

RESEARCH ARTICLE

Open Access



# The human olfactory transcriptome

Tsviya Olender<sup>1\*</sup>, Ifat Keydar<sup>1</sup>, Jayant M. Pinto<sup>2</sup>, Pavlo Tatarkyy<sup>1</sup>, Anna Alkelai<sup>1</sup>, Ming-Shan Chien<sup>3</sup>, Simon Fishilevich<sup>1</sup>, Diego Restrepo<sup>4</sup>, Hiroaki Matsunami<sup>3</sup>, Yoav Gilad<sup>5</sup> and Doron Lancet<sup>1</sup>

## Abstract

**Background:** Olfaction is a versatile sensory mechanism for detecting thousands of volatile odorants. Although molecular basis of odorant signaling is relatively well understood considerable gaps remain in the complete charting of all relevant gene products. To address this challenge, we applied RNAseq to four well-characterized human olfactory epithelial samples and compared the results to novel and published mouse olfactory epithelium as well as 16 human control tissues.

**Results:** We identified 194 non-olfactory receptor (OR) genes that are overexpressed in human olfactory tissues vs. controls. The highest overexpression is seen for lipocalins and bactericidal/permeability-increasing (BPI)-fold proteins, which in other species include secreted odorant carriers. Mouse-human discordance in orthologous lipocalin expression suggests different mammalian evolutionary paths in this family.

Of the overexpressed genes 36 have documented olfactory function while for 158 there is little or no previous such functional evidence. The latter group includes GPCRs, neuropeptides, solute carriers, transcription factors and biotransformation enzymes. Many of them may be indirectly implicated in sensory function, and ~70 % are over expressed also in mouse olfactory epithelium, corroborating their olfactory role.

Nearly 90 % of the intact OR repertoire, and ~60 % of the OR pseudogenes are expressed in the olfactory epithelium, with the latter showing a 3-fold lower expression. ORs transcription levels show a 1000-fold inter-paralog variation, as well as significant inter-individual differences. We assembled 160 transcripts representing 100 intact OR genes. These include 1–4 short 5' non-coding exons with considerable alternative splicing and long last exons that contain the coding region and 3' untranslated region of highly variable length. Notably, we identified 10 ORs with an intact open reading frame but with seemingly non-functional transcripts, suggesting a yet unreported OR pseudogenization mechanism. Analysis of the OR upstream regions indicated an enrichment of the homeobox family transcription factor binding sites and a consensus localization of a specific transcription factor binding site subfamily (Olf/EBF).

**Conclusions:** We provide an overview of expression levels of ORs and auxiliary genes in human olfactory epithelium. This forms a transcriptomic view of the entire OR repertoire, and reveals a large number of over-expressed uncharacterized human non-receptor genes, providing a platform for future discovery.

**Keywords:** Olfaction, RNAseq, Olfactory epithelium, Olfactory receptor

## Background

Olfaction, the sense of smell, is a versatile and sensitive mechanism for detecting and discriminating thousands of volatile odorants. Olfactory recognition is mediated by a large repertoire of olfactory receptors (ORs), which activate a G-protein-mediated transduction cascade [1–4]. The ORs are expressed on the ciliated dendrites of olfactory sensory neurons located in the olfactory epithelium.

Each sensory neurons expresses a single allele of a single OR gene locus, to ensure a distinct pattern of neuronal activation for every odorant [1, 2].

Olfactory epithelium, the tissue analyzed here, is heterogonous, containing besides the sensory neurons also epithelial supporting cells and progenitor basal cells, as well as sub-epithelial Bowman's glands cells that secrete the mucus within which olfactory cilia reside, microvillar cells, and fingerlike microvilli cells [5]. While in mouse this tissue is readily available, the human counterpart is harder to obtain, due to difficulties in dissection and in defining the exact anatomically boundaries [6]. This

\* Correspondence: [tsviya.olender@weizmann.ac.il](mailto:tsviya.olender@weizmann.ac.il)

<sup>1</sup>Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

Full list of author information is available at the end of the article



explains the relative lack of transcriptome information about human olfactory genes. Overcoming these difficulties, we provide here a whole genome expression view of the human olfactory tissue.

Olfactory transduction can be divided into ligand binding, signal generation and signal termination. The end result is triggering of action potentials conducted along the axon to the olfactory bulb. A large number of proteins take part in such processes, as well as in the development and maintenance of the relevant cellular components [7–9]. Such gene products have been termed “auxiliary”, as portrayed in a digital compendium, GOSdb database (<http://genome.weizmann.ac.il/GOSdb/>, [10]).

Not all olfactory auxiliary genes have been identified, and most of them have never been studied in humans. Physiological differences among mammalian species may be accompanied with differences at the signal transduction level as well. For example guanylate cyclase 2D (*Gucy2d*, *GC-D*) gene, which is expressed in specific subset of OSNs and involved in CO<sub>2</sub> detection [4, 11], was shown to be a pseudogene in humans, and signaling through this system appears to have been lost during primate evolution [12]. Here, we employ a broad transcriptome analysis to help fill some of these knowledge gaps, pertaining to olfactory auxiliary genes.

The human genome contains 857 OR genes, of which 391 are intact and 466 are pseudogenes with disrupting mutations in the open reading frame [13]. The repertoire of OR coding regions was deciphered mainly by computational genome data-mining [10, 14–16]. This information is reflected in genomic databases of both human and mouse, where most OR genes are portrayed with a partial gene structure that depicts only the coding region. The general reported ORs gene structure shows a single exon encompassing both a ~960 bp coding region and a 3' untranslated region (3'UTR), with additional 5' UTR short exons separated by long introns [17–20]. In mouse, cDNA sequencing and RNAseq, including single cell RNAseq of olfactory sensory neurons, have provided considerable specific information on OR gene structure [18, 20–25]. In contrast, information of the human OR gene structure is available only for a limited number of genes. Here we considerably expand this information.

The current study describes the transcriptome analysis of four human olfactory epithelial samples. We identified a set of 196 olfactory over expressed genes, composed of genes with known olfactory functions as well as novel olfactory candidates. This provides clues to a large number of uncharted genes which might have a role in the olfactory epithelium, including chemosensory function.

For OR genes we observed large variation in the expression intensity, as well as inter-individual differences in expression. We were also able to assemble the complete

gene structure of 100 OR intact genes, providing a fresh general view of the encoded human OR transcripts.

## Results

### Differential expression

We obtained epithelial samples from the nasal cavity of 14 human subjects and three autopsies. Four of them were identified by gene marker analysis as relatively high quality olfactory epithelium, and were selected for RNA sequencing (see Methods, Additional file 1: Table S1, Figure S1, Additional file 2: Table S2). The olfactory epithelial tissues portrayed a unique overall gene expression signature as compared to 16 control tissues from the illumina Body Map project (Fig. 1). For comparisons, we analyzed mouse olfactory epithelial RNAseq data from different sources (Additional file 2: Table S2), including a preparation of isolated sensory neurons. The latter provided specific information about the role of certain genes in this class of neuronal cells within olfactory epithelium. Although the mouse RNAseq came from different mouse strains, sex and age, the correlation values between the different strains as well as within strains was high and significant (MOE1-MOE2 0.945, MOE1-MOE3 0.947, MOE2-MOE3 0.978, Pearson).

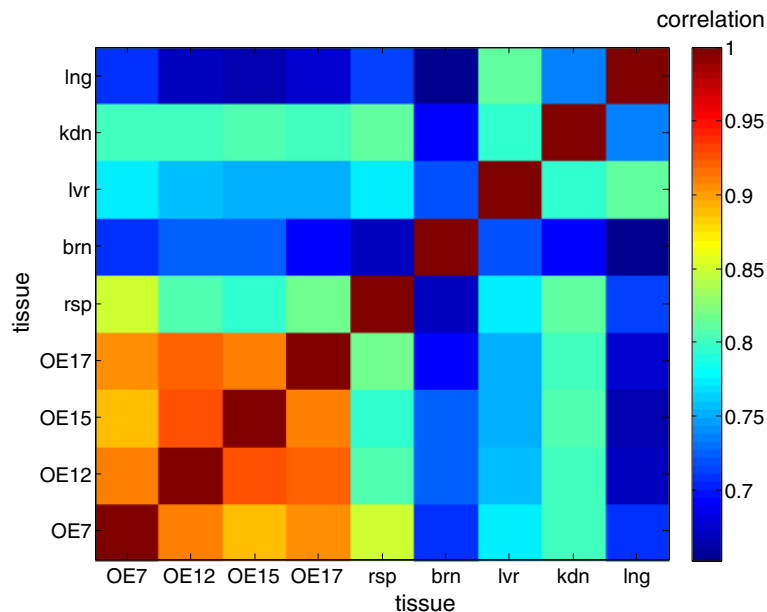
We next identified a set of 196 non-olfactory receptor genes that were over-expressed (>X6) in human olfactory epithelium relative to the illumina BodyMap control tissues but not over-expressed in respiratory epithelium relative to the same controls (Fig. 2a and Additional file 3: Table S3). Using database and literature searches we were able to discriminate between genes with previous evidence of an olfactory role (36 genes, class A) and genes with little or no previous knowledge on such functional involvement (158 genes, class B). Class A genes include the expected well-known olfactory signal transduction genes such as *OMP*, *CNGA2*, *CNGA4*, *GNG13*, *ANO2*, *RTP1*, and *RTP2*. Additional genes in class A include bio-transformation enzymes such as *UGT2A2*, and more (Additional file 3: Table S4). Notably, all class A genes are over expressed in mouse olfactory epithelium, except *NOS2*.

### New olfactory expression vistas

We subsequently examined the 158 class B genes and identified 109 genes that were overexpressed in human as well as mouse olfactory epithelium, highlighting their putative new olfactory role (Additional file 3: Table S5). These olfactory epithelium-enriched genes are described below, grouped by functional subgroups (Fig. 2b).

### Non-olfactory GPCRs

We identified three neuropeptide receptors, *NPBWR1*, *NPBWR2* and *NPFER1* (Additional file 1: Figure S2). *NPBWR1* and *NPBWR2* have a role in regulating feeding



**Fig. 1** Expression correlation of olfactory and other tissues. Pearson correlation of human olfactory epithelium and a selected set of control tissues. Correlation was calculated for each tissue pair based on vectors with the logarithmic FPKM value of 16,288 genes. Correlation values are on average  $0.91 \pm 0.0$  among the olfactory epithelium tissues and  $0.74 \pm 0.1$  between the olfactory epithelium tissues to the controls. Tissue abbreviations: OE, human olfactory epithelium; rsp, respiratory; brn, brain; lvr, liver; kdn, kidney; lng, lung

behavior, energy homeostasis, neuroendocrine function, and modulating inflammatory pain [26] as well as in the regulation of emotion-related responses that affect autonomic functions [27]. *NPFFR1* has a role in GnRH signaling of the reproductive axis [28] *NPBWR1* and *NPBWR2* are receptors of the *NPB* and *NPW* neuropeptides. The peptide *NPB*, but not *NPW* shows sufficient expression (1.6 FPKM) and overexpression (X2.77) in human olfactory epithelium, to warrant notice.

