 Human Molecular Genetics, 2000, Vol. 9, No. 2 **289–301**

A novel human odorant-binding protein gene family resulting from genomic duplicons at 9q34: differential expression in the oral and genital spheres

Eric Lacazette, Anne-Marie Gachon+ and Gilles Pitiot

Laboratoire de Biochimie Médicale (Prof. B. Dastugue), INSERM U384, Faculté de Médecine, 28 place Henri Dunant, 63000 Clermont-Ferrand, France

Received 17 September 1999; Revised and Accepted 23 November 1999 DDBJ/EMBL/GenBank accession nos AJ251020–AJ251029

Lipocalins are carrier proteins for hydrophobic molecules in many biological fluids. In the oral sphere (nasal mucus, saliva, tears), they have an environmental biosensor function and are involved in the detection of odours and pheromones. Herein, we report the first identification of human lipocalins involved in odorant binding. They correspond to a gene family located on human chromosome 9q34 produced by genomic duplications: two new odorant-binding protein genes (*hOBP*$_{IIa}$ and *hOBP*$_{IIb}$), the previously described tear lipocalin *LCN1* gene and two new *LCN1* pseudogenes. Although 95% similar in sequence, the two *hOBP*$_{II}$ genes were differentially expressed in secretory structures. *hOBP*$_{IIa}$ was strongly expressed in the nasal structures, salivary and lachrymal glands, and lung, therefore having an oral sphere profile. *hOBP*$_{IIb}$ was more strongly expressed in genital sphere organs such as the prostate and mammary glands. Both were expressed in the male deferent ducts and placenta. Surprisingly, alternatively spliced mRNAs resulting in proteins with different C-termini were generated from each of the two genes. The single *LCN1* gene in humans generated a putative odorant-binding protein in nasal structures. Finally, based on the proposed successive genomic duplication history, we demonstrated the recruitment of exons within intronic DNA generating diversity. This is consistent with a positive selection pressure in vertebrate evolution in the intron-late hypothesis.

INTRODUCTION

Olfaction involves the binding of small, hydrophobic, volatile molecules to receptors of the nasal neuroepithelia (1). It generates a cascade of neurological events that transmit the information to the olfactory bulbs projecting into the brain. The very first step in this process is the solubilization of these hydrophobic molecules in the hydrophilic nasal mucus. Odorant-binding proteins (OBPs) are thought to transport these molecules within the mucus (2). These proteins belong to the lipocalin family and so their biochemical structure is well suited for this function. This family, initially described by Pervaiz and Brew (3), comprises >100 small proteins secreted in various biological fluids. They contain eight consecutive β-sheets forming a barrel-shaped hydrophobic pocket (4).

Lipocalins in the mucus of the oral sphere epithelia (upper airway, mouth, orbital area) act as biosensor proteins for the detection of environmental signals. Odorants, which are chemically diverse, are distinguished at the neuroepithelium level using combinations of hundreds of receptors (5). It is unknown whether there are many OBPs transporting odorants within the nasal cavity with high binding specificities, or whether there are fewer with a broader spectrum of binding (2). Determining the number of OBPs secreted from lateral nasal glands could help to determine whether the OBPs have a discriminating function (6). To date, up to three different OBP genes have been identified in a single species (7), but at least eight proteins have been detected in porcupine (8). The overall sequence identity for OBPs is usually described as low both within a single species (6,7,9) and between species (10). However, mouse OBP_{Ia} is 64% similar to mouse OBP_{Ib} and mouse OBP_{II} is 80% similar to rat OBP_{I} (10). This suggests that the OBPs form a heterogeneous group of lipocalins (2,11). Another lipocalin from the oral sphere, the tear lipocalin (TL-VEG), mainly secreted in humans by the lachrymal (12) and salivary (submaxillary and von Ebner's) (13,14) glands, and the secretory units of the trachaea (15), has recently been found in nasal mucus (16). An additional complexity arises from the strong sequence similarity of some OBPs (7) to pheromone carriers such as the two vomeronasal secretory proteins, $VNSP_{I}$ and $VNSP_{II}$, present in the mucus covering the vomeronasal sensory epithelium (17), the major urinary protein subfamily (MUP) synthesized by the liver and excreted in urine (18), and the aphrodisin secreted by the genital tract of the female hamster that induces copulatory behaviour in males (19). MUPs are constitutively produced in the salivary and lachrymal glands. Hence, the relationships between lipocalin pheromone carriers, lipocalin odorant carriers (OBPs) and tear lipocalins are unclear.

Lipocalins are also present in the genital sphere. The tear lipocalin gene (*LCN1-VEGP*) is expressed in the prostate (20). The lactoglobulins are the most abundant lactation proteins in mammals (21), along with the late lactation protein (LALP) and trichosurin in marsupials (22). They are thought to transport retin-

+To whom correspondence should be addressed. Tel: +33 4 73 60 80 24; Fax: +33 4 73 27 61 32; Email: a-m-f.gachon@u-clermont1.fr

oids and fatty acids to neonates. Other lipocalins, including the mouse and rat epididymal retinoic acid-binding protein (E-RABP) (23), the three lizard epididymal secretory proteins (LESP) (24) and human PAEP/glycodelin protein (25) secreted from the genital tract, are involved in the maturation of spermatozoa.

The various physiological functions acquired by these proteins in vertebrates during evolution (26) are based on a binding capacity defined by their membership of the lipocalin family. Evolution has generated diversity, as shown by the low level of sequence identity (~20%), except in proteins from orthologous or recent paralogous genes. Genomic organization provides evidence for the evolutionary relationship of these genes: (i) exons are similar in size, with the corresponding introns identically spliced in phase (27); (ii) positions of intron–exon junctions are well conserved among members [except for retinol binding protein (*RBP4*) and apolipoprotein D (*APOD*) genes]; and (iii) eight genes of the lipocalin family are located on the long arm of human chromosome 9 (28,29), whereas *RBP4* and *APOD* genes occur on human chromosomes 10 and 3, respectively.

We investigated whether human tear lipocalins were produced from two active genes, as in the rat (30), and we found a new family of paralogous genes on human chromosome 9q34, created by recent genomic duplications. We describe two new OBP genes related to *LCN1*, the alternate splicing of their mRNAs and their expression patterns in secretory tissues involved in several functions (olfaction, respiration, taste, lactation and reproduction). We discuss the impact of these results on the classification of lipocalins based on sequence comparisons and expression patterns. Furthermore, the results show an evolutionary mechanism of acquisition of diversity by the recruitment of exons within previous intronic sequences. This provides evidence for positive selection pressure for an 'intron-late' process in vertebrates (31).

RESULTS

LCN1-homologous genes located on human chromosome 9

We previously reported the identification of the *LCN1* cDNA coding for the human tear lipocalin (32) and its mapping to chromosome 9q34 (33,34). Two genes code for the rat proteins homologous to *LCN1*, the von Ebner's gland proteins 1 and 2 (30), which raised the question as to whether additional genes coding for *LCN1* were present in the human genome. Chromosome *in situ* hybridization (33) and somatic hybrid analysis (35) indicated that, if they existed, the additional human genes were located in the 9q34 region. We screened the human chromosome 9-specific cosmid library generated at the Lawrence Livermore National Laboratory (LL09NC01) with the human *LCN1* cDNA probe and identified 26 cosmid clones. They were fingerprinted with *Eco*RI or *Pvu*II and hybridized successively with the *LCN1* cDNA and various oligonucleotides (Fig. 1). Cosmids were assigned to three groups. The first group (clones P32H3, P41B5, P63B6, P92H10, P109C6, P145H6, P195B4, P233G2, P233F2, P265D4 and P276H8) corresponded to the previously reported *LCN1* gene (GenBank accession no. L14927) consisting of seven exons (36). Sequence data for the *LCN1*-homologous region of clone P19E4 (GenBank accession no. Y10826), corresponding to the second group (clones P19E4, P19E7, P42H9, P98H5 and P142H8), demonstrated the presence of an *LCN1b* region that was similar to *LCN1* from the promoter to the sixth exon and divergent thereafter. A third cosmid group determined from partial sequencing of P181A9 (GenBank accession no. Y10827) (clones P110C1, P174E4, P174E5, P181A9, P181B10, P211A7, P238G6 and P291E1) contained an *LCN1c* region very similar to *LCN1* from the promoter to exon 2. Thus, *LCN1* was the only gene that possessed the seventh and final exon. In addition, the TATA boxes of the *LCN1b* and *LCN1c* promoters were degenerate.

Genomic duplications containing lipocalin genes and mapping to chromosome 9q34

At the time of the identification of the *LCN1* gene family, a large physical mapping project produced a cosmid contig map of human chromosome 9q34 (37–39) and the corresponding clones were sequenced by Dr Hawkins and colleagues (Whitehead Institute of the Massachussets Institute of Technology, Boston, MA). Searches of sequence databases with the *LCN1*, *LCN1b* and *LCN1c* sequences revealed strong similarity to the previously reported *LCN1* gene (GenBank accession no. L14927) and to three cosmid sequences: cosmid P161A1 (GenBank accession no. AC002098) P203H12 (AC000396) and P161G2 (AC002106). (From this point onwards, the numbers of the cosmid clones are those of the LL09NC01 library and the corresponding GenBank accession numbers are given in parentheses.) Analysis, in particular of the 3′ end of the *LCN1* genes, showed that the *LCN1c* sequence (Y10827) corresponded to the sequences found in cosmid P161G2 (AC002106), *LCN1b* (Y10826) to cosmid P161A1 (AC002098) and *LCN1* (L14927) to cosmid P203H12 (AC000396) except for a 60 bp insertion at position 12360 of AC000396 relative to L14927. Furthermore, sequence similarities were detected in a region larger than the *LCN1* genes. Dot-plot analyses (Fig. 2) showed that these three cosmids corresponded to areas of genome duplication: cosmid P161A1 (AC002098) and cosmid P203H12 (AC000396) sequences were similar over their entire length, whereas cosmid P161G2 (AC002106) was similar to the others only for the sequences upstream from *LCN1* intron 3.

