Wesleyan University
WesScholar

Division III Faculty Publications | Natural Sciences and Mathematics

January 2003

Odorant Receptor Expressed Sequence Tags Demonstrate Olfactory Expression of over 400 Genes, Extensive Alternate Splicing and Unequal Expression Levels

Robert Lane
Wesleyan University, rlane@wesleyan.edu

Follow this and additional works at: http://wesscholar.wesleyan.edu/div3facpubs

Part of the Life Sciences Commons

Recommended Citation
Lane, Robert, "Odorant Receptor Expressed Sequence Tags Demonstrate Olfactory Expression of over 400 Genes, Extensive Alternate Splicing and Unequal Expression Levels" (2003). *Division III Faculty Publications*. Paper 278.
http://wesscholar.wesleyan.edu/div3facpubs/278

This Article is brought to you for free and open access by the Natural Sciences and Mathematics at WesScholar. It has been accepted for inclusion in Division III Faculty Publications by an authorized administrator of WesScholar. For more information, please contact dschnaidt@wesleyan.edu, ljohnson@wesleyan.edu.

Research

Open Access

Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels

Janet M Young[*], Benjamin M Shykind[†], Robert P Lane[*‡], Lori Tonnes-Priddy[*§], Joseph A Ross[*], Megan Walker[*], Eleanor M Williams[*] and Barbara J Trask[*]

Addresses: [*]Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA 98109, USA. [†]Center for Neurobiology and Behavior, College of Physicians and Surgeons, Howard Hughes Medical Institute, Columbia University, 701 W 168th Street, New York, NY 10032, USA. [‡]Current address: Molecular Biology & Biochemistry Department, Wesleyan University, 237 High Street, Middletown, CT 06459, USA. [§]Current address: Epigenomics Inc., 1000 Seneca Street, Seattle, WA 98101, USA.

Correspondence: Janet M Young. E-mail: jayoung@fhcrc.org. Barbara J Trask. E-mail: btrask@fhcrc.org

Published: 7 October 2003

Genome ***Biology*** 2003, **4:**R71

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/11/R71

Received: 9 July 2003
Revised: 18 August 2003
Accepted: 27 August 2003

© 2003 Young *et al.*; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The olfactory receptor gene family is one of the largest in the mammalian genome. Previous computational analyses have identified approximately 1,500 mouse olfactory receptors, but experimental evidence confirming olfactory function is available for very few olfactory receptors. We therefore screened a mouse olfactory epithelium cDNA library to obtain olfactory receptor expressed sequence tags, providing evidence of olfactory function for many additional olfactory receptors, as well as identifying gene structure and putative promoter regions.

Results: We identified more than 1,200 odorant receptor cDNAs representing more than 400 genes. Using real-time PCR to confirm expression level differences suggested by our screen, we find that transcript levels in the olfactory epithelium can differ between olfactory receptors by up to 300-fold. Differences for one gene pair are apparently due to both unequal numbers of expressing cells and unequal transcript levels per expressing cell. At least two-thirds of olfactory receptors exhibit multiple transcriptional variants, with alternative isoforms of both 5' and 3' untranslated regions. Some transcripts (5%) utilize splice sites within the coding region, contrary to the stereotyped olfactory receptor gene structure. Most atypical transcripts encode nonfunctional olfactory receptors, but can occasionally increase receptor diversity.

Conclusions: Our cDNA collection confirms olfactory function of over one-third of the intact mouse olfactory receptors. Most of these genes were previously annotated as olfactory receptors based solely on sequence similarity. Our finding that different olfactory receptors have different expression levels is intriguing given the one-neuron, one-gene expression regime of olfactory receptors. We provide 5' untranslated region sequences and candidate promoter regions for more than 300 olfactory receptors, valuable resources for computational regulatory motif searches and for designing olfactory receptor microarrays and other experimental probes.

Background

The interaction of olfactory (or odorant) receptors with their odorant ligands is the first step in a signal transduction pathway that results in the perception of smell. The olfactory receptor gene family is one of the largest in the mammalian genome, comprising about 1,500 members in the mouse genome [1,2]. Olfactory receptors were originally identified in an elegant experiment based on the hypothesis that they would be seven-transmembrane-domain proteins encoded by a large, diverse gene family whose expression is restricted to the olfactory epithelium [3]. Subsequent studies have shown that some of these receptors do indeed respond to odorants and can confer that responsivity when expressed in heterologous cell types (for example [4]). Recent computational investigations have provided the almost complete human [5,6] and mouse [1,2] olfactory receptor-gene catalogs. However, the assignment of most of these genes as olfactory receptors is based solely on similarity to one of a relatively small number of experimentally confirmed mouse olfactory receptors or, worse, on similarity to a gene that in turn was defined as an olfactory receptor solely by similarity. While similarity-based genome annotation is a good initial method to identify genes and predict their function, in some cases it can be misleading, as genes of similar sequence can carry out different functions and be expressed in different tissues (for example, the sugar transporter gene family [7]).

A small subset of olfactory receptors appears to be expressed in non-olfactory tissues, principally the testis [8], but also taste tissues [9], prostate [10], erythroid cells [11], notochord [12] and perhaps other tissues. Expression in the testis has led some investigators to suggest that a subset of olfactory receptors may function as spermatid chemoreceptors [8]. Recent studies of one human testis-expressed olfactory receptor indicate that it does indeed function in sperm chemotaxis [13]. Due to the paucity of experimental evidence of the olfactory function of most genes in the family and suggestions of extra-olfactory roles, we embarked on an olfactory receptor expressed sequence tag (EST) project to confirm olfactory epithelial expression of hundreds of mouse odorant receptor genes.