#### Neuropeptides

This subgroup includes four neuropeptides: *UCN3* and *PTH2* (class B genes), and two more well-known (class A genes, *AVP*, *TAC4*) (Additional file 1: Figure S3). Two of these neuropeptides, *AVP* and *UCN3*, were respectively linked to the processing of social stimuli in sensory regions of the brain [29] and to social discrimination abilities via corticotropin-releasing hormone receptor type 2, being localized in nuclei functionally connected to the accessory olfactory system [30]. We note that *PTH2* is highly overexpressed in isolated olfactory sensory neurons (Additional file 3: Table S3 and Table S5), suggesting a yet undefined role in these cells.

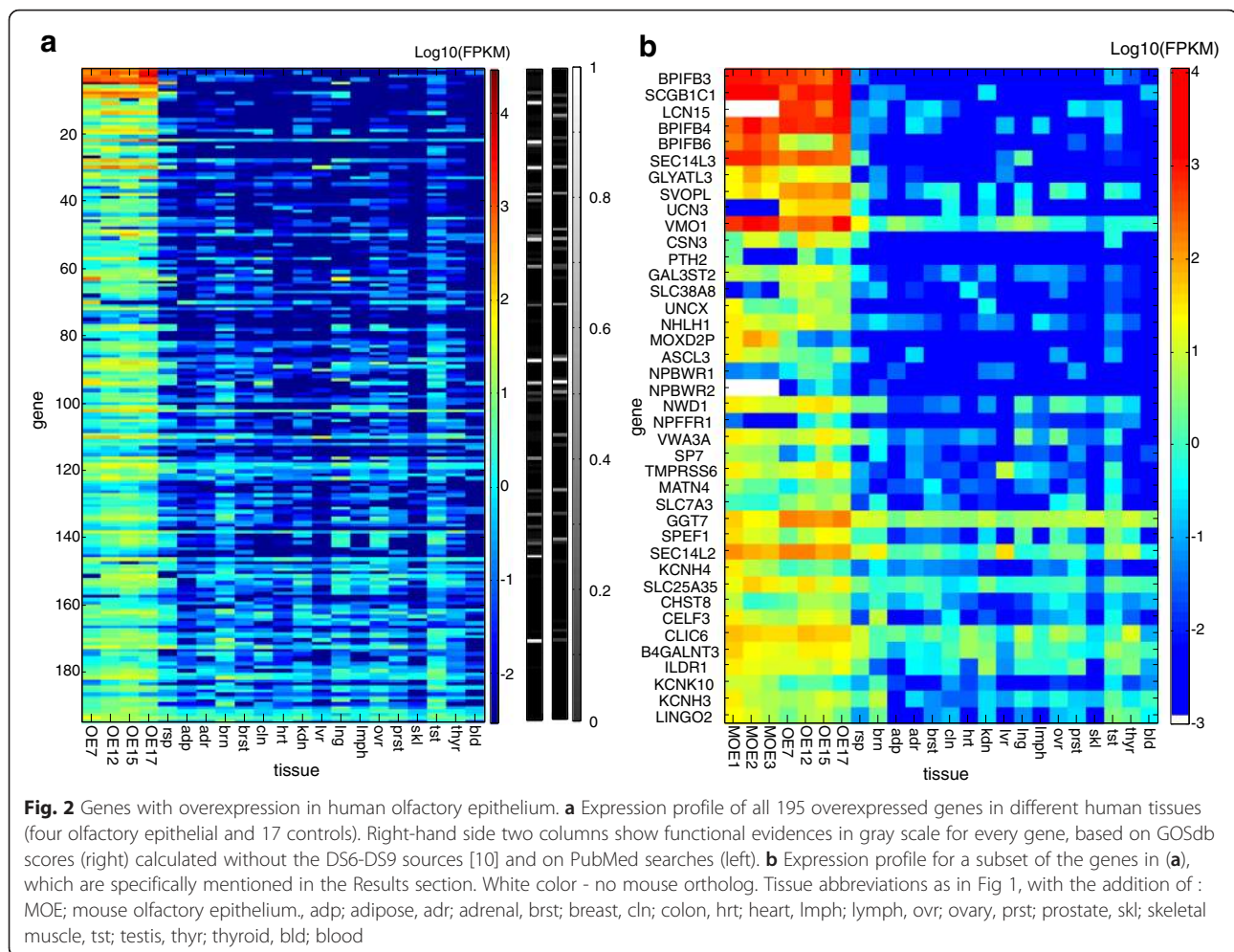
#### Ion channels

Three voltage gated potassium channels (*KCNK10*, *KCNH3* and *KCNH4*) and a chloride channel (*CLIC6*) appear among class B genes (Additional file 1: Figure S4). *KCNH* channels are voltage-gated potassium channels

with roles in cardiac repolarization, cellular proliferation and tumor growth [31]. All three genes show strong overexpression in isolated olfactory sensory neurons (Additional file 3: Table S3 and Table S5). Future scrutiny could uncover a possible role in olfactory epithelial differentiation or neurogenesis, as suggested [32]. *CLIC6* (X16.5 overexpressed) is a member of the chloride intracellular channel family, which functions as monomeric soluble proteins and as integral membrane chloride ion channels. In the soluble form they adopt a glutathione S-transferase (GST) fold, with an enzymatic activity [33]. In line with this, our data are consistent with function in olfactory epithelial cells other than the sensory neurons (Additional file 3: Table S3 and Table S5).

#### Solute carriers

Four proteins of this group are identified: *SVOPL*, member of the SLC22 family, *SLC25A35*, *SLC38A8* and *SLC7A3* (Additional file 1: Figure S5). *SVOPL* is a putative synaptic vesicle glycoprotein and its affiliation with the SLC22 family suggests a role as organic ion transporters. Enhanced expression of *SVOPL* in the olfactory bulb and cerebral cortex has been reported [34], consistent with an olfactory role. Yet, based on the analysis of isolated olfactory sensory neurons (Additional file 3: Table S3 and Table S5), these genes may have a role in non-sensory cells of the olfactory epithelium. *SLC38A8* functions as a neuronal transporter with a broad amino acid transport profile and



was suggested to have a key role in the glutamine/glutamate (GABA) brain cycle [35].

### Transcription factors

Thirteen transcription factors, with little or no known olfactory involvement are overexpressed (Additional file 1: Figure S6). Of these we note the gene *ASCL3*, member of the achaete-scute complex (ASC) family which has a role in cell fate determination and differentiation of numerous tissues, including neuronal tissues [36]. In drosophila achaete-scute gene complex (AS-C) is a key component in developing of the macrochaetes sensory organs [37]. The paralog *ASCL1* is required for early development of olfactory neuron [38]. Three additional transcription factors are *UNCX*, suggested to participate in the regulation of neural progenitor cells proliferation and neuronal survival in the olfactory epithelium [39], and *SP7* which plays a role in the olfactory glomerular layer [5] and *NHLH1*, a neurogenesis transcription factor whose expression in mouse olfactory epithelium and vomeronasal organ during development has been reported [40]. The expression data provide further support for such

roles. A specific mature olfactory sensory neuronal function is suggested only for one of the above, *UNCX* (Additional file 3: Table S3 and Table S5).

### Other genes

Other genes for which we suggest a novel olfactory role include: 1) matrilin 4 (*MATN4*), a member of the von Willebrand factor A domain-containing protein family, involved in the formation of filamentous networks in extracellular matrix. In zebrafish *MATN4* expression was significantly increased following exposure of olfactory epithelium to the odorant phenylethyl alcohol in a potential context of memory formation [41]; 2) Leucine Rich Repeat and Ig Domain Containing 2 (*LINGO2*), reported to be expressed during mouse embryogenesis in a population of cells lying adjacent to the epithelial lining of the olfactory pit [42], and to be involved in the development of the olfactory pathway of mouse and zebrafish embryos [42]. 3) *SPEF1*, a microtubule-associated protein, that plays a role in the structural integrity of auditory sensory epithelium [43]; 4) *VMO1* a protein of the outer layer of the vitelline membrane of eggs which has an essential

function in the antimicrobial barrier in avian eggs [44] and was suggested to function in tear integrity [45, 46]; 5) *EFR3B*, a palmitoylated plasma membrane protein, which is responsible for maintaining an active pool of the *PI4KA* enzyme at the plasma membrane. *EFR3B* was recently shown to function also as a direct regulator of GPCRs [47]. The high overexpression of three of the above genes, *LINGO2*, *SPEF1* and *EFR3B* in isolated sensory neurons is noteworthy (Additional file 3: Table S3 and Table S5).

Other than that we identified several RNA genes, linc RNA genes and antisense genes which might have a role in the regulation of the olfactory system, three transmembrane protein, two Squalene Transfer Proteins (*SEC14L2* and *SEC14L3*) and 24 other secreted proteins. Among the secreted proteins especially high over expression is observed for *SCGB1C1*.

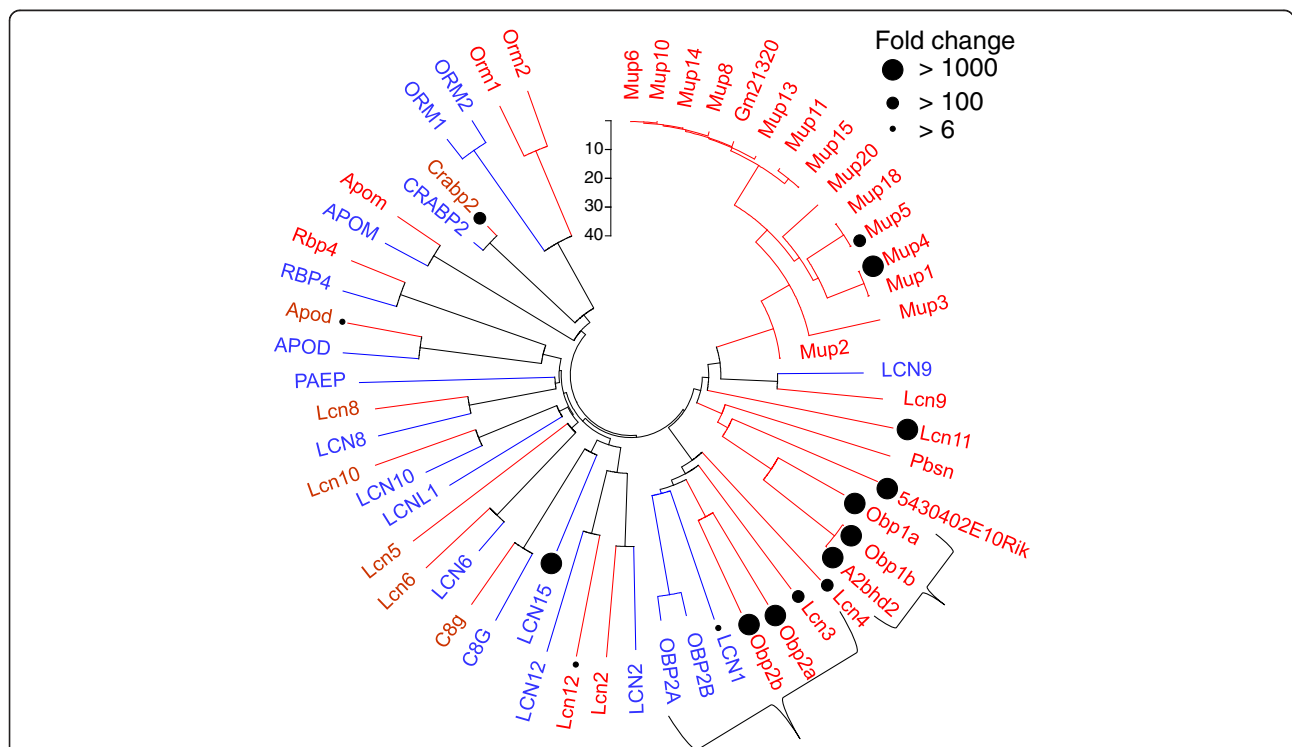
**Odorant binding proteins (OBPs)**

Our data indicate a complex situation with respect to olfactory epithelial overexpressed genes that might harbor odorant binding protein (OBP) function. In mouse 4 paralogs genes, *Obp1a*, *Obp1b*, *Obp2a* and *Obp2b* are highly overexpressed (Additional file 4: Table S6). These are “classical” OBPs, as indicated by their symbols, and indeed all four are highly expressed in the mouse sensory organ.