The positions of the duplicated areas on chromosome 9 were determined. We sequenced the cosmid insert extremities and compared them with sequences in databases. The sequence of cosmid P181A9 T3-extremity (cosmid group containing *LCN1c*) contained part of the *Surf5* gene (Fig. 3). This result, placing *LCN1c* upstream from the *Surfeit* locus, was confirmed by the presence of the sequence of cosmid P161G2 (AC002106) in a sequence contig between *ABO* and the *Surfeit* locus, and was consistent with the results of Hornigold *et al.* (39) using the same LL09NC01 library. The limitations of fingerprinting for duplicated areas probably explain the divergence that we observed for the P161A1 and P203H12 locations. P203H12 (AC000396) contained the *LCN1* gene that we previously mapped close to *D9S1826* (34); using *LCN1*-specific polymerase chain reaction (PCR) (based on sequences from AC000396 and L14927) on 150 genomic DNAs we showed that the 60 bp deletion in the previously reported *LCN1* sequence (L14927) does not exist (data not shown). Cosmid P161A1 extremity sequences (AC002098) and our *LCN1b* cosmid extremity sequences were not anchored to any sequence in the database. A new minisatellite (AJ251020) (Fig. 3) located in the *LCN1b* subfamily of cosmids (located at position 3177–3724 of AC002098) detected a rare polymorphism (PIC = 0.05 for 20 unrelated individuals), and was informative in CEPH reference family 1362. Linkage analysis

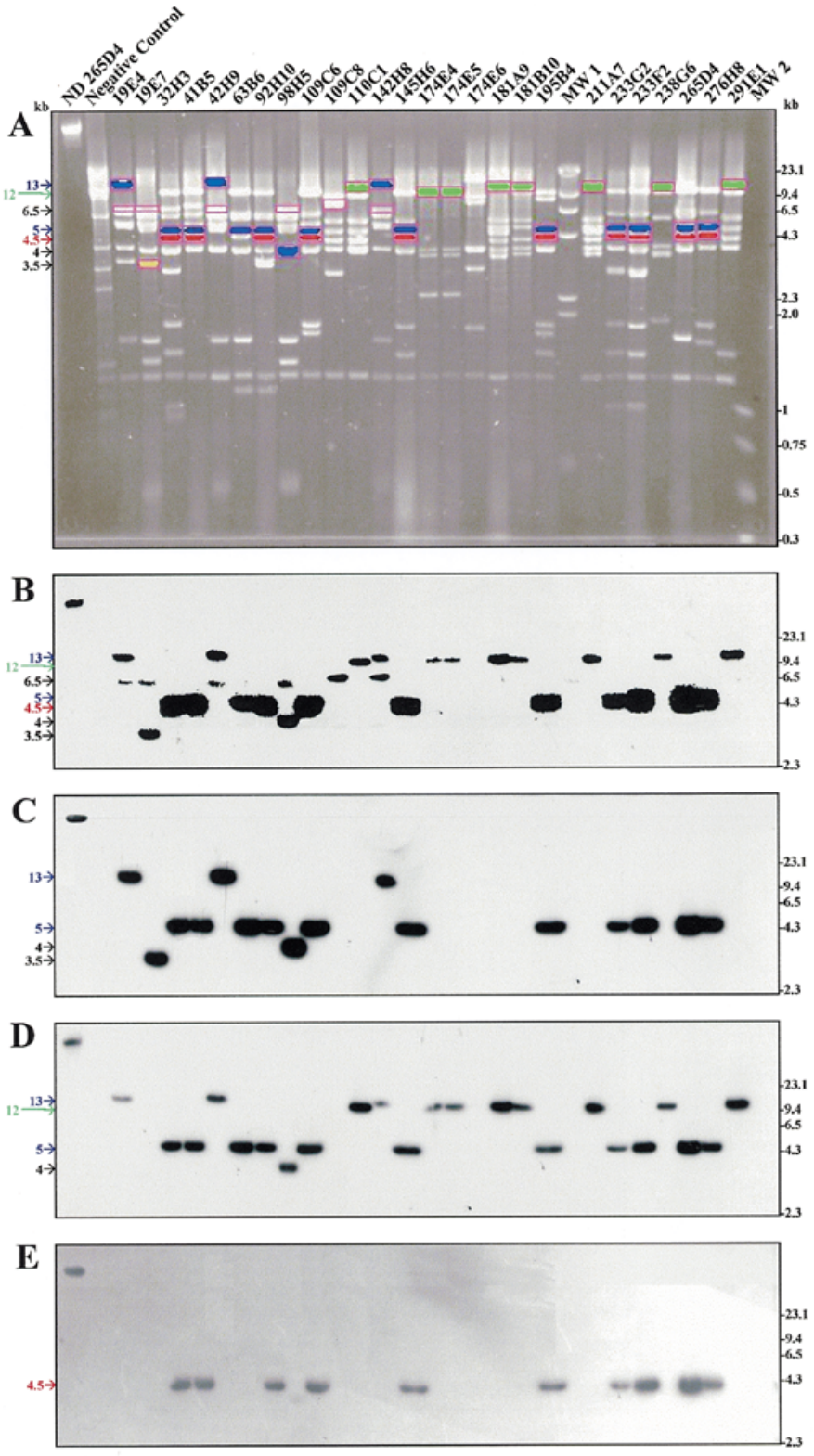


Figure 1. Fingerprinting analyses of '*LCN1*-type' cosmid clones. (**A**) PFGE analysis of cosmid clone DNA digested with *Eco*RI and stained with ethidium bromide. Coloured bands show the hybridization results presented in (B–E): yellow, oligonucleotide EL2; green, EL1; red, EL3; blue, co-detection with oligonucleotides EL1 and EL2; bands framed in pink correspond to the *LCN1* cDNA probe. (**B**) *LCN1* cDNA hybridization. (**C**) EL2 hybridization. (**D**) EL1 hybridization. (**E**) EL3 hybridization. The sizes indicated by coloured arrows correspond to the coloured bands of 1A.

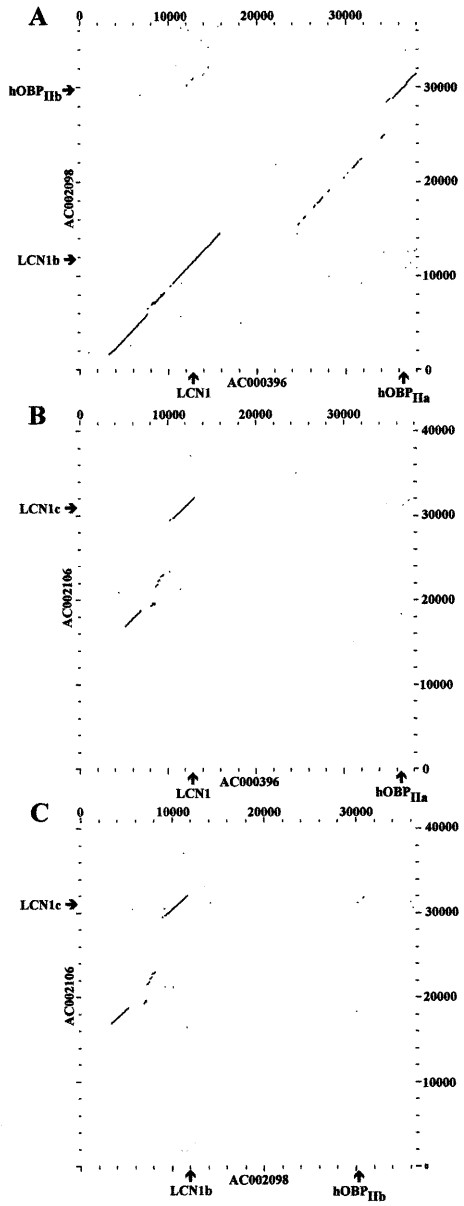


Figure 2. Dot-plot analysis of (**A**) *LCN1-hOBPIIa* locus sequence (AC000396 + AJ251029) versus *LCN1b-hOBPIIb* locus sequence (AC002098); (**B**) *LCN1-hOBPIIa* locus sequence (AC000396 + AJ251029) versus *LCN1c* locus sequence (AC002106); (**C**) *LCN1b-hOBPIIb* locus sequence (AC002098) versus *LCN1c* locus sequence (AC002106). Genomic sequences were filtered for human repetitive sequences using RepeatMasker, compared using a 25 word size with a stringency of 20, and dot-plot analysis was carried out with the GCG package.

revealed two-point lod scores >3 at q = 0 for *D9S275* and *D9S1818*. Haplotype reconstruction confirmed the location of the *LCN1b* gene between *D9S1811* and *D9S67* on chromosome 9q34 (Fig. 3).

Identification of two new OBP genes

By comparing sequences with those in databases, we found in sequences from cosmids P203H12 (AC000396) and P161A1 (AC002098) regions of similarity to lipocalin genes outside the *LCN1*, *LCN1b* and *LCN1c* gene areas (Fig. 3). We identified a sequence at position 2150 of cosmid P161A1 (AC002098), 20 kb downstream from *LCN1b*, identical to a human testis expressed sequence tag (EST AA460385) corresponding to four exons of a new lipocalin gene that was similar to rat OBP_{II}. Similarly, a putative 50 bp exon similar to the EST sequence was also found at the extremity of cosmid P203H12 (AC000396). Owing to the genomic duplication, we sequenced cosmid P233G2, which contained regions located downstream from *LCN1* (Fig. 3), with oligonucleotides corresponding to the sequence of the EST and identified another lipocalin gene 20 kb distal to *LCN1*. To identify the first exons in each of the two new genes, nested PCR was performed using cDNA clones from a testis library and oligomers corresponding to the 5′ region of the EST and the vector arms. PCR products were cloned and sequenced, and we identified three additional exons as compared with the genomic sequences (Fig. 3). A TATA box was present upstream from the first exon in both cases (Fig. 4a). The two mRNAs, *hOBP$_{IIa}$* corresponding to the gene located downstream from *LCN1* and *hOBP$_{IIb}$* located downstream from *LCN1b* (Fig. 4), were 97.5% identical to each other and 63% identical to *LCN1*. Their intron–exon organizations were consistent with those of seven-exon genes of the lipocalin family.