Within the olfactory epithelium, individual olfactory receptor genes show an intriguing expression pattern. Each olfactory receptor is expressed in a subset of cells in one of four zones of the epithelium [14,15]. Furthermore, each olfactory neuron expresses only one allele [16] of a single olfactory receptor gene [17,18], and the remaining approximately 1,499 genes are transcriptionally inactive. While the mechanism ensuring singular expression is unknown, many hypotheses have been proposed [14,16,19]. In one model, somatic DNA recombination would bring one olfactory receptor gene into a transcriptionally active genomic configuration, as observed for the yeast mating type locus [20] and the mammalian immunoglobulin genes [21]. Alternatively, a second model invokes a combinatorial code of transcription factor binding sites unique to each gene. This is unlikely, however, as even olfactory receptor transgenes with identical upstream regions are expressed in different neurons [18]. In a third model, there would be a limiting quantity of transcription factors - the cell might contain a single transcriptional 'machine' that is capable of accommodating the promoter of only one olfactory receptor gene, similar to the expression site body used by African trypanosomes to ensure singular expression of only one set of variant surface glycoprotein genes [22]. Finally, in a fourth model, transcriptional activity at one stochastically chosen olfactory receptor allele might send negative feedback to repress activity of all other olfactory receptors and/or positive feedback to enhance its own expression. In the latter three models, some or all olfactory receptor genes might share transcription factor binding motifs, and in the first model, olfactory receptor genes might share a common recombination signal. In order to perform computational and experimental searches for such signals, it is important to have a better idea of the transcriptional start site of a large number of olfactory receptor genes. Our olfactory receptor EST collection provides 5' untranslated region (UTR) sequences for many genes and, therefore, a large dataset of candidate promoter regions.

Olfactory receptor genes have an intronless coding region, simplifying both computational and experimental olfactory receptor identification. For a small number of olfactory receptors, gene structure has been determined. Additional 5' untranslated exons lie upstream of the coding region and can be alternatively spliced [19,23-26]. The 3' untranslated region is typically intronless. Exceptions to this stereotyped structure have been described for some human olfactory receptors, but are thought to be rare [25-27]. cDNA identification and RACE data have been used to determine gene structure for about 30 genes, see, for example, [19,23]. However, computational prediction of the location of 5' upstream exons and the extent of the 3' UTR from genomic sequence has been extremely difficult. A combination of splice site predictions and similarity to other olfactory receptors has allowed some investigators to predict 5' exon locations for around 15 genes [25,28]. Experimental validation shows that some, but not all, predictions are accurate [24,25]. The total number of olfactory receptors for which gene structure is known is vastly increased by our study.

In this report, we describe the isolation and analysis of over 1,200 cDNAs representing 419 odorant receptor genes. We screened a mouse olfactory epithelium library with degenerate olfactory receptor probes and obtained 5' end sequences (ESTs) from purified cDNAs. These clones confirm olfactory expression of over 400 olfactory receptors, provide their gene structure, demonstrate that not all olfactory receptors are expressed at the same level and show that most olfactory receptor genes have multiple transcriptional isoforms.

Table 1

Number of olfactory receptors in old (Release 12) and new (Release 13) Celera mouse genome assemblies

	Olfactory receptors in Release 12 mouse genome assembly [1]	Olfactory receptors in Release 13 mouse genome assembly
Total number of olfactory receptor sequences	1,468	1,490
Number of partial sequences (at end or gap in Celera scaffold)	262/1,468 (18%)	96/1,490 (6%)
Number of full olfactory receptor sequences	1,206/1,468 (82%)	1,394/1,490 (94%)
Interrupted by repeat sequence	134/1,206 (11%)	117/1,394 (8%)
Contains frameshift or stop codon	206/1,206 (17%)	170/1,394 (12%)
Intact ORF	866/1,206 (72%)	1,107/1,394 (79%)
Intact class I	104/866 (12%)	124/1,107 (11%)
Intact class II	762/866 (88%)	983/1,107 (89%)

Results

At least 419 mouse olfactory receptor genes are expressed in the olfactory epithelium

We have isolated 1,264 olfactory receptor cDNA clones, which together confirm the olfactory epithelial expression of 419 annotated olfactory receptor genes. We used low-stringency hybridization with degenerate olfactory receptor DNA probes to screen around 4,100,000 plaque-forming units (pfu) of an adult mouse olfactory epithelium cDNA library and around 640,000 pfu of an embryonic olfactory epithelium library. We obtained sequences from 1,715 hybridization-positive cDNAs following secondary screens to isolate single clones. Of these clones, 1,264 yielded olfactory receptor-containing sequences. The 26% false-positive rate is a consequence of using low-stringency hybridization to obtain maximal sensitivity. Continuing the screen would have resulted in cDNAs from additional olfactory receptors, but we reached a point of limiting returns: our final batch of 45 olfactory receptor-positive sequences represented 33 different olfactory receptors, of which only four had not been encountered previously in our screen.

Sequence analysis shows that the libraries are of high quality. Firstly, directional cloning was successful: only eight out of 1,430 cDNA sequences with any protein homology matched that protein on the reverse strand. Secondly, genomic contamination is rare: when the 83 olfactory receptor-containing sequences that had a 5' UTR of 400 bp or longer were aligned to genomic sequence, 80 spliced across at least one intron, leaving a maximum of three clones (3.6%) that potentially represent genomic contamination of the libraries. Thirdly, most clones are of a reasonable length: although we did not determine whether clones are full-length, 881 of 1,264 (70%) olfactory receptor cDNAs contain the gene's start codon and at least some 5' UTR sequence.

In order to match cDNAs to their genomic counterparts, first we updated our catalog of mouse olfactory receptor genes [1] based on Celera's most recent genome assembly (Release 13) [29]. Previous reports of the mouse olfactory receptor repertoire [1,2] were based on the Release 12 assembly. Release 13 consists of fewer, longer scaffold sequences containing fewer, smaller gaps than Release 12. Using the BLAST-based methods detailed previously [1], we identified 1,490 olfactory receptor sequences in the new assembly, including 1,107 intact olfactory receptor genes (compared to 866 intact olfactory receptors in the old assembly) reflecting the reduced sequence error rate and increased coverage of the new assembly (Table 1). We created a local database of genomic sequences including all olfactory receptor loci and 0.5 Mb flanking sequences (if available) and compared each cDNA sequence to this 'olfactory subgenome' database using sim4 [30].

cDNAs were assigned to individual genes based on their best match to an olfactory receptor coding region or its upstream region (see Materials and methods). Of the 1,264 olfactory receptor cDNAs, 1,176 matched a total of 419 olfactory receptor genes; the remaining cDNAs either matched an olfactory receptor below our 96% nucleotide identity threshold or had ambiguous matches encompassing more than one olfactory receptor gene region (see below).