However, in human only two of the four appear to have orthologs bearing identical symbols (*OBP2A*, *OBP2B*, Additional file 1: Figure S6). Surprisingly, these two genes do not show any enhanced expression in the human sensory tissue (under-expression of 0.1X and 0.32X respectively and a very low absolute values of 0.12 and 0.08 FPKM respectively). In fact *OBP2A*, *OBP2B* are overexpressed in testis and ovary (Additional file 1: Figure S8). This may indicate a case of functional misidentification.

The finding that human OBPs are likely to have no olfactory role is further corroborated by the recent *OBP2B* crystal structure, showing a structural features different from those of other mammalian OBPs, including a potentially reactive cysteine side chain within the binding pocket, which is most similar to human tear lipocalin [48].

Alternative human functional OBPs may be gleaned in the broad scope orthology dendrogram of Fig. 3, showing sequence relationships as well as overexpression traits. A second gene sub-family - lipocalins with explicit symbol prefix LCN - appears to be relevant to human olfaction. In human there are 9 LCN genes while in mouse there are 10 such genes. In human, only two enhanced sensory organ expression: *LCN1* and *LCN15*, the latter showing especially high overexpression. In contrast, a quite different subgroup of LCN genes shows



**Fig. 3** Phylogenetic analysis of the lipocalin family. The evolutionary branches of OBP1 and OBP2 are marked by curly brackets. Black circles indicate the fold change of a gene in olfactory epithelium relative to controls. The tree was constructed with MEGA6 [88] using the NJ method [87]. Red- mouse; blue - human

mouse sensory prominence: *Lcn3*, *Lcn4*, *Lcn10*, *Lcn11*, with the latter being the strongest.

In addition, several BPI fold containing proteins which were previously suggested to function as odorant binding proteins [49] are extremely overexpressed ( $\times 10,000$  or more) in human olfactory epithelium. Such striking result lends credence to a possible olfactory role, possibly in odorant carrying or removal. The null expression in the sensory neurons (Additional file 3: Table S5) is consistent with expression in non-neuronal secretory cells of the olfactory epithelium or sub-epithelium.

### Biotransformation enzymes

Biotransformation enzymes are involved in xenobiotic modification and clearance [50, 51]. Some of these enzymes have previously been reported to be expressed in the olfactory epithelium and suggested to play a role in odorant modification [52–54]. These include two cytochrome P450s (*CYP2G1*, *CYP2A13*) and one UDP glucuronosyl transferase (*UGT2A2*), which are indeed overexpressed in olfactory epithelium (Additional file 1: Figure S9, Additional file 3: Table S5). The mouse ortholog of *CYP2G1*, *Cyp2g1*, is a known olfactory-enriched protein [53, 55, 56] suggested to be involved in clearance of odorous compounds [52]. Another seven biotransformation enzymes show high ( $>\times 10$ ) olfactory over expression (Additional file 1: Figure S9). Among these, for the aldehyde dehydrogenase *ALDH3A1* and the glutathione peroxidase *GPX6* there are previous indications for olfactory involvement [50, 57]. Four others, a glycine acyl transferase (*GLYATL3*), a galactose sulfotransferase (*GAL3ST2*), a gamma glutamyltransferase (*GGT7*) and a carbohydrate sulfotransferase (*CHST8*) have no reported olfactory role, and their specific olfactory epithelial expression is worthy of future scrutiny. Of note is that the only one among the abovementioned genes that has a strong enrichment in isolated sensory neurons is *CHST8* (Additional file 3: Table S3 and Table S5).

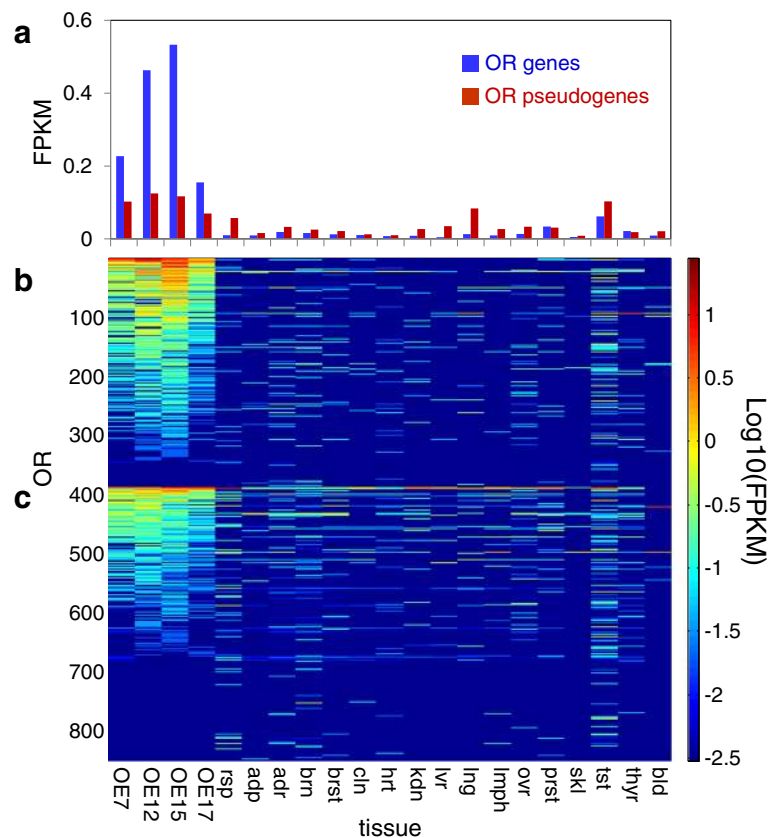
Another biotransformation enzyme overexpressed in human olfactory epithelium is the dopamine beta-hydroxylase-like monooxygenase *MOXD2*. This enzyme harbors frequent loss-of-function mutations in some apes, toothed whales and baleen whales [58]. In human the gene inactivated due to a 2 exons deletion [59] that is fixated in the population, as indicated by 1000 genomes and DGV scrutiny. This pseudogenization is possibly connected to the impaired olfactory faculties in apes and monodontidae [58] as well as in human [60].

### Olfactory receptors

We examined the expression profile of all mappable OR genes in human olfactory epithelium (Fig. 4, Additional file 5: Table S7). Using a cutoff of with FPKM  $\geq 0.01$  [61] we observed that 88.6 % of the intact OR genes were

expressed, while a much lower percentage (61.2 %) of the OR pseudogenes were expressed in at least one of the tissue samples. Further, intact OR genes had a significantly higher average expression level ( $0.35 \pm 1.08$  FPKM) as compared to OR pseudogenes ( $0.09 \pm 0.33$  FPKM,  $P = 3.7 \times 10^{-26}$ , Kolmogorov-Smirnoff test, Fig. 4). Moreover, we found a significant correlation between the predicted probability of the OR to encode a functional protein, computed by the CORP score [62] and its expression level ( $P = 3.62 \times 10^{-7}$ , Kolmogorov-Smirnoff test, Additional file 1: Figure S10). Interestingly, in control non-olfactory tissues the relationship is inverted, whereby OR pseudogenes have a higher average expression than intact genes (Fig. 4). We note that among non-olfactory tissues, the highest expression of both OR genes and pseudogenes is in testis, confirming a previous report [61].

We asked about trends in expression levels across the OR repertoire. Our data clearly indicate that expression levels of different OR loci are highly heterogeneous. While for intact human OR genes the median is about 0.1 FPKM, the top OR is expressed at levels 100 fold higher, and the overall span is around three orders of magnitude (Fig. 5a). Human OR pseudogenes show a similar trend, suggesting that protein functionality may not be underlying correlate. It appears that variations in the effectiveness of transcription regulation may be at work. In mouse, while the overall expression is  $\sim 5$ -10 fold higher across the board (probably due to tissue purity differences), the inter-OR heterogeneity is less pronounced. Interestingly, the discrepancy in expression levels between intact and pseudogenized loci is much higher in mouse, possibly because of the recently of pseudogenization in a large majority human OR genes [60]. Of note is that both species show a skew in the inter-OR expression variation, and that this is much more pronounced in human. While in mouse  $\sim 18$  % of the ORs contribute 50 % of to the cumulative OR expression, in human the value is  $\sim 5$  % (Fig. 5b). This may be rephrased as indicating that in the sensory organ, the effective repertoire, in particularly in human, is only a small fraction of the nominal repertoire. Interestingly, the position of a given OR on the expression intensity scale portrays no ortholog-pair correlation (Additional file 1: Figure S11). It might be argued that the above observations could be misleading due to inaccurate calling of genes and pseudogenes in human. This is because of the fact that individual human genomes contain a high number of deleterious variations that turn intact genes into pseudogenes in some individuals (segregating pseudogenes) [10]. We therefore redrew Figs. 4, 5 without all 282 OR loci reported to harbor segregating pseudogenes. The assignment of loci to segregating pseudogene status was done base on our previous data [10] that integrated 13 different resources, including the 1000 genomes project as well as exome



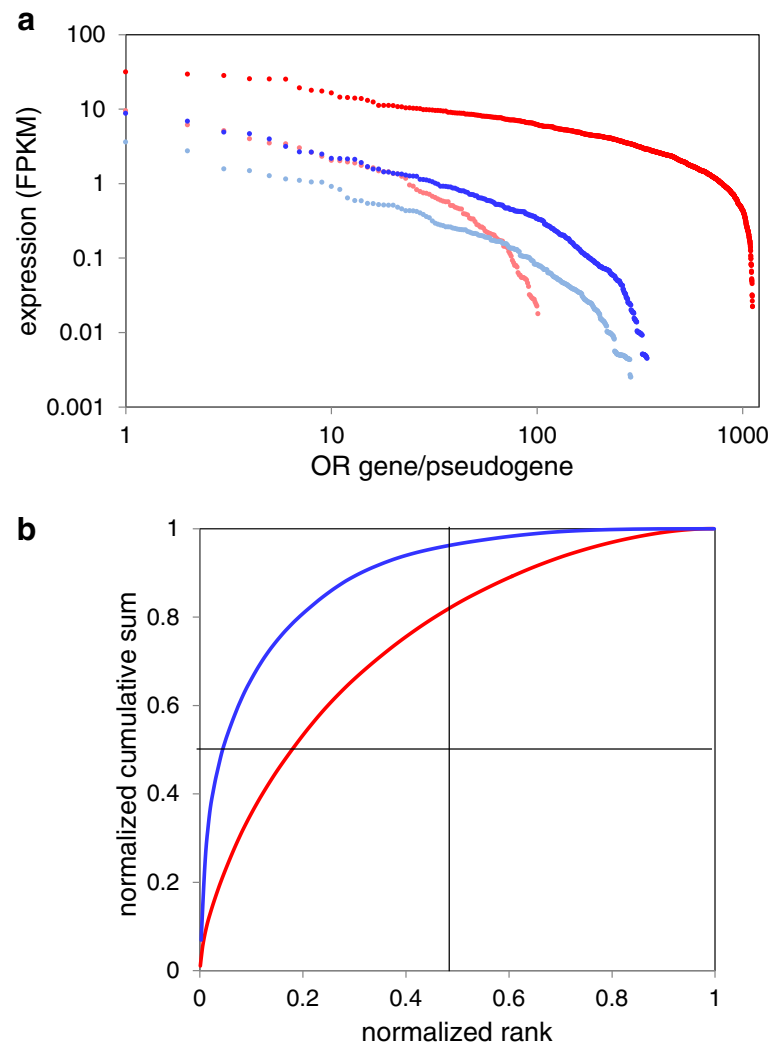
**Fig. 4** Expression profiles of OR genes. **a** the average expression of OR genes and pseudogenes per enumerated tissue. **b, c** respective expression profile for the same tissues for intact and pseudogenized genes. Tissue abbreviations as in Fig 2

