CREB (Ber-kowitz and Gilman, 1990). The latter was present only in the murine but not the human 5'-flanking region.

Mouse LANCL1 full-length cDNA

In man, two different transcripts have been described according to the usage of two different polyadenylation signals. The long transcript is the major one as can be seen in human Northern blot analyses (Mayer et al., 1998a). In mouse, only the sequence of the short transcript has been determined so far (Mayer et al., 1998b). Here we report the sequence of the long transcript (GenBank accession number AJ294535) originating

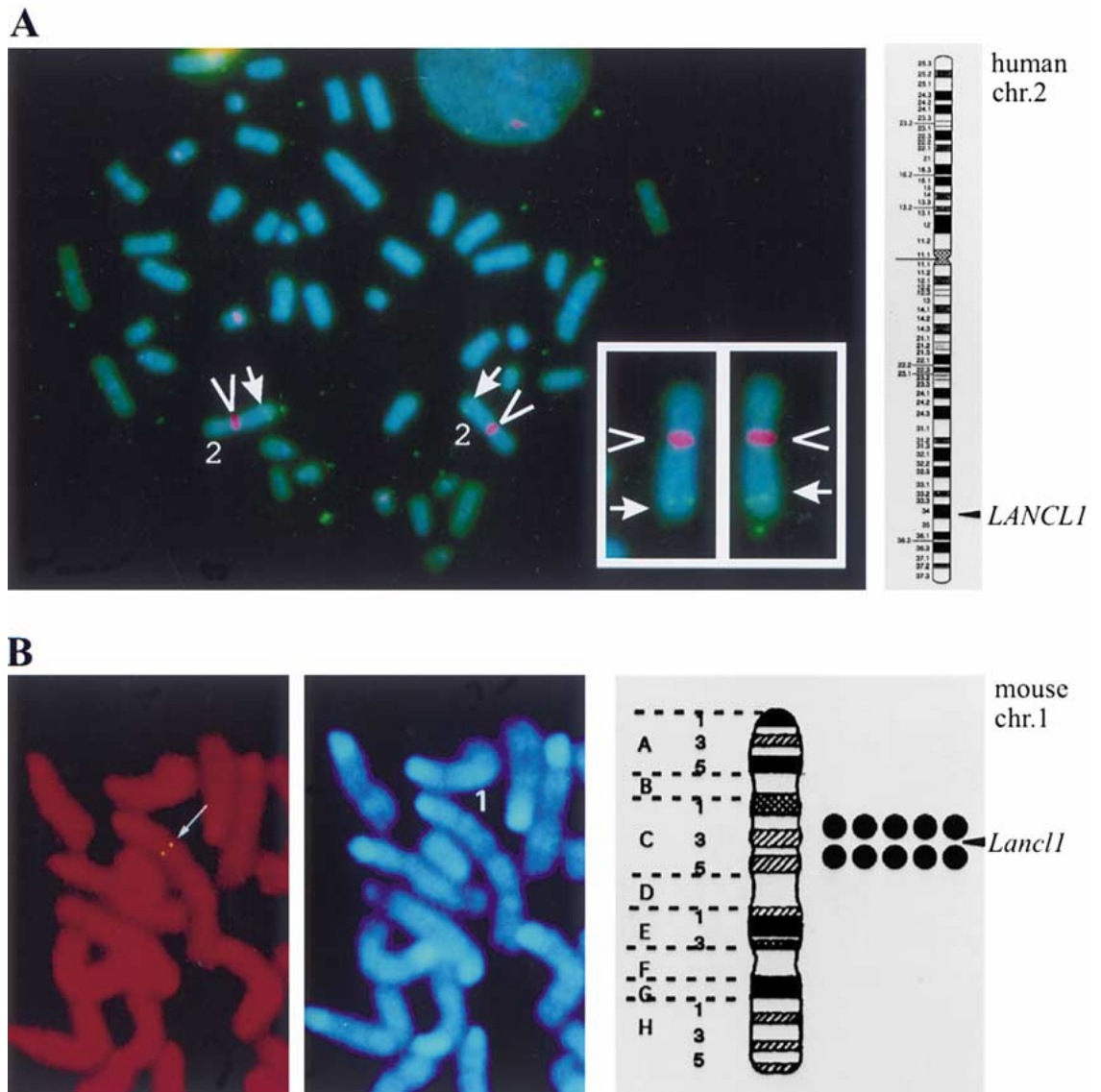


Fig. 2. Chromosomal localization of the human *LANCL1* and mouse *Lancl1* genes mapped by FISH analysis. **(A)** Human *LANCL1*. The specific signals of the P1 clone 11 38 B8 at the long arm of chromosome 2 (green) and those of the probe specific for the centromere of chromosome 2 (red) are indicated by arrows. The diagram on the right illustrates the position of the human *LANCL1* gene in region 2q34. **(B)** Mouse *Lancl1*. The left panel shows the FISH signals on the mouse chromosomes. The middle panel shows the same mitotic figure stained with DAPI to identify mouse chromosome 1. The diagram on the right illustrates the position of the murine *Lancl1* in region C2–C5. Each dot represents the double FISH signals detected on mouse chromosome 1.

from the use of a second polyadenylation signal, giving rise to an mRNA of 4,266 bp, excluding the poly(A)-tail. This signal matches the perfect consensus sequence (5'-AATAAA-3'). BLAST-searches (Altschul et al., 1997) within the mouse section of the dbEST (Expressed Sequence Tag) database revealed the presence of seven EST clones within this region (Gen Bank accession numbers AW124738, AI842791, AA796946, AW456619, BE134741, BE653099, and BF019324). Northern blot analyses show the predominance of the short transcript in mouse (Mayer et al., 1998b). The long mRNA is mainly found in brain suggesting tissue-specific regulatory functions within the 3'-UTR.

Chromosome assignment of human LANCL1 and mouse Lancl1