The hOBP$_{IIa\alpha}$ and hOBP$_{IIb\alpha}$ proteins are traditional lipocalins

The deduced protein sequences (170 amino acids) of hOBP$_{IIa\alpha}$ and hOBP$_{IIb\alpha}$ confirmed their membership of the lipocalin family. The two proteins, hOBP$_{IIa\alpha}$ (mol. wt = 17.8 kDa) and hOBP$_{IIb\alpha}$ (mol. wt = 18.0 kDa), were 89% identical. Each had a putative 15 amino acid signal peptide (Fig. 4c), the conserved lipocalin motif G-X-W at positions 27–30 (40). Their amino acid sequences were 45.5% identical to that of rat OBP_{II}, 43% identical to that of human tear lipocalin (TL-VEG) and much lower (15–25%) for other lipocalins. The calculated isoelectric points (pIs) of hOBP$_{IIa\alpha}$ and hOBP$_{IIb\alpha}$ were 7.85 and 8.72, respectively, whereas those of lipocalins are generally acidic (~4.5) except that of rat OBP_{II} (pI = 9.01). Eight β-sheets (possibly forming a barrel) followed by an α-helix and a final β-sheet were predicted for the two proteins with the DSC program using lipocalin multiple alignment, consistent with the data for other members of this family (Fig. 4c). We compared sequences with those of other lipocalins studied by crystallography using the automated Swiss-Model protein modelling service. β-lactoglobulin and RBP were used as matrices to identify a very first understanding of the three-dimensional (3D) structures for the hOBP$_{IIa\alpha}$ and hOBP$_{IIb\alpha}$ proteins (Fig. 5). We found that they consisted of eight anti-parallel β-sheets (A–H) defining a barrel and a final α-helix, consistent with the structure of other lipocalins. Their structures are presumably locked by a disulfide bridge between cysteines 58 and 150. In addition, previously described hydrophobic amino acids implicated in ligand interactions (41) are conserved in the hOBP$_{IIa\alpha}$ and hOBP$_{IIb\alpha}$ proteins, strongly suggesting that these two molecules have ligand-binding activity (Fig. 5, Phe 51, Phe 53, Ile 64, Tyr 78), like the orthologous rat OBP_{II} protein (42).

The two paralogous *hOBP$_{II}$* genes are expressed differently

Gene expression was investigated in 18 human tissues by RT–PCR using *LCN1*- or *hOBP$_{II}$*-type sets of primers and gene-specific oligonucleotide hybridizations (Fig. 6). *LCN1b* and

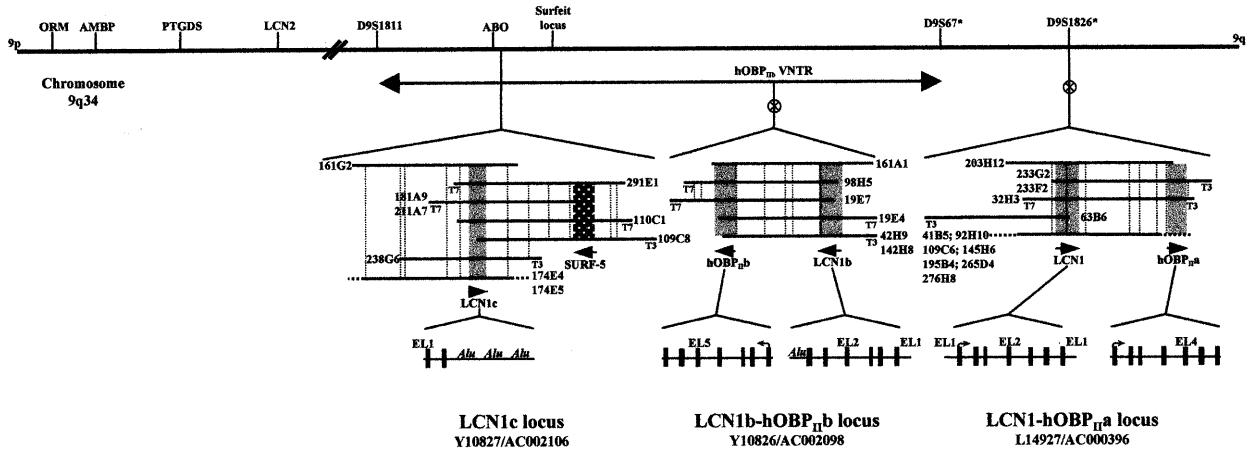


Figure 3. Genomic organization of the *LCN1*/*hOBP$_{II}$* duplicated loci. The top line represents the chromosome 9q34 region with anchored polymorphic loci. The double arrow indicates the location interval for the *hOBP$_{IIb}$* minisatellite (GenBank accession no. AJ251020) and the asterisk an uncertain relative position between *D9S67* and *D9S1826*. The middle line shows the partial cosmid organization at three different loci: *LCN1c*, *LCN1b-hOBP$_{IIb}$* and *LCN1-hOBP$_{IIa}$*. Cosmid clones (~40 kb) are represented by horizontal lines, with names and T3 or T7 vector arm indicated above the line. Arrows indicate the gene or pseudogene orientations within the cosmids and the vertical dotted lines the *Eco*RI sites. The cross in a circle indicates an uncertain locus orientation. The bottom line shows the intron–exon structure of the *LCN1* and *hOBP$_{II}$* genes: black boxes represent exons; arrows are transcription start sites, EL1, EL2, EL3, EL4 and EL5 are the oligonucleotide probes used for screening and *Alu* the repetitive sequences.

LCN1c were not expressed, whereas *LCN1* mRNA was detected in the lachrymal gland, the sweat and von Ebner's glands, the nasal septum and turbinate epithelia, the placenta and the mammary gland (Fig. 6a) and, at very low level, in the prostate. In addition, the *hOBP$_{IIa}$* and *hOBP$_{IIb}$* genes were expressed differently, despite their sequences being very similar, including the 1.5 kb promotor region (Fig. 6b). *hOBP$_{IIa}$* was strongly expressed in the nasal septum, middle meatus, turbinates, lung, testis and placenta, and less strongly in lachrymal, sweat and von Ebner's glands. In contrast, *hOBP$_{IIb}$* was expressed predominantly in the prostate, testis and mammary gland, and weakly in the submaxillary gland, nasal septum, middle meatus and lung.

Different alternatively spliced mRNAs generate diversity in the C-terminus of the proteins

Surprisingly for OBP genes, RT–PCR analyses detected large amounts of seven alternatively spliced mRNAs (Figs 4 and 6). For the *hOBP$_{IIa}$* gene, three different acceptor splice sites were detected for exon 5 (Fig. 4a and b). An alternative acceptor splice site for exon 5 (exon 5b) located 49 bp upstream from the known site generated a 725 nucleotide mRNA. The hOBP$_{IIa\beta}$ protein was 146 amino acids long, identical to hOBP$_{IIa\alpha}$ until the eight putative β-sheets, and different only for the 16 additional amino acids. A third exon 5 acceptor splice site located 65 bp (exon 5c) upstream from exon 5a generated an mRNA of 741 nucleotides. The resulting 228 amino acid hOBP$_{IIa\gamma}$ protein was identical to hOBP$_{IIa\alpha}$ for the first eight putative β-sheets and differed in its C-terminal region (Fig. 4c), predicted by Predator software to give a long coiled region with a ninth β-sheet. For the *hOBP$_{IIb}$* gene, there was an extra 106 bp exon (exon 3b) between exons 3 and 4 (Fig. 4a and b). The resulting mRNA (782 nucleotides) coded for hOBP$_{IIb\beta}$, a 165 amino acid protein which is identical to hOBP$_{IIb\alpha}$ up to the fifth putative β-sheet and different thereafter, with a predicted α-helix for the 'ALWEALAIDTLRK' motif downstream from the fifth β-sheet, followed by two additional β-sheets in the long C-terminal part. None of these alternative splice variants were detected for the other gene, although the putative acceptor and donor splice sites were present (Fig. 4a). In addition, low levels of alternatively spliced mRNAs missing exon 2 but with exon 5b for *hOBP$_{IIa\delta}$*, or with exon 5 for *hOBP$_{IIb\gamma}$*, generated putative secreted proteins of 147 and 85 amino acids, respectively (Fig. 4b and c), that diverged from typical lipocalin sequences after the 24th amino acid.