sequences of over 1000 individuals, including deleterious SNPs, indels and CNVs. This process thus captured all segregating pseudogenes with an allele frequency  $>0.001$  in the human population. The probability that an individual that underwent transcriptome analysis has rare private deleterious OR mutations is further diminished based on our statistics [10] that a typical individual genome has on average only 46 affected OR loci out of the population's 282. The results shown in Additional file 1: Figure S12 suggest a very small effect as compared to Fig. 4. Likewise, while the results relevant to Fig. 5a seen in Additional file 1: Figure S13A show somewhat different trends due to the different gene count, the normalized curves relevant to Fig. 5b (Additional file 1: Figure S13B) are nearly indistinguishable with and without omission of segregating pseudogenes.

Interestingly, within the mouse OR genes we find that class II ORs have significantly higher expression level than class I ORs ( $P = 7.2 \times 10^{-15}$ , Chi Square), consistent with the recent findings of [63]. Class II ORs have a higher expression also within the human OR expression data, although the difference was not significant (0.06, Chi Square), perhaps due to the low count of human class I receptors.

#### Inter-individual patterns of OR expression

The foregoing analyses were performed on values averaged across the four human samples employed. We asked whether there was inter-individual variations in the expression of OR genes. Such across-sample variations are indeed visible in Fig. 4. For further validation, we computed the six pairwise correlation values for the four human olfactory epithelial samples and indeed observed lower correlation values for OR genes as compared to the complete gene set (Fig. 6a, Additional file 1: Figure S14). However such results could stem from noise related to the fact that ORs typically have low expression values. To address this, we analyzed 10,000 random subsets of a similar size to that of the OR set which obey the same FPKM distribution as the OR genes. For OE12, OE15 and OE17 this simulation demonstrated that the correlation values for the OR gene set lie completely outside the distribution for the 10,000 control gene sets (Fig. 6b). Thus for these tissues the expression level of the ORs is significantly more variable than that of other genes. When OE7 was compared to each of the other three tissues, no significant inter-individual differences were observed (Fig. 6b legend). This is likely due to the 5 fold lower reads for this tissue.



**Fig. 5** Rank plots for the expression values of human and mouse ORs. **a** Ranked expression values. Red and pink respectively represent intact and pseudogenized mouse OR genes; Blue and light blue respectively represent intact and pseudogenized human OR genes. **b** Normalized cumulative expression values, colors as in (a)

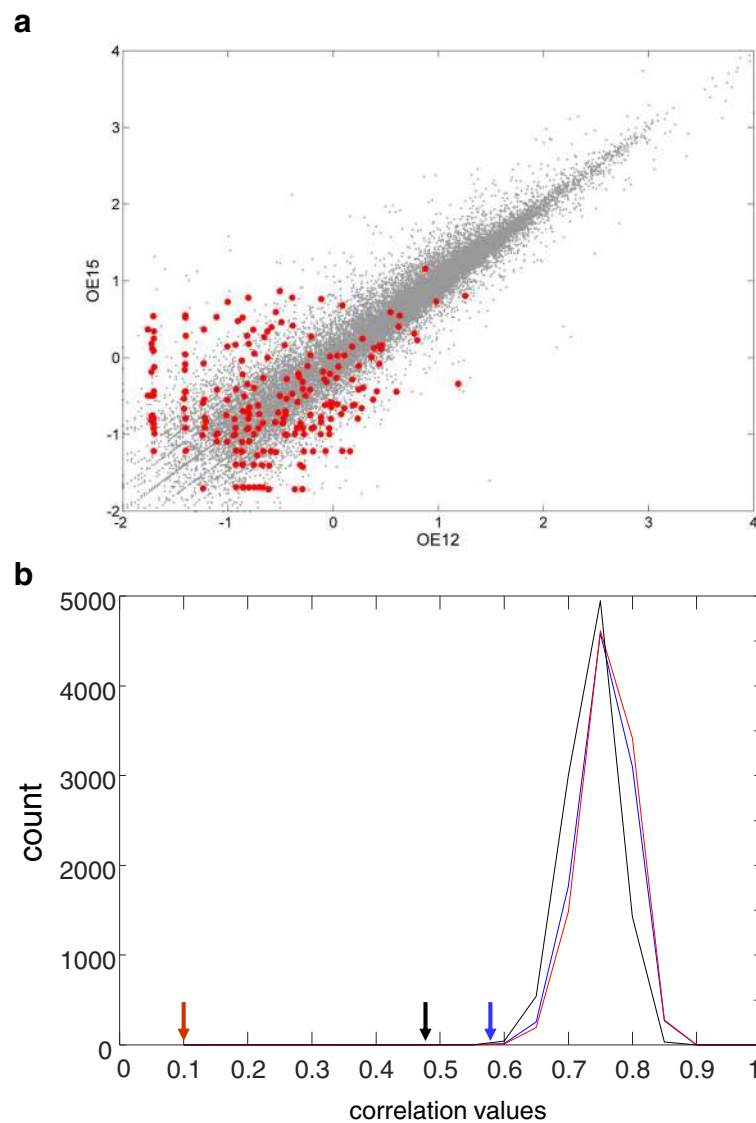
A potential weakness of the conclusions regarding inter-individual differences in OR expression patterns is that some such difference could arise due to differences in the counts of intact OR alleles in different individuals [10]. To address this, we redrew Fig. 6 and Additional file 1: Figure S14 in a version that excludes all gene loci previously reported [10] to harbor segregating pseudogenes (Additional file 1: Figures S15-16). The results were nearly identical to the original.

#### Genomic structure of OR transcripts

We used uniquely-mapped RNAseq reads from all human olfactory epithelial samples to assemble transcripts for intact OR genes. After a curation process we obtained 311 transcripts representing 210 intact OR genes. Of these, 120 transcripts with expression level  $<1.0$  FPKM were

excluded to avoid inaccuracies in transcript assembly, and the remaining 163 transcripts (encoded in 100 genes) were further analyzed (Fig. 7 and Additional file 1: Figure S17, Additional file 6: Table S8). The OR transcripts obtained were  $4097 \pm 2053$  bp long and spanned genomic lengths of  $8103 \pm 3464$  bp. The coding exon contains an open reading frame of  $940 \pm 16$  bp and 3' UTR of  $2777 \pm 2047$  bp (Fig. 8). The 5' UTR is 389 bp long, and contains 0–3 non-coding exons (Fig. 7). The number of splice variants per gene varies from 1 to 5 (Fig. 7). Of note are 10 OR genes with a seemingly non-functional transcript that skips the initiating methionine (Fig. 7), suggesting a yet unreported OR pseudogenization mechanism. Eight of these aberrant transcripts show the co-existence of both functional and non-functional transcripts, indicating the presence of two different alleles.





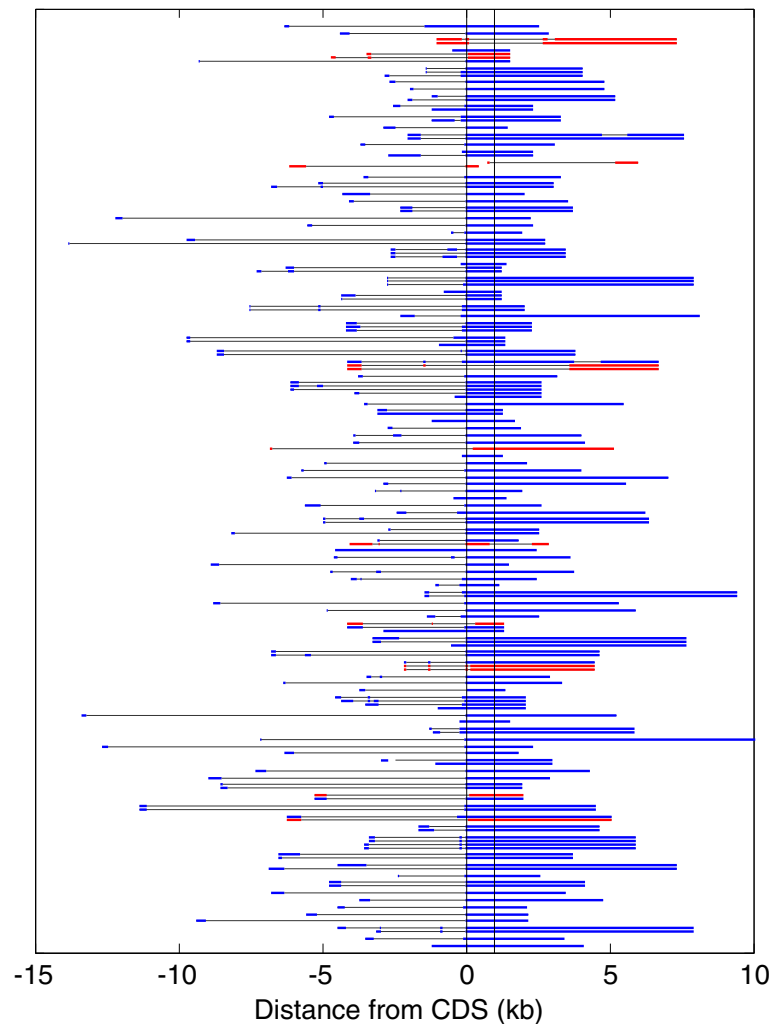
**Fig. 6** Inter-individual patterns of OR expression. **a** Inter-individual correlation of the ORs expression level (red) versus the whole genome (grey) for samples OE12 and OE15. Data are presented on a log<sub>10</sub> scale of the FPKM values. The Pearson correlation values are 0.35 for the OR genes and 0.9 for the whole genome. The complete data set is shown in Additional file 1: Figure S12. **b** Pearson correlation values for OE pairs shown by arrows: OE12,OE15 - red, OE12, OE17 - black, OE15, OE17 - blue. Also shown are distributions of Pearson correlation values obtained from 10,000 random sets for the respective pairs with same color code. The respective P values for the above pairs are <0.0001, 0.0004 and <0.0001. For the other three pairwise comparisons involving OE7 the P values were 0.0801 (OE7,OE12), <0.0001 (OE7, OE15), 0.46 (OE7-OE17), i.e. not significant (see text)

Our data provide repertoire-wide information on the gene structure of intact human OR genes. For comparison, we screened the AceView gene model repository [64], calculated from public expressed sequences and found 279 AceView transcripts that harbor also untranslated regions. However, only 36 of them (12.9 %) are spliced, with a substantial shorter the 3'UTR ( $159 \pm 352$  bp), probably due to the insufficient coverage in non-olfactory cDNA libraries. In the case of one gene, *OR51E2*, known to be highly expressed in prostate [65], we identified potential disparate tissue-specific promoters. The proximal promoter appears

to be active in olfactory epithelium, while the distal promoter – in prostate (Additional file 1: Figure S18). The latter is inferred from AceView data, as well as from the prostate Illumina Body map data.