A class I olfactory receptor degenerate primer broadens phylogenetic distribution of confirmed olfactory receptor genes

Previous analyses of the mammalian olfactory receptor family define two major phylogenetic clades, referred to as class I and II olfactory receptors, and suggest that class I olfactory receptors are more similar to fish olfactory receptors than are class IIs [5]. Figure 1 illustrates the phylogenetic diversity of our cDNA collection, showing that we have confirmed expression of at least one olfactory receptor gene in each major clade of the class II olfactory receptor genes, or 391 out of 983 (40%) of all intact class II olfactory receptor genes where full-length genomic sequence data are available (blue branches). The screen thus appears relatively unbiased in its coverage of class II olfactory receptors. However, our random screen provided cDNAs for only two out of 124 intact, full-length class I olfactory receptors. In an attempt to broaden the phylogenetic coverage of our hybridization screen, we used additional degenerate probes on the adult library and screened an

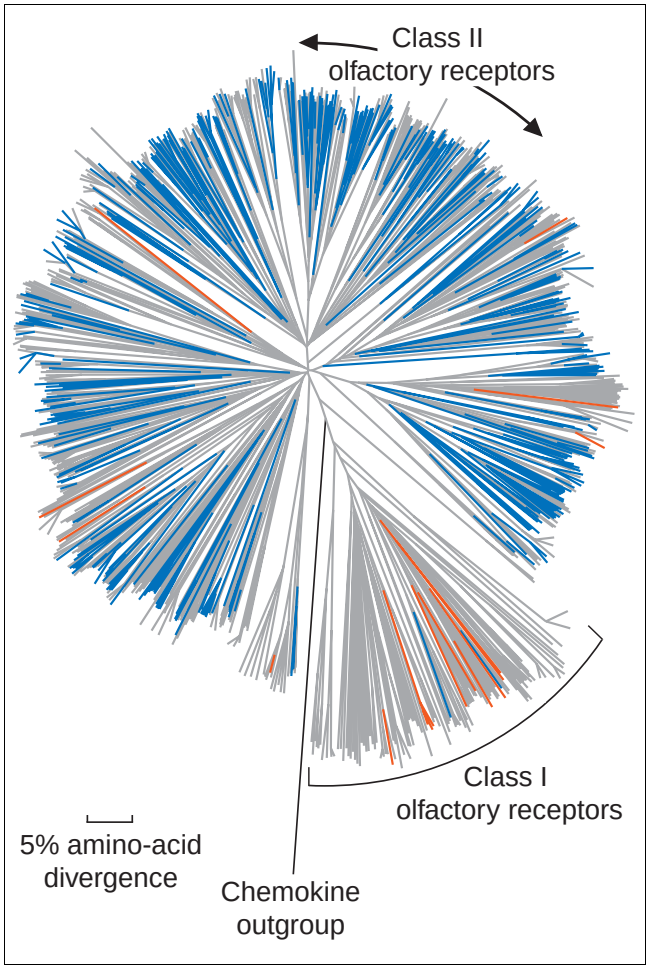

Figure 1
Olfactory receptor genes whose expression in the mouse olfactory epithelium was confirmed in this study. Genes whose expression has been confirmed by our cDNA screen are colored blue on a phylogenetic tree of 1,107 intact mouse olfactory receptors. Genes whose expression was confirmed by PCR methods are colored red (genes listed in Additional data file 1, available with the online version of this article, were confirmed by specific PCR of the cDNA library or reverse-transcribed RNA, and genes confirmed using the class I degenerate primer for RT-PCR are AY317681, AY317698, AY317700, AY317767, AY317773, AY317774, AY317797 and AY317923). Other olfactory receptors are colored gray, and a chemokine outgroup is colored black. Class I olfactory receptors are bracketed, and the remaining olfactory receptors are class II.

embryonic library (Table 2). These experiments did not increase the diversity of clones identified (not shown).

This severe class I underrepresentation could be due to experimental bias - a consequence of using degenerate primers to create our hybridization probe. Alternatively, class I genes might be expressed at extremely low levels in the olfactory epithelium. In order to determine whether class I olfactory receptors are expressed in the olfactory epithelium, we designed a reverse-strand degenerate primer to recognize a motif in transmembrane domain 7 (PP{V/M/A/T}{F/L/I/M}NP) enriched among class I olfactory receptor sequences. Most of the motif is shared among all olfactory receptors, but the first proline residue (at the primer's 3' end) is found in 121 out of 124 (98%) intact class I genes compared to only 37 out of 983 (4%) intact class II genes. When combined with another olfactory receptor degenerate primer, P26 [17], this primer preferentially amplifies class I olfactory receptors from mouse genomic DNA: of 33 sequenced, cloned PCR products, 17 represented seven different class I olfactory receptors, six represented three different class II olfactory receptors, and ten represented five different non-olfactory receptor contaminants. Degenerate PCR, cloning, and sequencing from reverse-transcribed olfactory epithelium RNA showed that at least seven class I olfactory receptors are expressed, as well as one additional class II gene (colored red in Figure 1). However, no products could be obtained from the adult or the fetal olfactory epithelium cDNA libraries using the class I primer, suggesting that the libraries contain very low levels of class I olfactory receptors. We also confirmed expression of nine additional olfactory receptors (three class I and six class II, colored red in Figure 1) from subclades that were poorly represented in our cDNA screen using gene-specific primer pairs to amplify cDNA library or reverse-transcribed RNA templates.

For two of the class I genes we had shown to be expressed, we determined relative transcript levels using quantitative RT-PCR (see below). Expression levels were similar to those observed for genes that were represented in our cDNA collection, suggesting that class I olfactory receptors are not underrepresented in the olfactory epithelium, and that the dearth of class I cDNAs in our screen is likely to be due to bias in the libraries and/or hybridization probes.