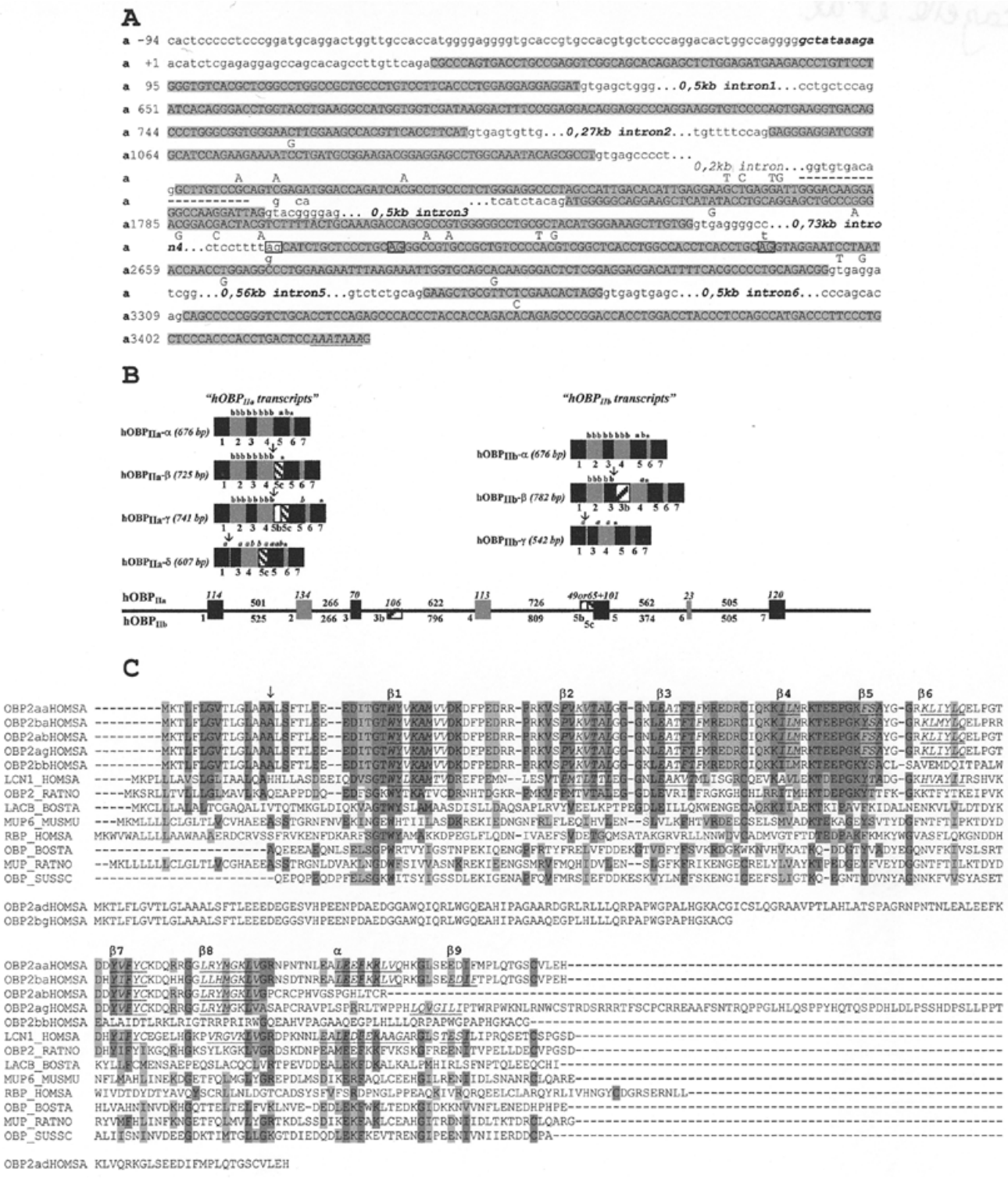
Secretory epithelia of the organs from the oral and genital spheres express *hOBP$_{II}$* genes

To identify the cell type in which *hOBP$_{II}$* transcripts were produced, *in situ* hybridization was performed on tissue sections. Sections were hybridized with digoxigenin-labelled sense or antisense *hOBP$_{II}$* probes and mRNAs were detected in acinar cells from the middle meatus and turbinates, and in epithelial cells from turbinates (Fig. 7), consistent with an olfactory function for hOBP$_{II}$ proteins. We also detected *hOBP$_{II}$* mRNAs in the genital sphere, namely in glandular cells from the prostate and breast, and epithelial secretory cells from the deferent duct and mammary gland. Combining these results with those from RT–PCR, we suggest that the five main hOBP$_{II}$ proteins (hOBP$_{IIa\alpha}$, hOBP$_{IIa\beta}$, hOBP$_{IIa\gamma}$, hOBP$_{IIb\alpha}$, hOBP$_{IIb\beta}$) are secreted by the epithelial cells of the male gonad ducts, the lung, the placenta and the acinar cells of the middle meatus and turbinates with large amounts of *hOBP$_{IIa}$* mRNAs in the nasal tissues. In the prostate and mammary glands, it may be that only the two major hOBP$_{IIb}$ proteins (hOBP$_{IIb\alpha}$ and hOBP$_{IIb\beta}$) are secreted by the epithelial cells.

DISCUSSION

We detected an *LCN1*-type gene family generated by genomic duplications on human chromosome 9q34, which contained in addition to *LCN1* two *LCN1* pseudogenes and two *hOBP$_{II}$* genes paralogous to *LCN1*. The two *hOBP$_{II}$* genes were expressed differently in the oral sphere (nasal epithelia, lung, von Ebner's glands, submaxillary glands, lachrymal glands) and in the genital sphere (deferent duct, vaginal epithelium, prostate and mammary glands). Three-dimensional modelling is consistent with a ligand-binding function, previously described for the orthologous rat OBP$_{II}$ (42). Various alternatively spliced forms were produced from each gene, generating proteins with different C-termini.

We found that the *hOBP$_{II}$-LCN1* family was produced by successive duplication events. The first was a tandem duplication of a seven-exon lipocalin ancestor with an exon 5a and no exon 3b (Fig. 8). This hypothesis is also supported by phylo-



genetic analyses (22, and unpublished data), indicating that the *LCN1-VEGP* and OBP_{II} genes correspond to a subfamily of lipocalin genes and probably have a common ancestor. *LCN1* exon 5 is 102 bp long, like exon 5a (101 bp). Exon 5a is present in the mRNAs generating the $hOBP_{IIa\alpha}$ and $hOBP_{IIb\alpha}$ proteins, the two $hOBP_{II}$ variants most closely related to *LCN1*. No exon 3b was detected for *LCN1*. This organization is consistent with that of the other seven-exon lipocalins (27). It suggests that the $hOBP_{II}$ proteins evolved by integrating additional surrounding intronic DNA into mRNAs via an upstream acceptor splice site for $hOBP_{IIa}$ exon 5 and the recruitment of an extra exon (exon 3b) for $hOBP_{IIb}$. The secondary events were the three complete or partial duplications of this 50 kb region on human chromosome 9q34. Two *VEGP* genes are expressed in rat, indicating that *LCN1b* was inactivated after the second duplication. The insertion of numerous Alu sequences downstream from *LCN1b* exon 6 may be the primary inactivation event. Clusters of lipocalin genes have been reported for *MUP*, with a 45 kb motif (43), and for *ORM* (44). We can wonder whether these genomic areas are paralogous. *MUP* genes are divergently oriented within and between two consecutive motifs, which was not the case here. In the *ORM* cluster there are three consecutive genes, whereas there were two here. Thus, these clusters are probably not paralogous, but instead correspond to independent duplication events in different ancestral genes. This is probably not the case for the milk proteins of marsupials (45), the sequences of which appear to be related to that of LCN1 in phylogenetic analysis (22). Preliminary results suggest that the late lactation protein and trichosurin genes of *Trichosurus vulpecula* are <20 kb apart (22), and that the late lactation protein and β-lactoglobulin in the tammar wallaby are closely linked (42). These data suggest that the duplication events described here occurred before the emergence of mammals. These milk proteins are thought to transport retinol or fatty acids from the mother to the young, and could participate in the discrimination between early- and late-lactating mammary glands. They may therefore represent a physiological and phylogenetic link between the traditional function of β-lactoglobulins and $hOBP_{IIb}$ proteins as retinol or fatty acid carriers in the mammary gland and smell or taste functions for the nasal OBP proteins ($hOBP_{IIa}$).

The evolution of the lipocalin gene family involved numerous tandem duplications, suggesting that such duplication may result in the acquisition of a new function or the production of a large amount of protein. Fattori *et al.* (16) counted the number of OBPs in a single species to investigate whether OBPs discriminate between different ligands. In humans, $hOBP_{IIb}$ was more highly expressed in the oral sphere than was $hOBP_{IIa}$, but both were present. *LCN1* was expressed in nasal structures and lachrymal glands, which are connected to the nasal cavity via the lachrymo-nasal duct. Based on the predicted barrel structure (Fig. 5) of $hOBP_{IIa\alpha}$ and $hOBP_{IIb\alpha}$ containing amino acids previously described as interacting with hydrophobic ligands (41) and the presence of these two proteins in the nasal mucus, we conclude that $hOBP_{IIa\alpha}$, $hOBP_{IIb\alpha}$ and TL-VEG proteins are OBPs. In addition, rat OBP_{II}, which is orthologous to the $hOBP_{IIa\alpha}$ and $hOBP_{IIb\alpha}$ proteins (data not shown), has been found to bind some odorants (46), and ligand-binding capacity has been demonstrated for TL-VEG (47). The $hOBP_{IIa\beta}$ and $hOBP_{IIa\gamma}$ proteins from alternatively spliced mRNAs have a lipocalin structure with eight β-sheets and possible cysteines for disulfide bridges, suggesting OBP function. The main differences in the C-terminal parts of the molecules may be due to differences in binding capacities to particular odorants. However, 3D ribbon-view predictions indicated that these C-terminal parts correspond to one side of the molecule and may be involved in protein–protein interactions such as dimerization (7,41) or interaction with specific receptors (48). A major difference from the traditional lipocalin structure is found with the $hOBP_{IIb\beta}$ protein, which retained only the first five β-sheets followed by an α-helix and additional β-sheets. The question arises as to whether the β-sheets located at the C-terminus of the molecule may replace the three missing β-sheets in the barrel. Such small variations around the traditional structure have been described for triabin (49). Conversely, the $hOBP_{IIa\delta}$ and $hOBP_{IIb\gamma}$ proteins, produced from mRNAs lacking exon 2, are not lipocalins and may have resulted from the transcription background. However, alternatively spliced forms of *PAEP* mRNAs lacking exon 2 have been described, leading to a markedly different protein structure (25). An immunosuppressive function that may not require the classical barrel has also been demonstrated (50).