#### The OR promoter region

We used the Genomatix RegionMiner tool to search for significant over representation of transcription factors binding sites (TFBS) immediately upstream from the inferred transcription start site. This was done jointly analyzing all 160 OR transcripts, with a comparison to the



**Fig. 7** Genomic maps of OR transcripts. Transcripts are presented on a genomic scale, phased by the OR open reading frame (vertical thin lines). Thick blue/red lines are exons and thin gray lines are introns. Narrowly spaced transcripts are for the same OR gene. Transcripts with a disrupted open reading frame are in red. These are: OR52K1, OR2V1, OR6C75, OR4M1, OR51L1, OR2A1, OR10H4, OR2K2, OR2J3 and OR6F1

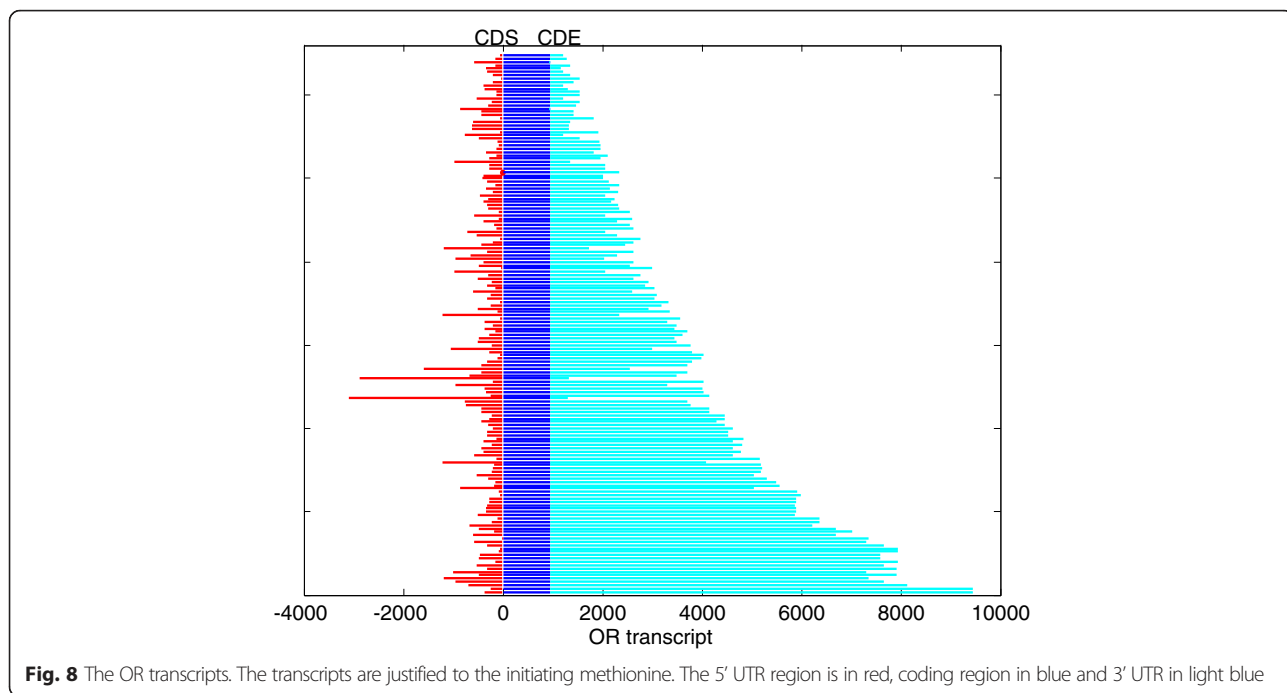
Genomatix genome-wide promoter region collection. The top 10 TFBS of all belonged to homeodomains class, including Genomatix families V\$LHXF and V\$HBOX. These respectively include the transcription factors LHX2 and EMX2, previously shown to be required to the expression of OR genes [66, 67]. The results are also in agreement with a similar analysis of mouse OR promoters [21, 68]. Repeating the analysis with a whole genome reference, or with promoters with matched low GC content (<60) to that of presumed OR promoters, identified an enrichment of the V\$NOLF family of the Early B-Cell Factor (EBF) proteins, involved in the expression of ORs [39]. These V\$NOLF TFBS are clustered in a distance of 100–300 bp from the transcription start site (Fig. 9), in broad agreement with the mouse data [21, 68]. The appearance of a distinct propensity peak for V\$NOLF

suggests that our OR transcription start site inference is adequate.

## Discussion

### Comprehensive transcriptome

The human olfactory epithelium is relatively difficult to obtain in a high quality, due to anatomical inaccessibility and heterogeneous boundaries. Specifically, there is variable contamination with non-sensory respiratory epithelium [6]. We succeeded in doing the analysis by examining biopsy and autopsy samples from 17 different human individuals, and selecting the ones most-enriched for known olfactory markers. Further, the gene-specific signals were filtered with respect to respiratory epithelial over-expression. This enables us to report a dependable

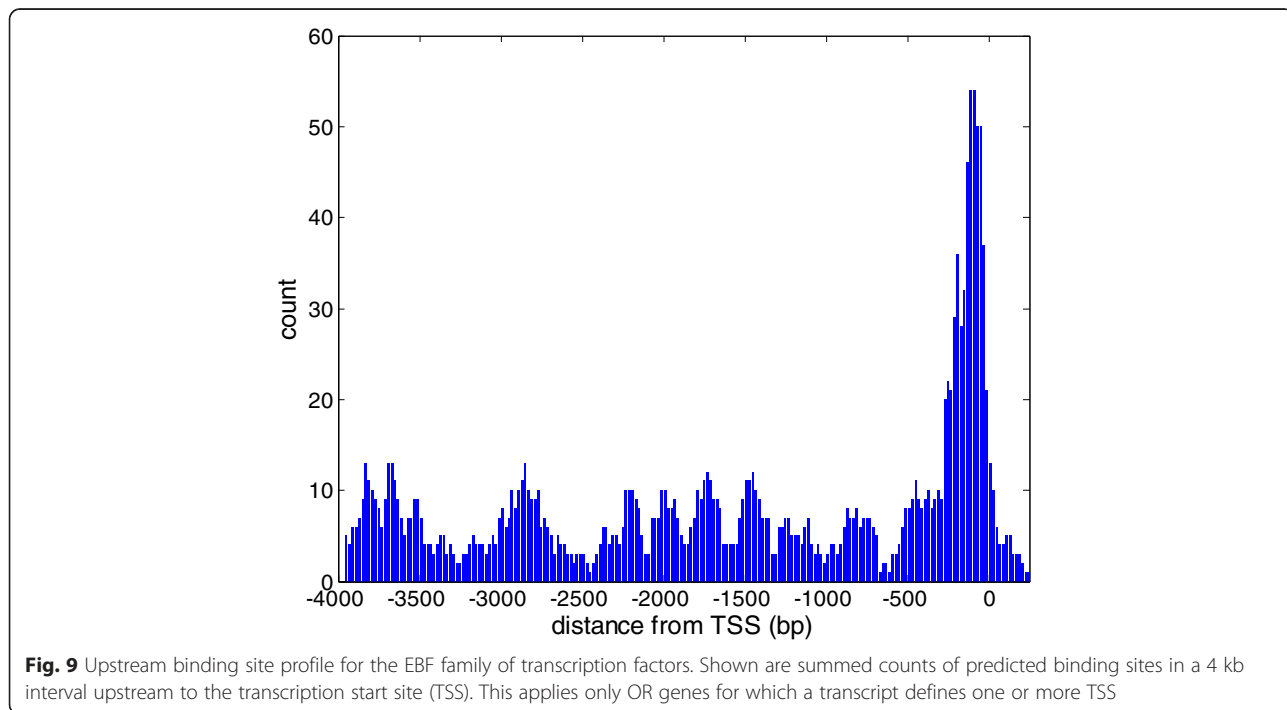


and comprehensive transcriptomic view of this human sensory tissue.

**New vista of olfactory auxiliary genes**

We identified a large number of olfactory auxiliary genes, those playing a role in transduction, development

and maintenance [10]) (Additional file 3: Table S3 and Additional file 7: Table S9). Our working hypothesis was that many of these would be highly expressed in human olfactory epithelium. It is clear that such correlation may not be perfect, i.e. genes that play a role in development and maintenance are not necessarily olfactory specific



and expressed at higher levels. Still, there is a relatively broad consensus that tissue-specific expression has functional implications [21, 69, 70].

It is also noteworthy that tissues are often heterogeneous, and a prominent example is olfactory mucosa, which includes sensory and supporting cells, basal progenitor cells and subepithelial Boman's gland cells. Further, tissue contamination may occur, e.g. with respiratory epithelium. However, as we set our overexpression cutoff of X6, a 10 % contaminating tissue would have to show X60 overexpression of a gene, and such high spurious expression signal would be easily excluded by our use of respiratory epithelium control. To summarize, equating olfactory epithelial over-expression with chemosensory function should be taken with caution, but is a relatively high probability premise.

In a first group are genes that are already well-known to play a key role in olfactory function. Some such genes are found to be significantly overexpressed in the sensory tissue (e.g. *OMP*, *GNG13* and *ANO2*), while others were not, due to a broader appearance in non-olfactory tissues (e.g. *GNAL* and *ADCY3*).

In the second group are genes that have very few or no previous publications on olfactory involvement. Some of these genes are also highly expressed in mouse olfactory epithelium, as also reported [32]. Such inter-species concordance increases the probability of olfactory involvement for those genes.

#### Apparent species-specific genes

Among the 194 olfactory over-expressed genes in human, we identified 58 that do not show over expression in mouse, or do not have a mouse ortholog. Such discordance suggests inter-species functional differences between mammals that need to be further elucidated. In an example, we showed the different patterns of expression of the broadly-defined lipocalin family members, where human and mouse show differential expression in disparate members of the family. Likewise, while in mouse several TAAR genes are overexpressed, in human only one paralog (*TAAR5*) shows such differential expression. These results may suggest that in evolution, different protein family members assume a chemosensory role. We note that some of these inter-species discrepancies could rest in the fact that the mouse data were obtained from samples pooled from four animals, which likely included a better representation of the olfactory epithelium, while the human anatomical sampling was more heterogeneous.

#### Expression of olfactory receptors

The RNAseq data accumulated in this study provide an informative view on the expression of OR genes in

olfactory epithelium and other tissues. In the olfactory tissue we observed the expression of nearly 90 % of the intact OR repertoire, but only ~60 % of all OR pseudogenes. Further, the expression distribution curve of pseudogenes is shifted to ~X4 lower intensity relative to intact genes (Fig. 4a). This may be due to the accumulation of promoter mutations or to nonsense mediated decay. In non-olfactory (ectopic) tissues the typical expression of intact OR goes down about X7, while at least in some such tissues the pseudogene expression is not much altered. This may reflect in part the functional feedback mechanism that selects against pseudogene expression [71].