Some olfactory receptor genes are expressed at higher levels than others

Our cDNA screen suggests that some olfactory receptor genes are expressed at higher levels than others. If all olfactory receptor genes were expressed at equal levels, and our screen and library were unbiased in their coverage of the class II olfactory receptors, the number of cDNAs detected per class II olfactory receptor should follow a Poisson distribution, calculated based on the assumption that all 983 intact class II olfactory receptors have an equal chance of being represented in the screen, but that class I olfactory receptors and pseudogenes cannot be found (Figure 2). We calculate a low probability (approximately one in 28) that we would observe any gene with at least eight matching cDNAs in the set of 1,176 cDNAs we assigned to single olfactory receptor sequences. However, for 17 olfactory receptors, we found ten or more matching cDNAs, suggesting that they might be expressed at higher levels than other olfactory receptor genes (Figure 2). The two genes for which we found most cDNAs (AY318726/MOR28 and AY318727/MOR10) are genomically adjacent and in the well-studied olfactory receptor cluster next to the T-cell receptor α/δ locus [18,31].

Table 2

Summary of cDNA screen for each library and probe.

Library	Probe	Number of plaques screened (× 10^3)	Number of sequences obtained	Number of real olfactory receptor sequences	True-positive rate	Olfactory receptor clone frequency	Number of olfactory receptor genes represented
Embryonic	OR5B_OR3B_40	640	58	37	64%	1/17,300	27
Adult	OR5B_OR3B_40	2,850	1,450	1,138	78%	1/2,500	394
	P24_P28_40 and TM3deg1_P28_45	200	23	3	13%	1/66,700	3
	P26_P27_45	700	135	58	43%	1/12,100	35
	P24_P28_45	200	39	22	56%	1/9,100	19
	OR5B_OR3B_45	150	10	6	60%	1/250,000	5
Total		4,740	1,715	1,264	74%	1/3,800	419

Probe names comprise the names of the two primers and the annealing temperature used during PCR to generate the probes, separated by underscores

Quantitative RT-PCR of six olfactory receptors confirms that expression levels do indeed vary considerably between genes. We used quantitative (real-time) PCR to measure olfactory epithelium transcript levels of six olfactory receptor genes and the ribosomal S16 gene in three mice of the same inbred strain (Figure 3). These genes include two olfactory receptors with more than 20 matching cDNAs, two with one or two matching cDNAs and two class I olfactory receptors with no matching cDNAs. In these assays, we measure transcript level per genomic copy of the gene by comparing how well a gene-specific primer pair amplifies reverse-transcribed RNA, relative to a standard curve of amplification of mouse genomic DNA. We find that expression levels can vary by almost 300-fold between genes (for example, genes A and D, Figure 3). However, cDNA numbers are not a good indicator of expression level, a discrepancy that is likely to be due to bias in the screen (we used degenerate primers to make the probes, which will favor some genes over others) and in the libraries (oligo-dT priming will favor genes with shorter 3' UTRs). For example, we observe large expression differences in all three mice between two genes for which similar numbers of cDNAs were found (genes A and B, Figure 3), and conversely, similar expression levels between two genes with a ten-fold difference in number of cDNAs found (genes B and C, Figure 3). Expression levels are mostly consistent between different mice: we find similar expression-level differences between olfactory receptor genes in all three mice examined (that is, the rank order of the six genes is similar among the three mice), although there is variation in expression level of some genes between mice (for example, gene E, Figure 3).

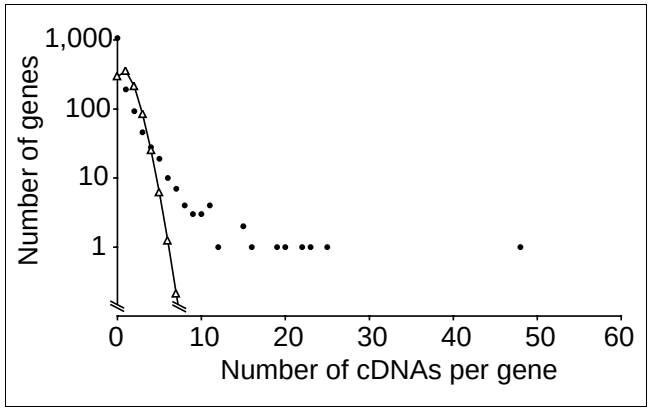

Figure 2
The cDNA screen suggests different expression levels for different olfactory receptors. Distribution of number of cDNAs observed (dots) and expected (triangles, line) per olfactory receptor gene among 1,176 olfactory receptor cDNAs identified, based on a Poisson distribution.

In situ hybridization (Figure 4) shows that increased numbers of expressing cells account for some, but not all, of the difference in transcript levels between two of the genes tested by real-time PCR (genes A and D in Figure 3). We hybridized alternate coronal serial sections spanning an entire olfactory epithelium of a young mouse (P6) with probes for gene A and gene D. Southern blot and BLAST analyses show that both probes are likely to hybridize to their intended target genes and no others (not shown). Gene A is expressed in zone 4 of the epithelium according to the nomenclature of Sullivan *et al.* [32] (Figure 4a). The expression pattern of gene D does not correspond to any of the four 'classical' olfactory epithelial zones [14,15,32]: positive cells are found in regions of endoturbinates II and III and ectoturbinate 3, resembling the expression pattern seen previously for the OR37 subfamily