The notion that different functions, such as immunosuppression or contraception, may be acquired by structurally divergent lipocalins is supported by our finding that the $hOBP_{II}$ proteins are produced by cells of the genital sphere. The $hOBP_{IIb}$ gene is expressed mainly in the prostate and deferent duct, whereas $hOBP_{IIa}$ gene expression in the genital sphere is limited to the deferent duct. The human glycodelin-S (25), the mouse and rat E-RABP (also called epididymal secretory protein 18.5 kDa) and the lizard epididymal secretory proteins are also lipocalins secreted into the seminal fluid (24,51). Other proteins from the CRISP and HE1 families are secreted by the epithelial cells of the genital glands (52). These secretions are known to coat spermatozoa and to be necessary for their maturation. Spermatozoa also express some olfactory receptors (53,54), and are probably the target cells for these lipocalins. The molecular function of the lipocalins present in the seminal

Figure 4. Genomic organization of the $hOBP_{II}$ genes, the corresponding mRNAs and sequence alignment of lipocalins. (**A**) Nucleotide sequences of the $hOBP_{II}$ genes. Upper lines represent $hOBP_{IIa}$ sequence (lines starting with a), lower lines indicate the nucleotides that differ in $hOBP_{IIb}$ and – indicates no corresponding sequence. Shaded capital letters are exon sequences and lower case letters intron sequences. Sizes indicated on the left are in base pairs. The TATA box is shown in bold and the polyadenylation signal is underlined. Boxes indicate acceptor splice sites for exons 5, 5b and 5c. (**B**) Schematic representation of the two $hOBP_{II}$ genes and their corresponding mRNA. The horizontal shows the exon–intron organization with sizes in base pairs. Solid shaded boxes numbered from 1 to 7 are the coding sequences of the main transcripts, b and c refer to additional exons. The various organizations of mRNA generated by alternative splicing are represented by the assembled boxes. ↓ indicates a frameshift resulting from the insertion or deletion of an exon and an asterisk a stop codon. a represents the α-helix and b the β-sheets predicted with DSC. Letters in italics are predictions made with Predator. (**C**) Sequence alignment of the human $hOBP_{IIa}$ and $hOBP_{IIb}$ proteins (aα:OBP2aaHOMSA, bα:OBP2baHOMSA, bβ:OBP2bbHOMSA, aγ:OBP2agHOMSA, aβ:OBP2abHOMSA) with human tear lipocalin (LCN1_HOMSA), rat OBP_{II} (OBP2_RATNO), bovine BLG (LACB_BOSTA), mouse MUP (MUP6_MUSMU), human RBP (RBP_HOMSA), bovine OBP (OBP_BOSTA), rat MUP (MUP_RATNO) and porcine OBP (OBP_SUSSC). Residues in dark grey boxes are identical and those in light grey boxes are similar. Secondary structure elements predicted with the DSC program are underlined and amino acid residues are in italics. β-sheets and α-helices are numbered for $hOBP_{IIa}$ and $hOBP_{IIb}$. The predicted cleavage site of the signal peptide is indicated by an arrow (AAA↓LS) at position 15. Unaligned sequences of divergent $hOBP_{II}$ gene spliced forms, $hOBP_{IIa\delta}$ (OBP2adHOMSA) and $hOBP_{IIb\gamma}$ (OBP2bgHOMSA), are at the bottom.

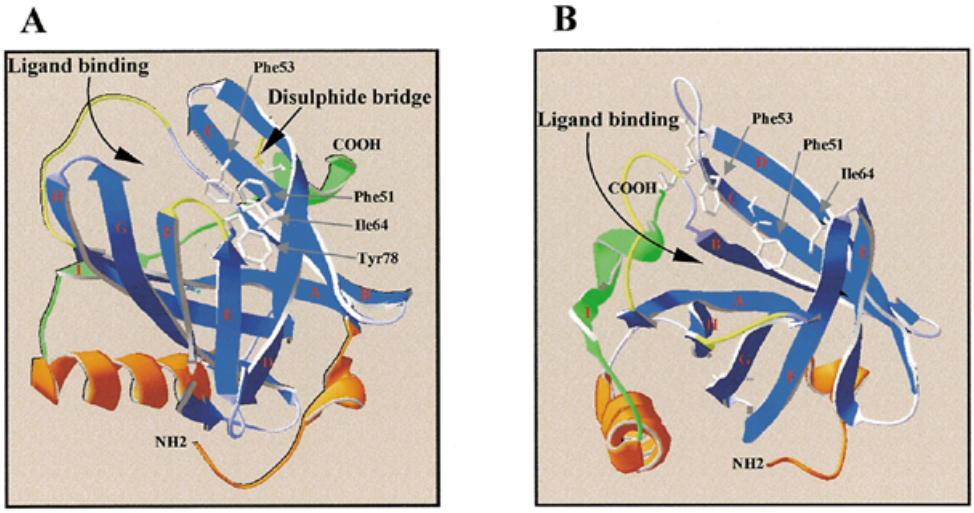


Figure 5. The Swiss-Model-deduced tertiary structure of (**A**) $hOBP_{IIa\alpha}$ and (**B**) $hOBP_{IIb\alpha}$. Ribbon-form view of the two predicted 3D structures of human $hOBP_{II}$ proteins. β-sheets are indicated in blue, α-helices in orange, predicted disulfide bridge in yellow, C-termini below the α-helices in green, and the loops between β-sheets predicted to contain acidic amino acids are in yellow. The eight β-sheets are labelled from A to H. Conserved hydrophobic amino acids contained in the pocket and previously described as interacting with the ligands are indicated (Phe51, Phe53, Ile64, Tyr78). The two $hOBP_{II}$ proteins were modelled separately by the Swiss-Model software. No crystallographic data were available for the orthologous protein or TL-VEG. The software therefore used the most similar sequences: RBP and β-lactoglobulin were used for matrices for $hOBP_{IIa\alpha}$ and $hOBP_{IIb\alpha}$, respectively. The 30% similarity threshold necessary for modelling was attained, so the models presented give a reasonable representation of reality, especially for a well conserved family such as that of the lipocalins. However, the differences in the models produced (e.g. difference in opening of the barrel, difference in disulfide bridge prediction) for two proteins 97.5% identical illustrate that this is only the first step towards understanding the real 3D structure of the proteins.

fluid is unknown, but they are probably involved in reproductive processes. Furthermore, the production of $hOBP_{IIb}$ proteins by the tubulo-acinar secretory cells of the mammary glands demonstrates the recruitment of the corresponding gene for lactation. To date, this is the only lipocalin described to be involved in lactation in humans, whereas β-lactoglobulin is known to transport retinoids and fatty acids to the newborn in many mammals.

The results presented illustrate that the biochemical characteristics of lipocalins have been applied to various physiological functions, mainly via new genes, but possibly also via the recruitment of previous genes acquiring new functions in different organs (26). This may result in confusion in terms of nomenclature: the $hOBP_{II}$ proteins could equally have been called tear lipocalins, odorant-binding proteins, lactation proteins and deferent duct secretory proteins, illustrating their pleiotropic capacity. This also indicates that many different lipocalins may be involved in a particular function. It is not known whether lipocalins bind the same ligand to fulfil different physiological functions. However, their genes have been duplicated frequently during evolution to generate proteins with different binding capacities, and their promoters have evolved for recruitment in different physiological functions. Furthermore, we found additional exons 5b and 5c within intron 4 of the *$hOBP_{IIa}$* gene, resulting in protein diversity. This intron was not present in *APOD*, the vertebrate lipocalin most similar to invertebrate lipocalins (55). This would support the intron-late hypothesis in vertebrates as a means of generating diversity.

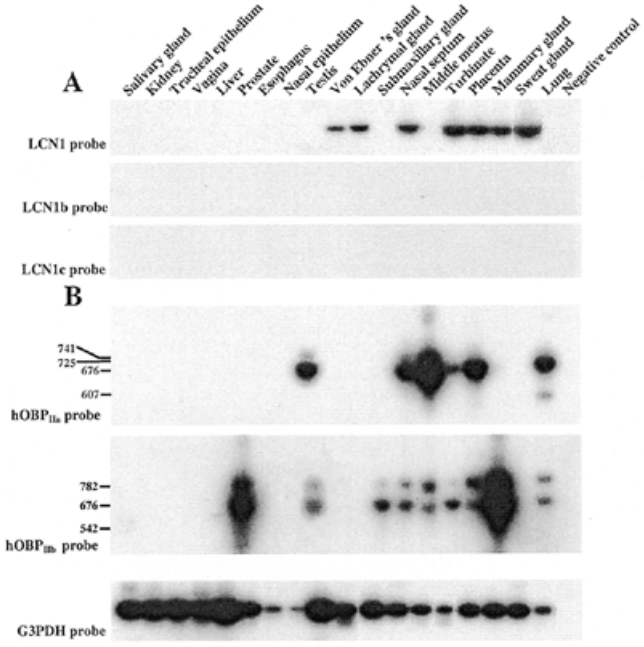


Figure 6. RT–PCR analysis. (**A**) *LCN1*, *LCN1b* and *LCN1c* or (**B**) *$hOBP_{IIa}$* and *$hOBP_{IIb}$* RT–PCR products were detected with their respective specific oligonucleotide probes. (**C**) RNA quality was checked by detection of *GAPD* RT–PCR products (G3PDH probe). Sizes are indicated in base pairs. Note that the autoradiographs presented correspond to exposures optimized for tissues with high levels of expression. As stated in the text, lower levels of tissue expression can be identified after a much longer exposure: *LCN1* in prostate and *$hOBP_{IIa}$* in lachrymal, sweat and von Ebner's glands.

MATERIALS AND METHODS

Genomic cloning

We used a copy of the chromosome 9-specific cosmid library LL09NC01 constructed by Dr J. Allmeman (Biochemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA) under the auspices of the National Gene Library Project sponsored by the US Department of Energy.