An intriguing finding is the three orders of magnitude difference in expression intensity of different members of the OR repertoire, both in human and in mouse. This phenomenon has been previously observed in mouse by other methodologies [20, 72], and may be accounted for in part by variations in transcription regulatory efficiency along OR gene clusters. Here we add a functionally relevant dimension to this observant, namely that the decline of OR expression in the rank curve is much faster in human as compared to mouse, hence the overall (integral) repertoire expression is considerable weaker in human (Fig. 5b).

Finally, we observe significant inter-individual differences in the across-repertoire pattern of OR expression as compared to whole genome measures, consistent with previous reports [73, 74]. It important to ask whether this phenomenon is a true inter-individual difference, with a potential genetic basis, or due to sampling artefact. The expression of different ORs is known to vary strongly among different anatomical zones of rodent olfactory epithelium. Thus, one possible explanation could be that in different individuals tissue samples were from different functional zones, though zonal OR expression has not yet reported in human. If the mechanism is genetic after all, one potential explanation is inter-individual difference in OR regulatory sequences. Alternatively, genetic variations elsewhere could be responsible. Clarifying these issues requires to perform extensive expression quantitative trait loci (eQTL) experiments, whereby genome sequencing a transcriptome analyses is concomitantly performed on the same individuals. As this far transcends the scope of the present study, it should be clearly stated that the inter-individual expression variation trends reported here should be taken with significant caution. Likewise, if one wishes to extend such results to mouse, this would necessitate comparing several different mouse strains, again beyond the present scope.

#### The OR transcript

Despite the low expression of individual OR genes, we succeeded in defining the exon-intron structure of 160

transcripts in 100 OR genes, those that show higher expression. We are aware of the possibility that because of the relatively low coverage, some of the features described herein represent transcript-mapping artifacts. The 160 transcripts elucidated may well be a representative sample, as a correlation between gene structure deviations and expression strength is not very likely, and as no OR family bias has been observed along the expression rank curve (not shown).

The OR transcripts have certain common features. These include 1–4 short 5' non-coding exons with considerable alternative splicing: 43 of the 100 genes having 2–4 splice variants; long last 3' exon which contains the entire OR coding region, i.e. no introns are seen within the coding region. At the same time, considerable structural variability is seen in 5' exon count, as well as intron length and 3'UTR length (Fig. 7). One interesting case is that of *OR51E2*, which in olfactory epithelium contains the coding exon spliced to one 5' non-coding exon. In prostate, the same coding exon is spliced to a remote 5' non coding exon of a neighboring OR pseudogene *OR51C1P*. This chimeric transcript is known to be highly expressed in prostate, and appears to bear a functional role in this tissue [75].

OR genes have undergone a massive recent evolutionary process of pseudogenization [60]. OR pseudogenes do not arise by retroposition, but rather by gradual accumulation of point mutations and indels that render the encoded protein aberrant. We report here an additional pseudogenization process that affects 15 transcripts in 10 OR genes with an intact open reading frame. These events stem from inappropriate slicing, eliminating part or the entire coding region. We note that only in three of the cases no intact OR-encoding splice variant remains.

The analysis performed allowed us to accurately define for the first time the transcription start site (TSS) of 100 human OR genes. We subsequently analyzed the transcription regulatory elements immediately upstream to the TSS. The results indicated an enrichment of transcription factor binding site signature belonging to the large homeobox family in that region. In addition, we found a consensus localization of a specific transcription factor binding site subfamily (Olf/EBF). In mouse, these binding sites are known to be present in promoter regions of the OR repertoire, and to play a key role in OR transcription [21, 68].

## Conclusions

We report the first RNAseq study of human olfactory epithelium, aimed at obtaining a whole transcriptome overview of the sensory tissue. Our work reveals nearly 200 olfactory-enriched non-receptor transcripts, 80 % of

which have not yet been implicated in chemosensory function, thus providing a platform for future discovery.

Our study further allowed us to quantitate the expression levels of most (90 %) of the ~400 intact human olfactory receptor (OR) genes. The expression levels of different paralogous members of the OR repertoire span a 1000-fold range, suggesting a strong imbalance among different odorant specificities. Further, we obtained convincing hints that different human beings show different cross-repertoire expression patterns. Having successfully assembled transcripts for 100 OR genes, we observed prevalent genome-encoded mutations that render some transcripts inactive despite the fact that the protein coding region is intact, an unreported mechanism for OR pseudogenization.

In sum, our transcriptome study offers important observations on genes that underlie olfactory function and provides a basis for significant future work.

## Methods

### Samples collection

#### Human olfactory epithelium samples

We obtained 14 biopsy tissues and three autopsy tissues of human olfactory epithelium biopsies dissected from the superior ethmoidal, the cribriform plate and the superior septum [73, 76] during elective surgeries of the samples donors. We used qPCR (see below) to characterize the olfactory content of 12 of the biopsies samples (Additional file 1: Figure S11A), and measure the expression level of olfactory markers (CNGA2, GNAL, OMP) versus respiratory markers (KRT13 [77] and TMPRSS11D [78]). Three samples (OE12, OE15 and OE17) were selected for RNAseq and analysis. A third biopsy sample (OE7) was selected for analysis based on the expression of the olfactory markers in the RNAseq results (Additional file 1: Figure S1B and Table S1). The other samples were excluded from the analysis, as they failed to show expression of olfactory markers.

The study was approved by the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia). All participants provided informed consent.

#### Mouse samples

C57BL/6 mouse olfactory epithelium and mouse olfactory bulb RNA (Additional file 1: Table S1 samples MOE1 and MOB) were collected in pool from four adult female *Trpm5*-GFP mice crossed with C57BL/6 for over 10 generations. These experiments were performed according to protocols approved by the University of Colorado Anschutz Medical Campus Institutional Animal Care and Use Committee. BALB/c olfactory epithelium RNAseq data were kindly provided by Wen-Hsiung Li, Academia Sinica, Taiwan [79]. Of note, the MOB sample

was used only for presentation in Additional file 1: Figures S2-S6, S8, and S9.

#### **Mouse olfactory sensory neurons**

Single dissociated cells of olfactory sensory neurons were randomly selected under a light microscope and seeded into a tube containing cell-lysis buffer by mouth pipetting. After reverse transcription and global amplification of cDNA, diagnostic PCR for four mature olfactory sensory neuron marker genes (olfactory marker protein (*Omp*); guanine nucleotide binding protein; alpha stimulating, olfactory type (*Gnal*); cyclic nucleotide gated channel alpha 2 (*Cnga2*) and adenylate cyclase 3 (*Adcy3*)) was carried out using the diluted amplified cDNA. Samples with the presence of all of four markers were selected for library preparation and sequencing by following the Illumina protocol. After aligning the read data to the mouse genome (mm9) an expression profile was generated using only the uniquely mapped reads and those that are mapped to exons. The number of reads per gene was further divided by the total number of the sample mapped reads (reads per million, RPM). The data presented in this manuscript is the average of nine olfactory sensory neurons collected from female *Mus musculus domesticus* (C57BL/6, B6) and an hybrid mice from cross of male *Mus spretus* (*Spretus*) and female B6.

#### **Control samples**

The Illumina human body map 2.0 project data containing RNAseq of sixteen normal human tissues were used in this study as controls. Data were obtained directly from the company.

#### **Experimental procedures**

##### **Reverse transcription reaction**

cDNA was synthesized with the SuperScript® First-strand Synthesis System for RT-PCR (Invitrogen, Carlsband, CA, USA) according to the manufacturer's instructions.

##### **Quantitative real-time polymerase Chain Reaction (qPCR)**

qPCR reactions were performed using TaqMan® Gene Expression Assays (Applied Biosystems, Foster City, CA, USA). A total of 2ul of cDNA was added with 2.5ul of water and 0.5ul TaqMan® Gene Expression Assay to 5 ul of TaqMan® universal PCR Master Mix (Applied Biosystems) and the resulting 10 ul reaction mixtures were loaded onto a 96-well PCR plate. We used eight different TaqMan® Gene Expression Assays including three housekeeping genes with the following assays IDs: Hs01087269\_s1 (OMP), Hs00181836\_m1 (GNAL), Hs00864448\_s1 (RT P1), Hs01377537\_m1 (CNGA2), Hs00975370\_m1 (TM PRSS11D), and ACTB (Hs99999903\_m1) and GAPDH (Hs99999905\_m1) as housekeeping genes.

Next generation sequencing-OE, MOE1 and the olfactory respiratory sample were sequenced at the biological services of the Weizmann institute using the Illumina Genome Analyzer platform (Illumina GA IIX) and the standard Illumina protocol at the experiment time. Samples OE12-OE17 were sequenced at Macrogen (Seoul, Korea). RNA samples were evaluated by ultraviolet spectroscopy for purity and concentration (NanoDrop, Wilmington, DE) and were assessed further for RNA integrity on the Agilent Bioanalyzer (Santa Clara, CA). Libraries were prepared using the Illumina mRNA-seq Sample Preparation Kit (San Diego, CA) and validated with an Agilent Bioanalyzer (Santa Clara, CA).

#### **Data analysis**

##### **Gene expression quantification**

Reads of human olfactory epithelium were aligned to the hg19 reference genome using Tophat 2.0.4 [80]. Quantification of expression level was performed with Cufflinks v2.1.1. [81] using Ensembl transcripts and HORDE build#43 annotation [10] of OR genes and pseudogenes. To enable downstream calculations, 0 FPKM values were set to 0.003. Mouse data were analyzed with the same procedures using Mm10 as a reference genome. The GeneAlaCart tool of the GeneCards database [82] was used to find the mouse ortholog of each human gene.

##### **Differential expression**

To test the statistical significance of over expression in the olfactory epithelium versus the controls we used HTSeq [83] to count the number of uniquely mapped reads per gene, following by the R package DESeq [84], using the illumina Body Map samples as control tissues. Genes with a fold change  $\geq 6$  and  $p < 0.001$  were considered as statistical significant. To overcome contamination from respiratory epithelium genes with expression in the respiratory epithelium that exceeded 0.1 of the expression in the olfactory epithelium were removed. We also included genes with a significant over expression in mouse olfactory epithelium provided that their over expression in human olfactory epithelium was above 6. The same analysis was applied to the mouse olfactory epithelium data. In the absence of mouse controls at the time that the analysis was performed, we used the illumina Body Map samples as controls [10]. Later on, with the progress of the mouse ENCODE project, we verified the over expression of the significant genes against a set of 11 mouse tissues (bone marrow, cerebellum, cortex, heart, intestine, kidney, liver, lung, spleen, testes, thymus) from the Mouse ENCODE project (<http://chromosome.sdsc.edu/mouse/download.html>).

### Prior information on overexpressed genes

To associate between a given gene and olfaction we performed an automatic search in PubMed with the gene symbols of the olfactory enriched genes and a set of olfactory related keywords (olfactory;olfaction;odorant;odor;chemosensory;smell). Followed by a manual curation of some genes, the number of PMID hits was used to classify genes into class A or B.

### Analysis of the lipocalin protein family

Most of lipocalin family members have no orthologs. Therefore over expression of mouse lipocalin genes were calculated using Mouse ENCODE data (<http://chromosome.sdsc.edu/mouse/download.html>) while over expression of human lipocalins was calculated relative to the human Illumina Body Map. Only tissues that are shared between the two control data sets were used in the calculation (adipose, brain, heart, kidney, liver, lung and testis), although the fold change of the over expressed genes did not altered when including the complete data set.