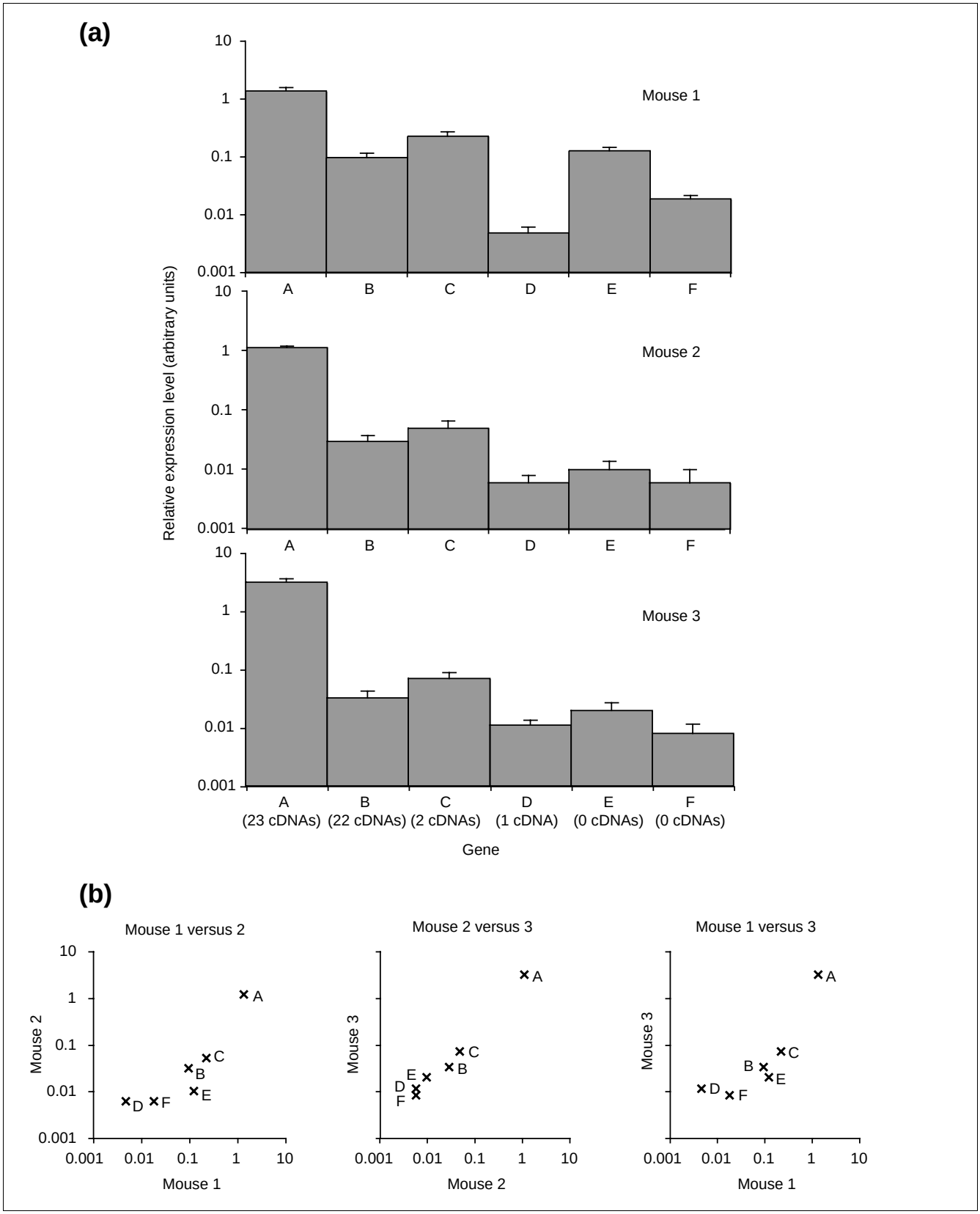

Figure 3 *(see legend on next page)*

Figure 3 *(see previous page)*
Differential expression levels among six olfactory receptor genes determined by quantitative PCR. **(a)** Expression levels of olfactory receptor genes can vary by almost 300-fold (for example, genes A and D). Relative expression levels of six selected olfactory receptor genes (A, AY318555; B, AY318107; C, AY318644; D, AY317365; E, AY317773; and F, AY317797) were determined in olfactory epithelium RNA samples from three mice. Expression levels for each gene were first determined relative to a standard curve made using mouse genomic DNA templates, and then values for each mouse were normalized so that a housekeeping gene, ribosomal S16, had a value of 1 (arbitrary units) (not shown). Error bars show one standard deviation (six replicate reactions). Genes E and F (AY317773 and AY317797) are class I olfactory receptors. Numbers of cDNAs observed in our screen are shown under each gene name. **(b)** Expression levels of each gene are similar, with some variation, among the three mice sampled. Graphs show pairwise comparisons between the three mice sampled, with relative expression levels (arbitrary units) in one mouse plotted along the *x*-axis and in a second on the *y*-axis.

and ORZ6 olfactory receptors [33,34] (Figure 4b). Counting the total number of positive cells in alternate sections across the entire epithelium, we find that gene A is expressed in 2,905 cells, about 12 times more cells than gene D, which is expressed in a total of 249 cells. This 12-fold difference in numbers of expressing cells does not account for the almost 300-fold difference in RNA levels observed by real-time PCR, implying that the transcript level per expressing cell for gene A is about 25 times higher than transcript level in each expressing cell for gene D. We note that hybridization intensities per positive neuron appear stronger for gene A than gene D after comparable exposure times, in accordance with the idea that transcript levels are higher per cell. Thus, we suggest that expression in more cells and in higher levels per cell together account for the almost 300-fold higher olfactory epithelial RNA levels of gene A relative to gene D (Figure 3).

Most olfactory receptor genes have several transcriptional isoforms

Our cDNA collection reveals that at least two thirds of the olfactory receptors sampled show alternative splicing of their 5' untranslated exons. Using a custom script to process sim4 alignments of cDNA and genomic sequences, we find two to eight different splice forms for 85 (45%) of the 191 genes for which we have had some opportunity to observe alternate splicing (that is, a minimum of two cDNAs, at least one of which is spliced), and 55 (67%) of the 82 genes for which we have four or more cDNAs (and thus a higher chance of observing any alternate splicing) (Figure 5). These alternative splice events are almost all restricted to the 5' UTR and include exon skipping and alternate splice-donor and -acceptor use.

At least half of the olfactory receptors represented in our cDNA collection utilize more than one polyadenylation site, resulting in alternative 3' UTR isoforms. We have crudely estimated 3' UTR size for 1,169 cDNA clones by combining approximate insert size information with 5' sequence data. More than one 3' UTR isoform is predicted for 43 of the 77 (56%) genes for which there are at least four cDNAs with 3' UTR size information. We confirmed the alternative polyadenylation isoforms of four out of five selected genes by sequencing the 3' end of 14 cDNA clones. These 14 sequences also revealed one cDNA where the poly(A) tail was added 27 bp before the stop codon, and another where an intron was spliced out of the 3' UTR, contrary to the conventional stereotype of olfactory receptor gene structure.