The chromosome assignment of human *LANCL1* and mouse *Lancl1* was carried out by FISH analyses. A human *LANCL1*-positive P1 clone was cohybridized with a probe specific for the centromere of chromosome 2. This experiment resulted in the specific labeling of the centromere in red and the long arm of chromosome 2 in green (Fig. 2A). Measurements of ten specifically labeled chromosomes 2 demonstrated that the *LANCL1* gene is located at a position which is 79% the distance from the centromere to the telomere of chromosome arm 2q, an area which corresponds to 2q34. A total of 80 metaphase

cells were analyzed with 71 exhibiting specific labeling. Independently, radiation hybrid mapping results have been deposited in the databases for human LANCL1, represented by the STS entries WI-13674, Cda0gh01, SHGC-36761, and A004N22. The gene is thereby assigned to position 2q33→q35, interval 210.9–217.5 cM.

The genomic location of the mouse *Lancl1* gene was determined by FISH using a 5-kb genomic PCR product as a probe. Under the conditions used, the detection efficiency was 85% for the probe. Since the DAPI banding was performed to identify the specific chromosome, the assignment between signals from the probe and mouse chromosome 1 was obtained (Fig. 2B). The detailed position of the *Lancl1* gene was mapped to mouse chromosome 1, region C2–C5. These data are in good accordance with the Human-Mouse Homology Map (<http://www.ncbi.nlm.nih.gov/Homology/>) where a conserved synteny between human chromosome 2q33→q37 and murine chromosome 1 is described. In close proximity to the human LANCL1

gene, two genes encoding the mitochondrial enzymes carbamoylphosphate synthetase I (CPS1) and long-chain acyl-CoA dehydrogenase (ACADL) are located, as determined by the NCBI MapViewer, which have also been mapped to mouse chromosome 1. It is currently unknown, whether LANCL1 is also expressed in mitochondria. While diseases have been described affecting these mitochondrial enzymes, there are no comparable data concerning LANCL1. Several single nucleotide polymorphisms (SNPs) for LANCL1 are defined in the SNP database (NCBI). They are located in various introns and the 3'-UTR. However, there is no correlation between these SNPs and any known hereditary disease.