### Assembly of OR isoforms

Uniquely mapped reads of all four studied olfactory epithelium samples were joined to assemble the OR transcripts. We used all reads from the OR territories defined as HORDE OR clusters [10] extended by 300 kb to the 5' and 3' of each cluster as an input to cufflinks, and applied the parameters `-A 0.15 -trim-3-dropoff-frac 0.15`. Cuffcompare [81] was used for gene identification. Transcripts encoded on the opposite strand of the OR genes and transcripts that are suspected as polymerase run-on fragment (cuffcompare classcodes `x`, `s` and `p`) were removed. We also removed isoforms with expression level  $<0.15\%$  of the major isoform of a gene. Transcripts with a single exon of  $>5$  kb length were curated by manual inspection.

### Testing inter-individual differences of

To test the significance of OR expression we used all 264 OR genes that are expressed in all olfactory epithelium tissues and compared their expression to a control set of 15,846 protein coding genes which are also expressed in all tissues. The distribution of the ORs was calculated. We then randomly selected 264 control genes such that their distribution will follow that of the ORs tested by Kolmogorov-Smirnov  $p > 0.05$ . The process was repeated 10,000 times. Samples OE12-17, which proved to contain the highest olfactory epithelium content, and show the highest similarity, were used for this test.

### Data-mining of AceView gene model

AceView transcripts [64] which overlap the OR coding regions were downloaded using the UCSC TableBrowser

tool of the [85]. We then applied cuffcompare software against a gtf file of HORDE genes to assign the transcripts to the OR genes. Transcripts that overlap the opposite strand (cuffcompare class code `x`) were removed.

### Promoter analysis

The tool RegionMiner of Genomatix was used to search for over representation of TFBS in the OR upstream regions. We used an interval of  $-750$  bp and  $+250$  of the TSS of each transcript. Transcripts representing the same OR gene were included if their upstream regions did not overlap by  $\geq 500$  bp. The analysis compared the sequences to a library of RefSeq promoters.

### Additional files

**Additional file 1:** Table S1, Figures S1-S14.(PPTX 2640 kb)

**Additional file 2: Table S2.** Whole genome expression of the studied samples (in FPKM). FC\_OE, the fold change of the human olfactory epithelium relative to the controls; FC\_resp, the fold change of the human respiratory epithelium relative to the controls; FC\_MOE, the fold change of mouse olfactory epithelium relative to the controls; FC\_MOB, the fold change of mouse olfactory bulb relative to the controls. NoOrth- no mouse ortholog. OSN- olfactory sensory neurons in RPM (reads per million). (XLSX 22311 kb)

**Additional file 3: Table S3.** Olfactory over expressed genes. Tissue abbreviation: HOE, human olfactory epithelium; MOE: mouse olfactory epithelium; MOB: mouse olfactory bulb; RES: respiratory epithelium. Numbers in these columns are the averaged fold change relative to the illumina body map control samples (CTRL; FPKM), MOE\_encode: the average fold change of mouse olfactory epithelium relative to mouse ENCODE data. OSN- olfactory sensory neurons. #pub- number of associated publications that were found in the automatic literature search. **Table S4.** Olfactory over expressed genes with prior evidences of an olfactory role. HOE: the fold change of human olfactory epithelium relative to the control tissues. MOE: the fold change of mouse olfactory epithelium relative to the control tissues. The 16 illumina Body Map tissues were used as controls in the calculation. MOE\_encode: the fold change of mouse olfactory epithelium relative to 11 mouse ENCODE tissues. OSN- the fold change in mouse olfactory neurons relative to mouse olfactory epithelium. In bold- genes mentioned in the manuscript text. Gene symbols are from HGNC (<http://www.genenames.org/>). **Table S5.** Olfactory over expressed genes with little or no previous knowledge on olfactory role. See Additional file 3: Table S4 for more details. Symbols with \* are from GeneCards (<http://www.genecards.org/>), with an informal gene name. (XLSX 1879 kb)

**Additional file 4: Table S6.** Expression of the lipocalin family members. Mouse was calculated relative to Mouse ENCODE data (<http://chromosome.sdsc.edu/mouse/download.html>), see methods. (XLSX 14 kb)

**Additional file 5: Table S7.** Expression profile of the OR repertoire. (XLSX 161 kb)

**Additional file 6: Table S8.** Genomic coordinates of the OR transcripts (in gtf format, hg19). (GTF 70 kb)

**Additional file 7: Table S9.** Genes that are discussed in the text. (XLSX 37 kb)

### Abbreviations

OR, olfactory receptor; TFBS, transcription factor binding site; TSS, transcription start site; UTR, untranslated region

### Acknowledgements

Not applicable.

### Funding

This work is supported by the Crown Human Genome Center of the Weizmann Institute, Life Map Sciences, Inc. (USA), NHGRI Grant U41HG003345 to DL, NIH grant DC014253 to DR and NIH grants DC012095 and DC014423 to HM.

### Availability of data and materials

The data sets supporting the results of this article are included within the article (and its additional files). The raw data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [86] and are accessible through GEO Series accession number GSE80249 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80249>). The phylogenetic data have been deposited in TreeBASE [87], (submission number 19597).

### Authors' contributions

TO and DL were involved in the project design, data analysis and manuscript preparation, IK was involved in the data analysis, JP and YG were responsible for obtaining human olfactory epithelium, PT was responsible for the cDNA preparation and qPCR, SF contributed to the classification of the olfactory over expressed genes to gene families, AA was involved as an advisor, MS and HM performed mouse OSN single cell transcriptomics, DR contributed mouse olfactory epithelium. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

Samples collection was approved by the Institutional Review Board at The University of Chicago, the Ottawa Health Science Network Research Ethics Board and the Ethics Board of University Rehabilitation Institute (Ljubljana, Slovenia). All participants provided written informed consent. Mouse samples were collected according to protocols approved by the University of Colorado Anschutz Medical Campus Institutional Animal Care and Use Committee.

### Author details

<sup>1</sup>Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. <sup>2</sup>Section of Otolaryngology-Head and Neck Surgery, University of Chicago, Chicago, IL, USA. <sup>3</sup>Department of Molecular Genetics and Microbiology, Department of Neurobiology, Duke Institute for Brain Sciences, Duke University Medical Center, Durham, NC, USA. <sup>4</sup>Department of Cell and Developmental Biology, Neuroscience Program, and Rocky Mountain Taste and Smell Center, University of Colorado School of Medicine, Aurora, CO, USA. <sup>5</sup>Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Received: 20 January 2016 Accepted: 21 July 2016

Published online: 11 August 2016

### References

- Hasin-Brumshtein Y, Lancet D, Olender T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet.* 2009;25(4):178–84.
- Hatt H. Molecular and cellular basis of human olfaction. *Chem Biodivers.* 2004;1(12):1857–69.
- Hayden S, Teeling EC. The molecular biology of vertebrate olfaction. *Anat Rec (Hoboken).* 2014;297(11):2216–26.
- Ihara S, Yoshikawa K, Touhara K. Chemosensory signals and their receptors in the olfactory neural system. *Neuroscience.* 2013;254:45–60.
- Chen J, Shi Y, Regan J, Karuppaiah K, Ornitz DM, Long F. Osx-Cre targets multiple cell types besides osteoblast lineage in postnatal mice. *PLoS One.* 2014;9(1):e85161.
- Escada PA, Lima C, da Silva JM. The human olfactory mucosa. *Eur Arch Otorhinolaryngol.* 2009;266(11):1675–80.
- Rospars JP, Gu Y, Gremiaux A, Lucas P. Odour transduction in olfactory receptor neurons. *Chin J Physiol.* 2010;53(6):364–72.
- Kaupp UB. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat Rev Neurosci.* 2010;11(3):188–200.
- Mori K, Sakano H. How is the olfactory map formed and interpreted in the mammalian brain? *Annu Rev Neurosci.* 2011;34:467–99.
- Keydar I, Ben-Asher E, Feldmesser E, Nativ N, Oshimoto A, Restrepo D, et al. General olfactory sensitivity database (GOSdb): candidate genes and their genomic variations. *Hum Mutat.* 2013;34(1):32–41.
- Sun L, Wang H, Hu J, Han J, Matsunami H, Luo M. Guanylyl cyclase-D in the olfactory CO2 neurons is activated by bicarbonate. *Proc Natl Acad Sci U S A.* 2009;106(6):2041–6.
- Young JM, Waters H, Dong C, Fulle HJ, Liman ER. Degeneration of the olfactory guanylyl cyclase D gene during primate evolution. *PLoS One.* 2007;2(9):e884.
- Olender T, Lancet D, Nebert DW. Update on the olfactory receptor (OR) gene superfamily. *Hum Genomics.* 2008;3(1):87–97.
- Olender T, Feldmesser E, Atarot T, Eisenstein M, Lancet D. The olfactory receptor universe—from whole genome analysis to structure and evolution. *Genet Mol Res.* 2004;3(4):545–53.
- Zozulya S, Echeverri F, Nguyen T. The human olfactory receptor repertoire. *Genome Biol.* 2001;2(6):RESEARCH0018.
- Niimura Y, Nei M. Evolution of olfactory receptor genes in the human genome. *Proc Natl Acad Sci U S A.* 2003;100(21):12235–40.
- Sosinsky A, Glusman G, Lancet D. The genomic structure of human olfactory receptor genes. *Genomics.* 2000;70(1):49–61.
- Ibarra-Soria X, Levitin MO, Saraiva LR, Logan DW. The olfactory transcriptomes of mice. *PLoS Genet.* 2014;10(9):e1004593.
- Hoppe R, Breer H, Strotmann J. Promoter motifs of olfactory receptor genes expressed in distinct topographic patterns. *Genomics.* 2006;87(6):711–23.
- Young JM, Shykind BM, Lane RP, Tonnes-Priddy L, Ross JA, Walker M, et al. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biol.* 2003;4(11):R71.
- Plessy C, Pascarella G, Bertin N, Akalin A, Carrieri C, Vassalli A, et al. Promoter architecture of mouse olfactory receptor genes. *Genome Res.* 2012;22(3):486–97.
- Scholz P, Kalbe B, Jansen F, Altmueller J, Becker C, Mohrhardt J et al.: Transcriptome Analysis of Murine Olfactory Sensory Neurons during Development Using Single Cell RNA-Seq. *Chemical Senses.* 2016;41(4):313–23. doi: 10.1093/chemse/bjw003.
- Saraiva LR, Ibarra-Soria X, Khan M, Omura M, Scialdone A, Mombaerts P, et al. Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Sci Rep.* 2015;5:18178.
- Tan L, Li Q, Xie XS. Olfactory sensory neurons transiently express multiple olfactory receptors during development. *Mol Syst Biol.* 2015;11(12):844.
- Hanchate NK, Kondoh K, Lu Z, Kuang D, Ye X, Qiu X, et al. Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science.* 2015;350(6265):1251–5.
- Hondo M, Ishii M, Sakurai T. The NPB/NPW neuropeptide system and its role in regulating energy homeostasis, pain, and emotion. *Results Probl Cell Differ.* 2008;46:239–56.
- Matsuoka Y, Furuyashiki T, Yamada K, Nagai T, Bito H, Tanaka Y, et al. Prostaglandin E receptor EP1 controls impulsive behavior under stress. *Proc Natl Acad Sci U S A.* 2005;102(44):16066–71.
- Leon S, Garcia-Galiano D, Ruiz-Pino F, Barroso A, Manfredi-Lozano M, Romero-Ruiz A, et al. Physiological roles of gonadotropin-inhibitory hormone signaling in the control of mammalian reproductive axis: studies in the NPFF1 receptor null mouse. *Endocrinology.* 2014;155(8):2953–65.
- Bester-Meredith JK, Fancher AP, Mammarella GE. Vasopressin proves Es-sense-tial: vasopressin and the modulation of sensory processing in mammals. *Front Endocrinol (Lausanne).* 2015;6:5.
- Deussing JM, Breu J, Kuhne C, Kallnik M, Bunck M, Glasl L, et al. Urocortin 3 modulates social discrimination abilities via corticotropin-releasing hormone receptor type 2. *J Neurosci.* 2010;30(27):9103–16.
- Morais-Cabral JH, Robertson GA. The enigmatic cytoplasmic regions of KCNH channels. *J Mol Biol.* 2015;427(1):67–76.
- Kanageswaran N, Demond M, Nagel M, Schreiner BS, Baumgart S, Scholz P, et al. Deep sequencing of the murine olfactory receptor neuron transcriptome. *PLoS One.* 2015;10(1):e0113170.
- Al Khamici H, Brown LJ, Hossain KR, Hudson AL, Sinclair-Burton AA, Ng JP, et al. Members of the chloride intracellular ion channel protein family demonstrate glutaredoxin-like enzymatic activity. *PLoS One.* 2015;10(1):e115699.
- Jacobsson JA, Haitina T, Lindblom J, Fredriksson R. Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family. *Genomics.* 2007;90(5):595–609.