A subset of olfactory receptors shows unusual splicing

We identified 62 cDNAs (5% of all olfactory receptor clones) from 38 intact olfactory receptors and one olfactory receptor pseudogene where a splice site within the protein-coding region is used. For two genes (top two cDNAs, Figure 6), the predicted protein appears to be an intact olfactory receptor with three or ten amino acids, including the initiating methionine, contributed by an upstream exon. A similar gene structure was described previously for a human olfactory receptor [25]. One of these two mouse genes has no start codon in its otherwise intact main coding exon. The unusual splicing thus rescues what would otherwise be a dysfunctional gene. In most cases (60 out of 62 cDNAs), the unusual transcript appears to be an aberrant splice form - the transcript would probably not encode a functional protein because the splice introduces a frameshift or removes conserved functional residues (Figure 6). For two clones (bottom two cDNAs, Figure 6), exon order in the cDNA clone is inconsistent with the corresponding genomic sequence. It is difficult to imagine what kind of cloning artefact resulted in these severely scrambled cDNAs: we suggest that they derive from real but rare transcripts. However, their low frequency in our cDNA collection suggests that splicing contrary to genomic organization does not contribute significantly to the olfactory receptor transcript repertoire. For 21 of the 26 genes for which unusually spliced cDNAs were found, we also observe an alternative ('normal') isoform that does not use splice sites within the coding region. (For the remaining 13 of the 3' genes showing odd splicing, we have identified only one cDNA so have not determined whether normal isoforms are present.)

We were intrigued both by previous reports of splicing of human olfactory receptors near the major histocompatibility complex (MHC) cluster, where several genes splice over long distances to a common upstream exon [26,27] and by the idea that olfactory receptor transcriptional control could be achieved by DNA recombination mechanisms, perhaps with the result that transcripts would contain some sequence from another locus. We therefore verified that the entire sequence of each olfactory receptor EST matches the corresponding gene's genomic 'territory' (defined for this purpose as from 1

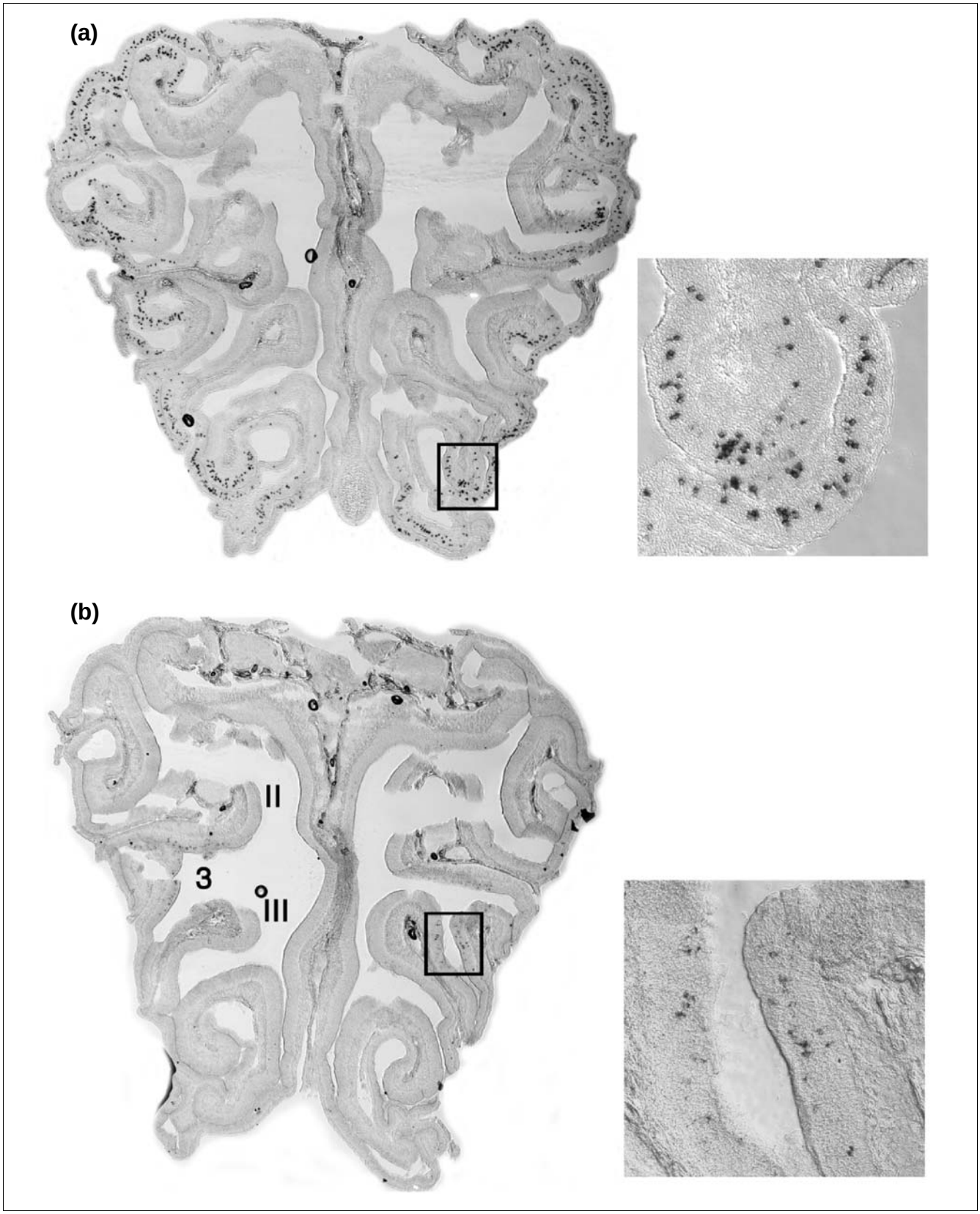

Figure 4 *(see legend on next page)*

Figure 4 *(see previous page)*
Olfactory receptors showing different expression levels. Different expression levels of one pair of olfactory receptors is due to different numbers of expressing cells and different transcript levels per cell. RNA *in situ* hybridization with digoxigenin-labeled probes for **(a)** gene A (AY318555) and **(b)** gene D (AY317365) on coronal sections of the olfactory turbinates of an adult mouse, shown at low magnification and inset (boxed) at high magnification. Endoturbinates II and III and ectoturbinate 3 are labeled in (b).

kb after the preceding gene to 1 kb after the gene's stop codon). We found no cDNAs where introns encompassed other olfactory receptor genes, as reported for olfactory receptors in the human MHC region [26,27]. Six cDNAs do extend further than a single gene's 'territory' and appear not to be artifacts of the sequencing or analysis process. In each of these cases, the clones use splice sites within the 3' UTR and thus extend further than the arbitrary 1 kb downstream of the stop codon. Five of these six cDNAs also use splice-donor sites within the coding region and encode disrupted olfactory receptors (Figure 6). In the sixth cDNA, a 2.6-kb intron is spliced out of the 3' UTR, leaving the coding region intact.