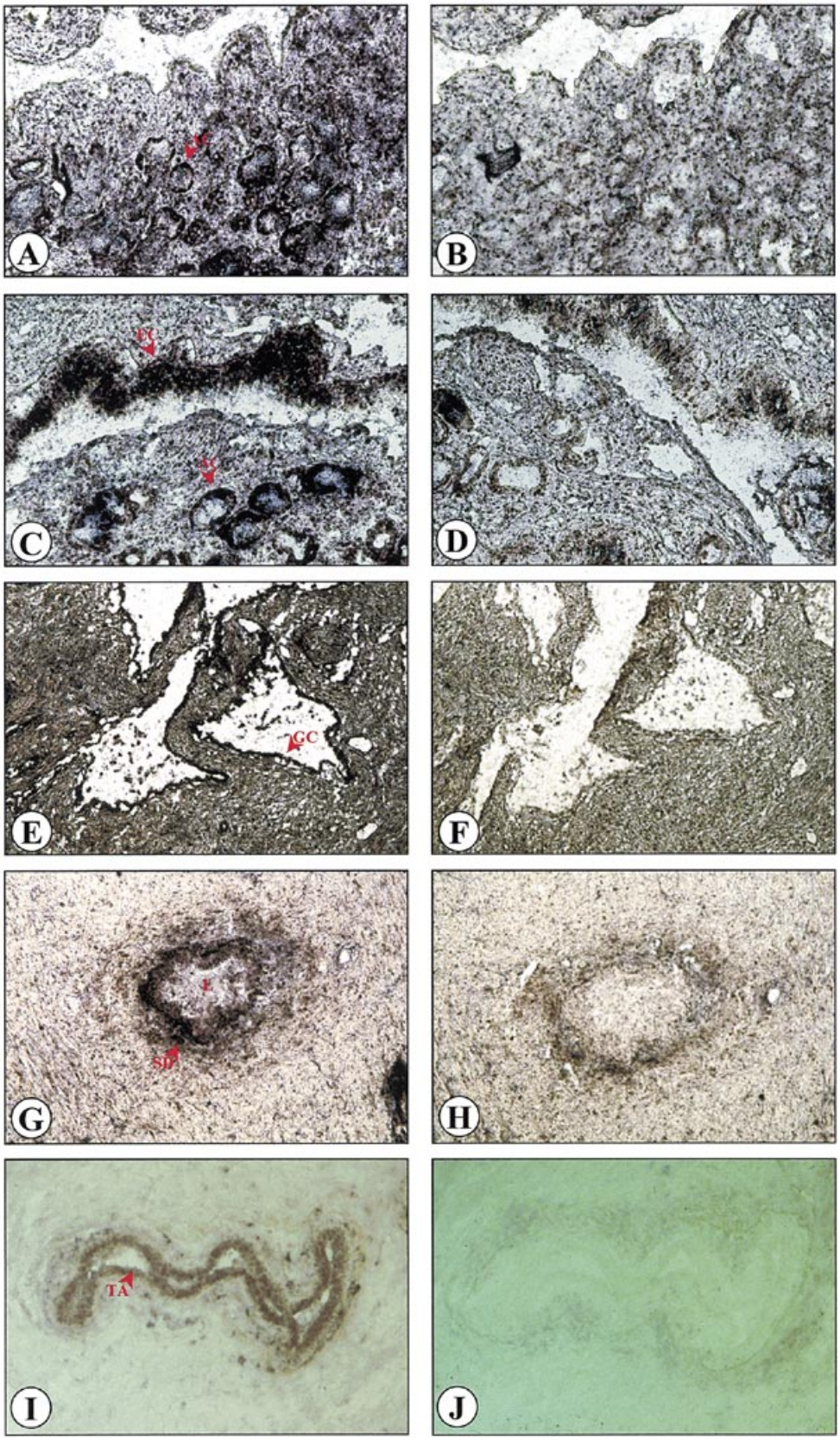


Figure 7. Tissue section *in situ* detection of *hOBPII* mRNAs. Middle meatus (**A** and **B**), turbinates (**C** and **D**), prostate (**E** and **F**), deferent duct (**G** and **H**) and mammary gland (**I** and **J**) sections were hybridized with digoxigenin-11-UTP-labelled $hOBP_{II}$ riboprobes. No signal was obtained with the sense *hOBPII* riboprobe (B, D, F, H and J). Note the specific signal for antisense riboprobe (A, C, E, G and I). Arrowheads indicate the various structures: AC, acinar cells; EC, epithelial cells; GC, glandular cells; SD, secretory duct; L, lumen; TA, tubuloacinar cells (×100).

Screening and clone analyses were performed as described previously (34).

Cloning and sequencing analysis

Thirty PCR cycles (94°C for 45 s, 54°C for 45 s, 72°C for 1 min 30 s) with 10^7 p.f.u. of a Clontech (Montigny-le-Bretonneux, France) λgt11 human testis cDNA library were performed with oliEST58 CCTGCAGGTACATGAGCTTCC and 5′ or 3′ insert screening amplimers (Clontech) located on λgt11 vector arms. Nested PCR was performed with oliEST26 CGCTGTATTTGCCAGGCTCC and vector arm oligonucleotides. PCR products were ligated into the pGEM-T vector, giving the 5′ part of the $hOBP_{II}$ cDNAs. The products of sequencing reactions with standard pGEM-T oligonucleotide and dye terminator cycle sequencing ready reaction mix (Applied Biosystems, Courtabœuf, France) were subjected to electrophoresis on an ABI PRISM 377 automatic sequencer (Perkin Elmer, Courtabœuf, France), and analysed with the Sequence Navigator 1.0.1 software (Perkin Elmer). Full-length cDNA clones ($hOBP_{IIa\alpha}$, $hOBP_{IIa\beta}$, $hOBP_{IIa\gamma}$, $hOBP_{IIa\delta}$, and $hOBP_{IIb\alpha}$, $hOBP_{IIb\beta}$, $hOBP_{IIb\gamma}$) were obtained by RT–PCR (described below) by purification of the bands of interest with a gel extraction kit according to the manufacturer's protocol (Qiagen, Courtabœuf, France) or by subcloning nested PCR products for weakly expressed alternative forms, and ligation into pGEM-T vector.

RT–PCR analysis

Tissue samples were collected from 45- to 55-year-old Caucasian individuals in accordance with French law. Total RNA was extracted by the single-step method using the RNA NOW mixture according to the manufacturer's protocol (Biogentex, Montigny-le-Bretonneux, France). Total RNA (5 µg) was reverse transcribed in a final volume of 20 µl containing 0.5 µg of oligonucleotide GACTCGAGTCGACATCGATTTTTTTTTTTT-TTTTT with the Superscript pre-amplification system (Gibco BRL, Gaithersburg, MD). The products of this reaction (3 µl) were used for subsequent PCR. Specific mRNAs were determined by PCR using primers: TL, CCTCTCCCAGCCCCAGCAAG, and AP, GACTCGAGTCGACATCG, for *LCN1*-type genes (*LCN1*, *LCN1B*, *LCN1C*), and DE, CGCCCAGTGACCT-GCCGAGGTC, and FI, CTTTATTTGGAGTCAGGT-GGGTG, for $hOBP_{II}$-type genes. As controls, we used primers: G3PDH1, CTCTGCCCCCTCTGCTGATG, and G3PDH2, CCTGCTTCACCACCTTCTTG, specific for the G3PDH gene, which is considered to be expressed constitutively in all cell types. Thirty-two PCR cycles (94°C for 45 s, 54°C for 45 s, 72°C for 2 min 30 s) were performed and the amplification products were separated by electrophoresis in a 1% agarose gel. DNA was transferred to a Hybond N^+ membrane.

We detected the expression of the various genes using several specific oligonucleotides:

olLCN1, GACTCAGACTCCGGAGATGA,
olLCN1b, AACTCAGACACCAGAGATGA,
olLCN1c, GACTCAGATCCCGGAGATGA, and
EL5, CCAGGAGGGACCACTACA, specific for the $hOBP_{IIb}$ gene,
EL4, CCGGGACGGACGACTACG, specific for the $hOBP_{IIa}$ gene, and
G3PDH3, CTCATGACCACAGTCCATGC.

Hybridization with oligonucleotides phosphorylated with [γ-^{32}P]ATP using T4 kinase (Applied Biosystems) was performed at 42°C using Hybond N^+ conditions and washing with increasing stringency, with a final wash in 0.1× SSC–0.1% SDS at 48°C for 20 min. The specificity of oligonucleotide binding was checked with samples of digested cosmid DNA (P233G2 for *LCN1* and $hOBP_{IIa}$, P19E7 for *LCN1b* and $hOBP_{IIb}$, P181A9 for *LCN1c*) loaded onto the gel with RT–PCR products.

Genotyping study and linkage analysis

Genotyping was performed by PCR with 100 ng of genomic DNA from the eight reference CEPH families using oligonucleotides oli9, TGTTCGGGAACGCAGCTT, and oli10bis, TGCCGCTGTCCCCACGTCGG. Thermocycling parameters were as follows: an initial cycle at 94°C for 10 min followed by 30 cycles at 94°C for 30 s, 55°C for 30 s and 70°C for 45 s. There was a final elongation step of 10 min at 70°C. PCR products were analysed by electrophoresis in a 3% agarose gel. Genotypes for the chromosome 9 markers were obtained from the chromosome 9 homepage organized by Prof. S. Povey and Dr J. Attwood (http://galton.ucl.ac.uk) and analyses were performed with the linkage package as described previously by Lacazette *et al.* (34). Haplotypes were reconstructed manually according to the previously described recombination events in family 1362 (56).

Protein structure predictions

Multiple alignment of lipocalin protein sequences for which crystallographic structures have been described (57) and for $hOBP_{IIa}$ and $hOBP_{IIb}$ proteins was achieved with ClustalW software (ftp://ftp.infobiogen.fr). This was used to determine putative secondary structures with the DSC program (discrimination of protein secondary structure class) developed by Drs R.D. King and M.J.E. Sternberg (http://bioweb.pasteur.fr/seqanal/interfaces/dsc-simple.html). The secondary structures of proteins corresponding to alternatively spliced forms were assumed to be identical to the classical forms before the frameshift and, after, prediction with single sequences was performed with Predator software (http://pbil.ibcp.fr/cgi-bin).

The tertiary structures of $hOBP_{IIa\alpha}$ and $hOBPII_{IIb\alpha}$ were obtained using the automated Swiss-Model protein modelling service (http://www.expasy.ch/swissmod/SWISS-MODEL.html), after multiple alignment with the sequences of lipocalins of known 3D structure. RBP and β-lactoglobulin (Brookhaven Protein Data Bank accession nos 11BSO, 11B0O, 11BSQ, 11BEB and 11EPA) were used as matrices for $hOBP_{IIa\alpha}$ and $hOBPII_{IIb\alpha}$, respectively. Protein models were viewed with Swiss pdb viewer software.

In situ hybridization

Serial cryostat sections (8 µm thick) were collected on Super-Frost Plus slides (Menzel Glazer, Nemours, France) and stored at –80°C. Antisense and sense RNA probes were transcribed by standard T7 or SP6 polymerase reactions using DIG-11-UTP (Boehringer Mannheim, Meylan, France) after restriction digestion (*Nco*I or *Pst*I) of the $phOBP_{IIa}P2$ cDNA clone (probe length ~150 nucleotides). *Pst*I-digested matrices transcribed with T7 RNA polymerase corresponded to the antisense probe and *Nco*I-digested matrices transcribed with SP6 probe corresponded to the sense probe.