35. Hagglund MG, Hellsten SV, Bagchi S, Philippot G, Lofqvist E, Nilsson VC, et al. Transport of L-glutamine, L-alanine, L-arginine and L-histidine by the neuron-specific Slc38a8 (SNAT8) in CNS. *J Mol Biol*. 2015;427(6 Pt B):1495–512.
36. Jonsson M, Bjornorp Mark E, Brantsing C, Brandner JM, Lindahl A, Asp J. Hash4, a novel human achaete-scute homologue found in fetal skin. *Genomics*. 2004;84(5):859–66.
37. Golubiatnikov VP, Bukharina TA, Furman DP. A model study of the morphogenesis of *D. melanogaster* mechanoreceptors: the central regulatory circuit. *J Bioinform Comput Biol*. 2015;13(1):1540006.
38. Guillemot F, Lo LC, Johnson JE, Auerbach A, Anderson DJ, Joyner AL. Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell*. 1993;75(3):463–76.
39. Sammeta N, Hardin DL, McClintock TS. Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium. *Mol Cell Neurosci*. 2010;45(4):398–407.
40. Suzuki Y, Tsuruga E, Yajima T, Takeda M. Expression of bHLH transcription factors NSCL1 and NSCL2 in the mouse olfactory system. *Chem Senses*. 2003;28(7):603–8.
41. Harden MV, Newton LA, Lloyd RC, Whitlock KE. Olfactory imprinting is correlated with changes in gene expression in the olfactory epithelia of the zebrafish. *J Neurobiol*. 2006;66(13):1452–66.
42. Haines BP, Rigby PW. Expression of the Lingo/LERN gene family during mouse embryogenesis. *Gene Expr Patterns*. 2008;8(2):79–86.
43. Dougherty GW, Adler HJ, Rzadzinska A, Gimona M, Tomita Y, Lattig MC, et al. CLAMP, a novel microtubule-associated protein with EB-type calponin homology. *Cell Motil Cytoskeleton*. 2005;62(3):141–56.
44. Lim W, Song G. Differential expression of vitelline membrane outer layer protein 1: hormonal regulation of expression in the oviduct and in ovarian carcinomas from laying hens. *Mol Cell Endocrinol*. 2015;399:250–8.
45. Wang Z, Chen Z, Yang Q, Jiang Y, Lin L, Liu X, et al. Vitelline membrane outer layer 1 homolog interacts with lysozyme C and promotes the stabilization of tear film. *Invest Ophthalmol Vis Sci*. 2014;55(10):6722–7.
46. Shamsi FA, Chen Z, Liang J, Li K, Al-Rajhi AA, Chaudhry IA, et al. Analysis and comparison of proteomic profiles of tear fluid from human, cow, sheep, and camel eyes. *Invest Ophthalmol Vis Sci*. 2011;52(12):9156–65.
47. Bojjireddy N, Guzman-Hernandez ML, Reinhard NR, Jovic M, Balla T. EFR3s are palmitoylated plasma membrane proteins that control responsiveness to G-protein-coupled receptors. *J Cell Sci*. 2015;128(1):118–28.
48. Schiefner A, Freier R, Eichinger A, Skerra A. Crystal structure of the human odorant binding protein, OBPIIa. *Proteins*. 2015;83(6):1180–4.
49. Andraut JB, Gaillard I, Giorgi D, Rouquier S. Expansion of the BPI family by duplication on human chromosome 20: characterization of the RY gene cluster in 20q11.21 encoding olfactory transporters/antimicrobial-like peptides. *Genomics*. 2003;82(2):172–84.
50. Heydel JM, Coelho A, Thiebaud N, Legendre A, Le Bon AM, Faure P, et al. Odorant-binding proteins and xenobiotic metabolizing enzymes: implications in olfactory perireceptor events. *Anat Rec (Hoboken)*. 2013; 296(9):1333–45.
51. Minn A, Leclerc S, Heydel JM, Minn AL, Denizcot C, Cattarelli M, et al. Drug transport into the mammalian brain: the nasal pathway and its specific metabolic barrier. *J Drug Target*. 2002;10(4):285–96.
52. Lazard D, Tal N, Rubinstein M, Khen M, Lancet D, Zupko K. Identification and biochemical analysis of novel olfactory-specific cytochrome P-450IIA and UDP-glucuronosyl transferase. *Biochemistry*. 1990;29(32):7433–40.
53. Lazard D, Zupko K, Poria Y, Nef P, Lazarovits J, Horn S, et al. Odorant signal termination by olfactory UDP glucuronosyl transferase. *Nature*. 1991; 349(6312):790–3.
54. Lancet D. Olfactory reception: from transduction to human genetics. *Soc Gen Physiol Ser*. 1992;47:73–91.
55. Nef P, Heldman J, Lazard D, Margalit T, Jaye M, Hanukoglu I, et al. Olfactory-specific cytochrome P-450. cDNA cloning of a novel neuroepithelial enzyme possibly involved in chemoreception. *J Biol Chem*. 1989;264(12):6780–5.
56. Yu TT, McIntyre JC, Bose SC, Hardin D, Owen MC, McClintock TS. Differentially expressed transcripts from phenotypically identified olfactory sensory neurons. *J Comp Neurol*. 2005;483(3):251–62.
57. Brigelius-Flohe R. Glutathione peroxidases and redox-regulated transcription factors. *Biol Chem*. 2006;387(10–11):1329–35.
58. Kim DS, Wang Y, Oh HJ, Lee K, Hahn Y. Frequent loss and alteration of the MOXD2 gene in catarrhines and whales: a possible connection with the evolution of olfaction. *PLoS One*. 2014;9(8):e104085.
59. Hahn Y, Jeong S, Lee B. Inactivation of MOXD2 and S100A15A by exon deletion during human evolution. *Mol Biol Evol*. 2007;24(10):2203–12.
60. Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome. *Genome Res*. 2001;11(5):685–702.
61. Flegel C, Manteniots S, Osthold S, Hatt H, Gisselmann G. Expression profile of ectopic olfactory receptors determined by deep sequencing. *PLoS One*. 2013;8(2):e55368.
62. Menashe I, Aloni R, Lancet D. A probabilistic classifier for olfactory receptor pseudogenes. *BMC Bioinformatics*. 2006;7:393.
63. Shum EY, Espinoza JL, Ramaiah M, Wilkinson MF. Identification of novel post-transcriptional features in olfactory receptor family mRNAs. *Nucleic Acids Res*. 2015;43(19):9314–26.
64. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*. 2006;7 Suppl 1:S12. 11–14.
65. Xu LL, Stackhouse BG, Florence K, Zhang W, Shanmugam N, Sesterhenn IA, et al. PSGR, a novel prostate-specific gene with homology to a G protein-coupled receptor, is overexpressed in prostate cancer. *Cancer Res*. 2000; 60(23):6568–72.
66. McIntyre JC, Bose SC, Stromberg AJ, McClintock TS. Emx2 stimulates odorant receptor gene expression. *Chem Senses*. 2008;33(9):825–37.
67. Hirota J, Omura M, Mombaerts P. Differential impact of Lhx2 deficiency on expression of class I and class II odorant receptor genes in mouse. *Mol Cell Neurosci*. 2007;34(4):679–88.
68. Clowney EJ, Magklara A, Colquitt BM, Pathak N, Lane RP, Lomvardas S. High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res*. 2011;21(8):1249–59.
69. Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, et al. A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol*. 2008;6:49.
70. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13(2):397–406.
71. Monahan K, Lomvardas S. Monoallelic expression of olfactory receptors. *Annu Rev Cell Dev Biol*. 2015;31:721–40.
72. Vaes E, Khan M, Mombaerts P. Statistical analysis of differential gene expression relative to a fold change threshold on NanoString data of mouse odorant receptor genes. *BMC Bioinformatics*. 2014;15:39.
73. Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y. Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol*. 2007;8(5):R86.
74. Verbeurgt C, Wilkin F, Tarabichi M, Gregoire F, Dumont JE, Chatelain P. Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS One*. 2014;9(5):e96333.
75. Xu LL, Sun C, Petrovics G, Makarem M, Furusato B, Zhang W, et al. Quantitative expression profile of PSGR in prostate cancer. *Prostate Cancer Prostatic Dis*. 2006;9(1):56–61.
76. De la Cruz O, Blekhan R, Zhang X, Nicolae D, Firestein S, Gilad Y. A signature of evolutionary constraint on a subset of ectopically expressed olfactory receptor genes. *Mol Biol Evol*. 2009;26(3):491–4.
77. Noruddin NA, Saim AB, Chua KH, Idrus R. Human nasal turbinates as a viable source of respiratory epithelial cells using co-culture system versus dissociation technique. *Laryngoscope*. 2007;117(12):2139–45.
78. Lee HJ, Yang YM, Kim K, Shin DM, Yoon JH, Cho HJ, et al. Protease-activated receptor 2 mediates mucus secretion in the airway submucosal gland. *PLoS One*. 2012;7(8):e43188.
79. Shiao MS, Chang AY, Liao BY, Ching YH, Lu MY, Chen SM, et al. Transcriptomes of mouse olfactory epithelium reveal sexual differences in odorant detection. *Genome Biol Evol*. 2012;4(5):703–12.
80. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
81. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–78.
82. Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, et al. In-silico human genomics with GeneCards. *Hum Genomics*. 2011;5(6):709–17.
83. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9.
84. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.

85. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform.* 2013;14(2):144–61.
86. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10.
87. Wang JT, Shan H, Shasha D, Piel WH. Fast structural search in phylogenetic databases. *Evol Bioinformatics Online.* 2005;1:37–46.
88. Senthil R, Angel KJ, Malathi R, Venkatesan D. Isolation, identification and computational studies on *Pseudomonas aeruginosa* sp. strain MPC1 in tannery effluent. *Bioinformation.* 2011;6(5):187–90.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