If olfactory receptor transcriptional control is achieved by DNA recombination, the beginning of each transcript might derive from a donated promoter region, with the rest of the transcript coming from the native ORF-containing locus. In order to examine the recombination hypothesis, we analyzed 115 cDNA clones for which sim4 failed to align 20 bp or more to the corresponding genomic locus. In most cases, the missing sequence was explained by gaps in the genomic sequence or by matches that fell below our percent identity-based cutoff for reporting matches. For three cDNAs (from three different olfactory receptors), we found that the missing piece of sequence matched elsewhere in the genome. Comparison with the public mouse genome assembly confirmed the distant matches. With such a small number of cDNAs exhibiting a possible sign of DNA recombination (a sign that could also be interpreted as chimeric cDNA clones), we conclude that such rearrangement is unlikely to occur. However, the possibility remains that DNA recombination is responsible for olfactory receptor transcriptional regulation, with the donated region contributing only promoter sequences but no part of the transcript.

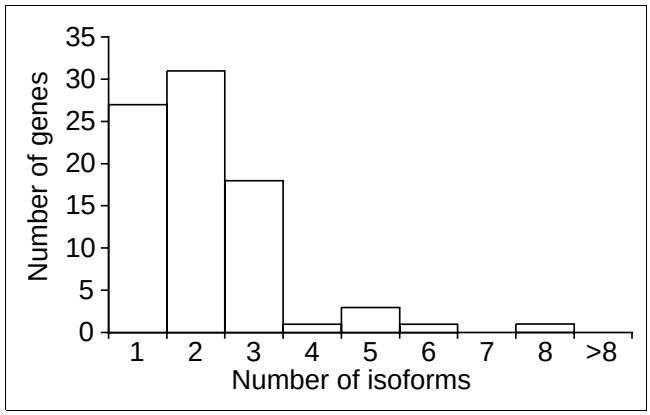

Figure 5
Many olfactory receptor genes show alternate splicing. Distribution of the number of transcriptional isoforms observed for the 82 olfactory receptors for which we have identified at least four cDNAs.

Both unclustered olfactory receptors and olfactory receptor pseudogenes can be expressed

We were interested in whether olfactory receptors need to be part of a cluster in the genome in order to be transcribed, or if the clustered genomic organization of olfactory receptors is simply a consequence of the fact that local duplication is the major mechanism for expanding the gene family [1]. 'Singleton' olfactory receptors (defined as full-length olfactory receptors without another olfactory receptor within 0.5 Mb) are more often pseudogenes than are olfactory receptors in clusters (8 out of 16 versus 271 out of 1,358; $\chi^2 = 8.8$, $P < 0.005$). Of the eight intact singleton olfactory receptors, two have matching cDNAs in our collection, a similar proportion as found for olfactory receptors in clusters, showing that clustering is not an absolute requirement for olfactory receptor expression. However, it is possible these two expressed singleton genes are part of 'extended' olfactory receptor clusters - their nearest olfactory receptor neighbors are 1.7 Mb and 2.6 Mb away, respectively.

We also find that some olfactory receptor pseudogenes are expressed, albeit with a lower probability than intact olfactory receptors. Considering the 1,392 olfactory receptor gene sequences for which reliable full-length data are available, 15 out of 285 (5%) apparent pseudogenes are represented in our cDNA collection, compared to 393 out of 1,107 (36%) intact olfactory receptors. However, three of these 15 'expressed pseudogenes' are intact genes in the public mouse genome sequence. The defects in Celera's version of these genes may be due to sequencing errors or true polymorphism. Publicly available mouse sequence confirms that 11 of the 12 remaining expressed pseudogenes are indeed pseudogenes. No public sequence matches the 12th 'expressed pseudogene' with 99% identity or more.

Discussion

We have identified and sequenced 1,264 odorant receptor cDNAs from 419 olfactory receptor genes, confirming their expression in the olfactory epithelium. We have thus validated the similarity-based prediction of over one-third of the intact olfactory receptor genes annotated in the mouse