The availability of the human LANCL1 and mouse *Lancl1* gene structures and organizations will be useful in the characterization of respective gene defects, in the determination and analysis of the function, and in studies of the tissue-specific regulation of LANCL1 and *Lancl1* gene expression.

References

- Akira S, Isshiki H, Sugita T, Tanabe O, Kinoshita S, Nishio Y, Nakajima T, Hirano T, Kishimoto T: A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. *EMBO J* 9:1897–1906 (1990).
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25:3389–3402 (1997).
- Bauer H, Mayer H, Marchler-Bauer A, Salzer U, Prohaska R: Characterization of p40/GPR69A as a peripheral membrane protein related to the lantibiotic synthetase component C. *Biochem biophys Res Commun* 275:69–74 (2000).
- Berkowitz LA, Gilman MZ: Two distinct forms of active transcription factor CREB (cAMP response element binding protein). *Proc natl Acad Sci, USA* 87:5258–5262 (1990).
- Borregaard N, Elsbach P, Ganz T, Garred P, Svejgaard A: Innate immunity: from plants to humans. (Published erratum appears in *Immunol Today* 21:191, 2000). *Immunol Today* 21:68–70 (2000).
- Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH: Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc natl Acad Sci, USA* 93:12355–12358 (1996).
- Ganz T, Lehrer RI: Antimicrobial peptides of vertebrates. *Curr Opin Immunol* 10:41–44 (1998).
- Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A: The Ikaros gene is required for the development of all lymphoid lineages. *Cell* 79:143–156 (1994).
- Harroch S, Revel M, Chebath J: Induction by interleukin-6 of interferon regulatory factor 1 (IRF-1) gene expression through the palindromic interferon response element pIRE and cell type-dependent control of IRF-1 binding to DNA. *EMBO J* 13:1942–1949 (1994).
- Heng HH, Squire J, Tsui LC: High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc natl Acad Sci, USA* 89:9509–9513 (1992).
- Heng HH, Tsui LC: Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* 102:325–332 (1993).
- Kiesau P, Eikmanns U, Gutowski-Eckel Z, Weber S, Hammelmann M, Entian KD: Evidence for a multimeric subtilin synthetase complex. *J Bacteriol* 179:1475–1481 (1997).
- Mayer H, Breuss J, Ziegler S, Prohaska R: Molecular characterization and tissue-specific expression of a murine putative G-protein-coupled receptor. *Biochim biophys Acta* 1399:51–56 (1998b).
- Mayer H, Salzer U, Breuss J, Ziegler S, Marchler Bauer A, Prohaska R: Isolation, molecular characterization, and tissue-specific expression of a novel putative G protein-coupled receptor. *Biochim biophys Acta* 1395:301–308 (1998a).
- Perkins ND: Achieving transcriptional specificity with NF-kappa B. *Int J Biochem Cell Biol* 29:1433–1448 (1997).
- Quandt K, Frech K, Karas H, Wingender E, Werner T: MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl Acids Res* 23:4878–4884 (1995).
- Sahl HG, Bierbaum G: Lantibiotics: biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *A Rev Microbiol* 52:41–79 (1998).
- Schmidt WM, Mueller MW: CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucl Acids Res* 27:e31 (1999).
- Schnell N, Entian KD, Schneider U, Gotz F, Zahner H, Kellner R, Jung G: Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* 333:276–278 (1988).
- Siegers K, Heinzmann S, Entian KD: Biosynthesis of lantibiotic nisin. Posttranslational modification of its prepeptide occurs at a multimeric membrane-associated lanthionine synthetase complex. *J Biol Chem* 271:12294–12301 (1996).
- Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN: A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244 (1990).
- Verrijzer CP, van Oosterhout JA, van Weperen WW, van der Vliet PC: POU proteins bend DNA via the POU-specific domain. *EMBO J* 10:3007–3014 (1991).