Tissue sections were fixed in 4% paraformaldehyde for 15 min and rinsed for 5 min in cold 2× phosphate-buffered saline.

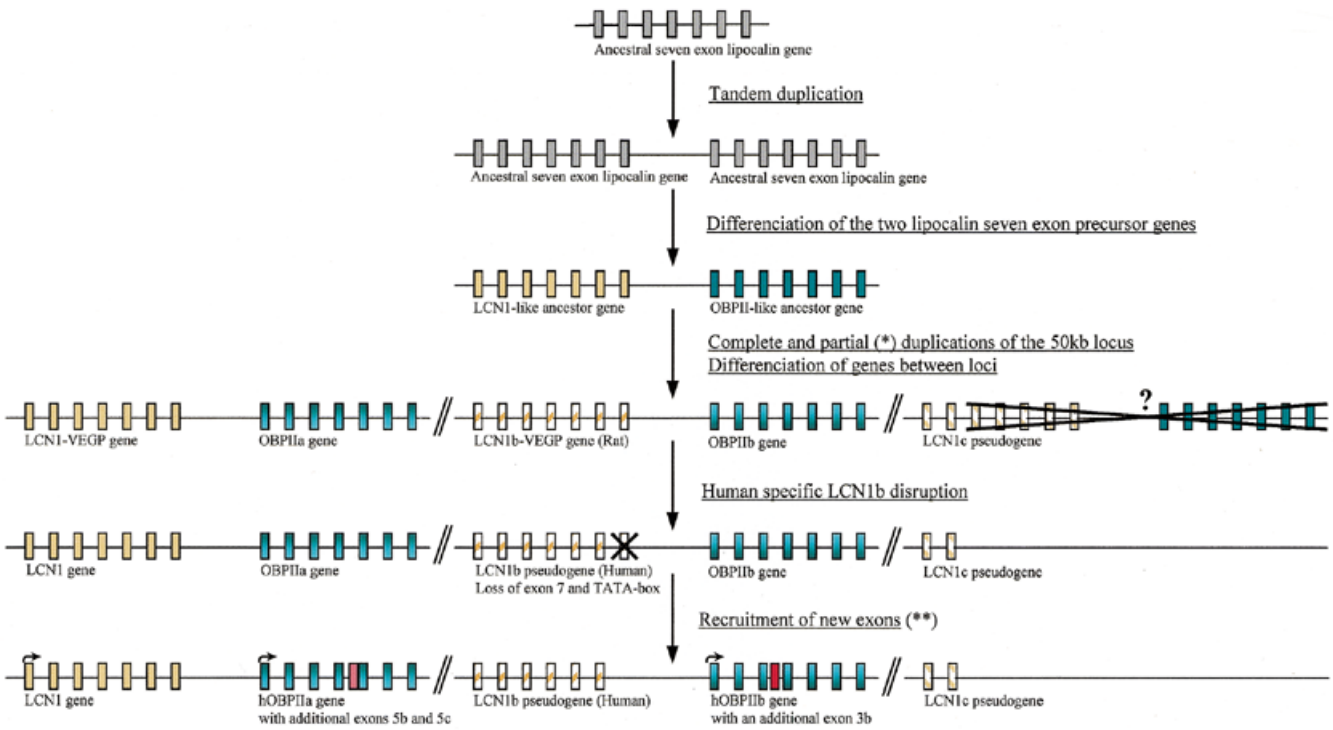


Figure 8. Proposed model for the evolution of the *LCN1-hOBPII* subfamily. Coloured boxes indicate exons, positioned by a line representing genomic DNA. Double slashes between lines indicate that the loci are not consecutive, irrespective of their order. The large cross under the question mark illustrates a partial duplication event, or a complete duplication with a further genomic deletion, for the *LCN1c* locus (choice indicated with an asterisk). The smaller cross indicates disruption of the seventh exon of *LCN1b*, which seems to be specific to humans because numerous Alu repeats are located in the corresponding area and rat has two *VEGP* genes. **Recruitment of new exons may have occurred any time after the locus duplications and may have been sequential.

Tissue sections were acetylated [twice for 5 min each with triethanolamine buffer pH 8.0, containing 0.25% (v/v) acetic anhydride] and incubated at 60°C for 15 min in 1× SSC/50% formamide. Labelled probes were applied to each section in 50 µl of hybridization buffer (50% formamide, 1× Denhardt's solution, 500 µg/ml total tRNA, 10% dextran sulfate, 10 mM dithiothreitol). Sections were covered and incubated in humidified chambers at 50°C overnight. After hybridization, the slides were immersed in washing buffer (50% formamide, 1× SSC) at 55°C for 2 h. They were rinsed twice for 5 min each in 2× SSC at room temperature, treated for 30 min with 10 mg/ml RNase A at 37°C, and immersed for 2 h at 55°C in washing solution (50% formamide, 2× SSC). The slides were then incubated for 15 min in 0.1× SSC at 55°C.

Immunological detection was performed with a sheep anti-DIG–alkaline phosphatase (Fab fragments) antibody according to the Boehringer Mannheim protocol. Sections were examined at various magnifications with an Axiophot (Zeiss, Lyon, France) microscope.

Accession numbers

The following EMBL accession numbers have been attributed to the newly described sequences: AJ251029 for the *hOPB*$_{IIa}$ gene, AJ251021 for hOBP$_{IIa\alpha}$, AJ251022 for hOBP$_{IIa\beta}$, AJ251024 for hOBP$_{IIa\gamma}$, AJ251023 for hOBP$_{IIa\delta}$, AJ251025 for the *hOBP*$_{IIb}$ gene, AJ251026 for hOBP$_{IIb\alpha}$, AJ251027 for hOBP$_{IIb\beta}$, AJ251028 for hOBP$_{IIb\gamma}$ and AJ251020 for the *hOBP*$_{IIb}$ gene minisatellite.

ACKNOWLEDGEMENTS

We are very grateful to Prof. Bernard Dastugue and the U384 INSERM research groups for helpful discussions and free access to their facilities, and to the CHU of Clermont-Ferrand (Recherche Biomédicale et Clinique) for its financial support. We thank Drs Julie Knight, Olga Corti, Jean-Louis Couderc, Kryztov Jagla, Loïc Blanchon and Caroline Conte for critical reading of the manuscript, and Dr Vincent Sapin for helpful discussions concerning *in situ* analysis. We thank Dr Gingrich of the LLNL, who gave us access to the LL09NC01 library, Drs Obermayer and Frischauf for duplicating it at ICRF, and Dr Soularue at Généthon for spotting it. We also thank Drs Howard Cann and Gilles Vergnaud for the CEPH samples, Stéphane Gouttesoulard, Drs Jean-Louis Kemeny, Laurent Gilain, Christophe Guichard and Monique Delatour for experimental support. We would like to thank the 'infobiogen' team with Drs Philippe Dessen and Guy Vayssex for its considerable support. E.L. was awarded a fellowship from the Fédération des Aveugles et Handicapés Visuels de France.

REFERENCES

1. Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.
2. Pelosi, P. (1996) Perireceptor events in olfaction. *J. Neurobiol.*, **30**, 3–19.
3. Pervaiz, S. and Brew, K. (1987) Homology and structure–function correlations between α 1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB J.*, **1**, 209–214.
4. Flower, D.R. (1996) The lipocalin protein family: structure and function. *Biochem. J.*, **318**, 1–14.

5. Malnic, B., Hirono, J., Sato, T. and Buck, L.B. (1999) Combinatorial receptor codes for odors. *Cell*, **96**, 713–723.
6. Pevsner, J., Sklar, P.B. and Snyder, S.H. (1986) Odorant-binding protein: localization to nasal glands and secretions. *Proc. Natl Acad. Sci. USA*, **83**, 4942–4946.
7. Garibotti, M., Navarrini, A., Pisanelli, A.M. and Pelosi, P. (1997) Three odorant-binding proteins from rabbit nasal mucosa. *Chem. Senses*, **22**, 383–390.
8. Felicioli, A., Ganni, M., Garibotti, M. and Pelosi, P. (1993) Multiple types and forms of odorant-binding proteins in the Old-World porcupine *Hystrix cristata*. *Comp. Biochem. Physiol. B*, **105**, 775–784.
9. Dear, T.N., Campbell, K. and Rabbitts, T.H. (1991) Molecular cloning of putative odorant-binding and odorant-metabolizing proteins. *Biochemistry*, **30**, 10376–10382.
10. Pes, D., Mameli, M., Andreini, I., Krieger, J., Weber, M., Breer, H. and Pelosi, P. (1998) Cloning and expression of odorant-binding proteins Ia and Ib from mouse nasal tissue. *Gene*, **212**, 49–55.
11. Marchese, S., Pes, D., Scaloni, A., Carbone V. and Pelosi, P. (1998) Lipocalins of boar salivary glands binding odours and pheromones. *Eur. J. Biochem.*, **252**, 563–568.
12. Redl, B., Holzfeind, P. and Lottspeich, F. (1992) cDNA cloning and sequencing reveals human tear prealbumin to be a member of the lipophilic-ligand carrier protein superfamily. *J. Biol. Chem.*, **267**, 20282–20287.
13. Blaker, M., Kock, K., Ahlers, C., Buck, F. and Schmale, H. (1993) Molecular cloning of human von Ebner's gland protein, a member of the lipocalin superfamily highly expressed in lingual salivary glands. *Biochim. Biophys. Acta*, **1172**, 131–137.
14. Ressot, C., Lassagne, H., Kemeny, J.L. and Gachon, A.M. (1998) Tissue expression of tear lipocalin in humans. *Adv. Exp. Med. Biol.*, **438**, 69–73.
15. Redl, B., Wojnar, P., Ellemunter, H. and Feichtinger, H. (1998) Identification of a lipocalin in mucosal glands of the human tracheobronchial tree and its enhanced secretion in cystic fibrosis. *Lab. Invest.*, **78**, 1121–1129.
16. Fattori, B., Castagna, M., Megna, G., Casani, A. and Pelosi, P. (1998) Immunohistochemical localisation of tear lipocalin in human nasal mucosa. *Rhinology*, **36**, 101–103.
17. Miyawaki, A., Matsushita, F., Ryo, Y. and Mikoshiba, K. (1994) Possible pheromone-carrier function of two lipocalin proteins in the vomeronasal organ. *EMBO J.*, **13**, 5835–5842.
18. Szoka, P.R., Gallagher, J.F. and Held, W.A. (1980) *In vitro* synthesis and characterization of precursors to the mouse major urinary proteins. *J. Biol. Chem.*, **255**, 1367–1373.
19. Henzel, W.J., Rodriguez, H., Singer, A.G., Stults, J.T., Macrides, F., Agosta, W.C. and Niall, H. (1988) The primary structure of aphrodisin. *J. Biol. Chem.*, **263**, 16682–16687.
20. Holzfeind, P., Merschak, P., Rogatsch, H., Culig, Z., Feichtinger, H., Klocker, H. and Redl, B. (1996) Expression of the gene for tear lipocalin/von Ebner's gland protein in human prostate. *FEBS Lett.*, **395**, 95–98.
21. Jamieson, A.C., Vandeyar, M.A., Kang, Y.C., Kinsella, J.E. and Batt, C.A. (1987) Cloning and nucleotide sequence of the bovine β-lactoglobulin gene. *Gene*, **61**, 85–90.
22. Piotte, C.P., Hunter, A.K., Marshall, C.J. and Grigor, M.R. (1998) Phylogenetic analysis of three lipocalin-like proteins present in the milk of *Trichosurus vulpecula* (Phalangeridae, Marsupialia). *J. Mol. Evol.*, **46**, 361–369.
23. Lareyre, J.J., Zheng, W.L., Zhao, G.Q., Kasper, S., Newcomer, M.E., Matusik, R.J., Ong, D.E. and Orgebin-Crist, M.C. (1998) Molecular cloning and hormonal regulation of a murine epididymal retinoic acid-binding protein messenger ribonucleic acid. *Endocrinology*, **139**, 2971–2981.
24. Morel, L., Depeiges, A. and Dufaure, J.P. (1991) Molecular cloning and characterization of a cDNA encoding for the mature form of a specific androgen dependent epididymal protein. *Cell. Mol. Biol.*, **37**, 757–764.
25. Koistinen, H., Koistinen, R., Kamarainen, M., Salo, J. and Seppala, M. (1997) Multiple forms of messenger ribonucleic acid encoding glycodelin in male genital tract. *Lab. Invest.*, **76**, 683–690.
26. Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
27. Igarashi, M., Nagata, A., Toh, H., Urade, Y. and Hayaishi, O. (1992) Structural organization of the gene for prostaglandin D synthase in the rat brain. *Proc. Natl Acad. Sci. USA*, **89**, 5376–5380.
28. Chan, P., Simon-Chazottes, D., Mattei, M.G., Guenet, J.L. and Salier, J.P. (1994) Comparative mapping of lipocalin genes in human and mouse: the four genes for complement C8 γ chain, prostaglandin-D-synthase, oncogene-24p3 and progestagen-associated endometrial protein map to HSA9 and MMU2. *Genomics*, **23**, 145–150.
29. Dewald, G., Cichon, S., Bryant, S.P., Hemmer, S., Nothen, M.M. and Spurr, N.K. (1996) The human complement C8G gene, a member of the lipocalin gene family: polymorphisms and mapping to chromosome 9q34.3. *Ann. Hum. Genet.*, **60**, 281–291.
30. Kock, K., Ahlers, C. and Schmale, H. (1994) Structural organization of the genes for rat von Ebner's gland proteins 1 and 2 reveals their close relationship to lipocalins. *Eur. J. Biochem.*, **221**, 905–916.
31. Logsdon Jr, J.M., Stoltzfus, A. and Doolittle, W.F. (1998) Molecular evolution: recent cases of spliceosomal intron gain? *Curr. Biol.*, **8**, R560–R563.
32. Lassagne, H. and Gachon, A.M. (1993) Cloning of a human lacrimal lipocalin secreted in tears [letter]. *Exp. Eye Res.*, **56**, 605–609.
33. Lassagne, H., Ressot, C., Mattei, M.G. and Gachon, A.M. (1993) Assignment of the human tear lipocalin gene (LCN1) to 9q34 by *in situ* hybridization. *Genomics*, **18**, 160–161.
34. Lacazette, E., Pitiot, G., Jobert, S., Mallet, J. and Gachon, A.M. (1997) Fine genetic mapping of LCN1/D9S1826 within 9q34. *Ann. Hum. Genet.*, **61**, 449–455.
35. Lassagne, H., Nguyen, V.C., Mattei, M.G. and Gachon, A.M. (1995) Assignment of LCN1 to human chromosome 9 is confirmed. *Cytogenet. Cell Genet.*, **71**, 104.
36. Holzfeind, P. and Redl, B. (1994) Structural organization of the gene encoding the human lipocalin tear prealbumin and synthesis of the recombinant protein in *Escherichia coli*. *Gene*, **139**, 177–183.
37. Nahmias, J., Hornigold, N., Fitzgibbon, J., Woodward, K., Pilz, A., Griffin, D., Henske, E.P., Nakamura, Y., Graw, S., Florian, F. *et al*. (1995) Cosmid contigs spanning 9q34 including the candidate region for TSC1. *Eur. J. Hum. Genet.*, **3**, 65–77.
38. van Slegtenhorst, M., Janssen, B., Nellist, M., Ramlakhan, S., Hermans, C., Hesseling, A., van den Ouweland, A., Kwiatkowski, D., Eussen, B., Sampson, J. *et al*. (1995) Cosmid contigs from the tuberous sclerosis candidate region on chromosome 9q34. *Eur. J. Hum. Genet.*, **3**, 78–86.
39. Hornigold, N., van Slegtenhorst, M., Nahmias, J., Ekong, R., Rousseaux, S., Hermans, C., Halley, D., Povey, S. and Wolfe, J. (1997) A 1.7-megabase sequence-ready cosmid contig covering the TSC1 candidate region in 9q34. *Genomics*, **41**, 385–389.
40. Godovac-Zimmermann, J. (1988) The structural motif of β-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules? *Trends Biochem. Sci.*, **13**, 64–66.
41. Bianchet, M.A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S.H., Monaco, H.L. and Amzel, L.M. (1996) The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nature Struct. Biol.*, **3**, 934–939.
42. Woodlee, G.L., Gooley, A.A., Collet, C. and Cooper, D.W. (1993) Origin of late lactation protein from β-lactoglobulin in the tammar wallaby. *J. Hered.*, **84**, 460–465.
43. Clark, A.J., Hickman, J. and Bishop, J. (1984) A 45-kb DNA domain with two divergently orientated genes is the unit of organisation of the murine major urinary protein genes. *EMBO J.*, **3**, 2055–2064.
44. Dente, L., Pizza, M.G., Metspalu, A. and Cortese, R. (1987) Structure and expression of the genes coding for human α 1-acid glycoprotein. *EMBO J.*, **6**, 2289–2296.
45. Nicholas, K.R., Messer, M., Elliott, C., Maher, F. and Shaw, D.C. (1987) A novel whey protein synthesized only in late lactation by the mammary gland from the tammar (*Macropus eugenii*). *Biochem. J.*, **241**, 899–904.
46. Lobel, D., Marchese, S., Krieger, J., Pelosi, P. and Breer, H. (1998) Subtypes of odorant-binding proteins—heterologous expression and ligand binding. *Eur. J. Biochem.*, **254**, 318–324.
47. Glasgow, B.J., Abduragimov, A.R., Yusifov, T.N., Gasymov, O.K., Horwitz, J., Hubbell, W.L. and Faull, K.F. (1998) A conserved disulfide motif in human tear lipocalins influences ligand binding. *Biochemistry*, **37**, 2215–2225.
48. Boudjelal, M., Sivaprasadarao, A. and Findlay, J.B. (1996) Membrane receptor for odour-binding proteins. *Biochem. J.*, **317**, 23–27.
49. Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W.D., Huber, R. and Bode, W. (1997) Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. *Proc. Natl Acad. Sci. USA*, **94**, 11845–11850.
50. Morrow, D.M., Xiong, N., Getty, R.R., Ratajczak, M.Z., Morgan, D., Seppala, M., Riittinen, L., Gewirtz, A.M. and Tykocinski, M.L. (1994)

Hematopoietic placental protein 14. An immunosuppressive factor in cells of the megakaryocytic lineage. *Am. J. Pathol.*, **145**, 1485–1495.

51. Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P. and Tiver, K.K. (1986) Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kilodaltons synthesized by the rat epididymis. *J. Biol. Chem.*, **261**, 4956–4961.
52. Schambony, A., Gentzel, M., Wolfes, H., Raida, M., Neumann, U. and Topfer-Petersen, E. (1998) Equine CRISP-3: primary structure and expression in the male genital tract. *Biochim. Biophys. Acta*, **1387**, 206–216.
53. Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J. *et al.* (1992) Expression of members of the putative olfactory receptor gene family in mammalian germ cells. *Nature*, **355**, 453–455.
54. Vanderhaeghen, P., Schurmans, S., Vassart, G. and Parmentier, M. (1997) Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species. *Genomics*, **39**, 239–246.
55. Bishop, R.E., Penfold, S.S., Frost, L.S., Holtje, J.V. and Weiner, J.H. (1995) Stationary phase expression of a novel *Escherichia coli* outer membrane lipoprotein and its relationship with mammalian apolipoprotein D. Implications for the origin of lipocalins. *J. Biol. Chem.*, **270**, 23097–23103.
56. Attwood, J., Chiano, M., Collins, A., Donis-Keller, H., Dracopoli, N., Fountain, J., Falk, C., Goudie, D., Gusella, J., Haines, J. *et al.* (1994) CEPH consortium map of chromosome 9. *Genomics*, **19**, 203–214.
57. Spinelli, S., Ramoni, R., Grolli, S., Bonicel, J., Cambillau, C. and Tegoni, M. (1998) The structure of the monomeric porcine odorant binding protein sheds light on the domain swapping mechanism. *Biochemistry*, **37**, 7913–7918.