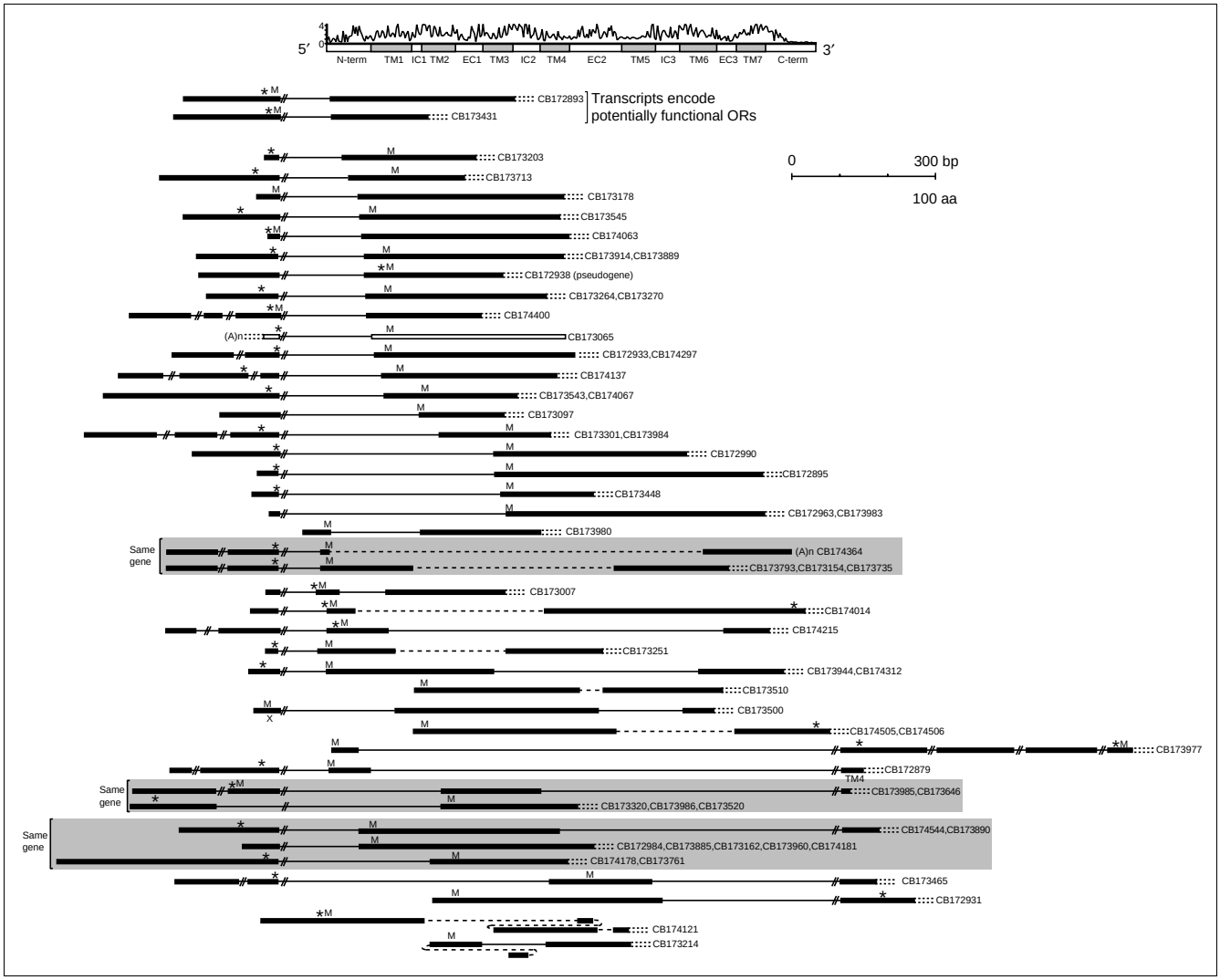

Figure 6

Sixty-two olfactory receptor cDNAs use splice sites within the coding region. The bar at the top represents an alignment of all olfactory receptor proteins, with transmembrane (TM) regions shaded gray and intracellular (IC) and extracellular (EC) loops in white. Above the bar, the jagged line plots information content [51] for each alignment position, with higher values representing residues conserved across more olfactory receptors. cDNAs with atypical splicing are plotted below, aligned appropriately to the consensus representation. Genbank accessions for each cDNA are shown on the right, and where more than one clone represents the same isoform, both names are given, but a composite sequence is drawn. Multiple isoforms from the same gene are grouped by gray background shading. Thick black lines represent cDNA sequence, and thin lines represent intronic sequence (with diagonal slash marks if not drawn to scale). The uppermost two cDNAs encode potentially functional olfactory receptors. A single cDNA drawn as white boxes (CB173065) is cloned into the vector in the reverse orientation. Introns that result in a frameshift relative to the olfactory receptor consensus are drawn as single dashed lines. The first in-frame methionine in the cDNA is marked with an 'M', and the first stop codon 5' to this methionine (if any) is marked with *. Most sequences are incomplete at the 3' end, as represented by paired dotted lines, although two sequences (CB174400 and CB174364), marked with '(A)n', contain the cDNA's poly(A) tail. The 'X' on sequence CB173500 marks an exon that does not align with genomic sequence near the rest of the gene or anywhere else in Celera's mouse genome sequence, and 'TM4' on sequence CB172879 notes an exon that matches to the reverse-complement of the fourth transmembrane domain of the next downstream olfactory receptor gene. For the two lowermost cDNAs, exon order in the cDNA clone is inconsistent with the corresponding genomic sequence, as represented by the curved intron lines.

genome [1,2], thereby vastly increasing the proportion of the family for which experimental evidence of olfactory function is available. We have not found cDNAs for all olfactory receptor genes or an even phylogenetic distribution of cDNAs, probably because the libraries and/or our screen are biased toward certain olfactory receptor subfamilies. Using RT-PCR with both degenerate and specific primers, we have confirmed olfactory expression of a number of additional olfactory receptors, bringing the total number of olfactory receptor genes verified in this study to 436, and ensuring that almost all phylogenetic clades have at least one representative with evidence of olfactory function.

Results of our cDNA library screen suggested that some olfactory receptors are expressed at significantly higher levels than others. We used quantitative PCR to show that expression levels are indeed highly variable, with one olfactory receptor expressed at almost 300 times the level of another. Higher expression levels could be due to increased transcript number per cell and/or a greater number of olfactory neurons 'choosing' those genes. For one pair of genes we tested, expression level differences appear to be due to both factors. It would be interesting to collect data for additional genes to determine how the numbers of expressing cells and transcript levels per cell vary across the olfactory receptor family. Data from a number of previous studies also show that different olfactory receptor genes, or even copies of the same olfactory receptor transgene in different genomic locations are expressed in different numbers of cells [14,18,35], but do not address the issue of transcript level per cell. The fact that some genes are chosen more frequently, and when chosen may be expressed at higher levels per cell, is intriguing given each olfactory neuron's single-allele expression regime. The observation of unequal expression leads to a number of questions. It is known that each olfactory receptor is expressed in one of four zones of the olfactory epithelium [14,15]; do some zones choose from a smaller olfactory receptor sub-repertoire and thus express each olfactory receptor in a larger number of cells? We note that several apparently highly expressed olfactory receptors (gene A, this study, and MOR10 and MOR28 [36]) are expressed in zone 4 of the olfactory epithelium. Does activity-dependent neuronal competition [37] contribute to increased representation of the olfactory receptors that respond to common environmental odorants? Do the favored olfactory receptors have stronger promoter sequences? Are some olfactory receptor mRNAs more stable than others, leading to higher transcript levels per expressing cell? Are the favored olfactory receptors in more open chromatin conformation or more accessible genomic locations? Transcription of apparent 'singleton' olfactory receptor genes (0.5 Mb or more from the nearest other olfactory receptor gene) suggests that there is no absolute requirement for genomic clustering for an olfactory receptor to be transcribed, consistent with observations that small olfactory receptor transgenes can be expressed correctly when integrated outside native olfactory receptor clusters [35]. However, the high pseudogene count among singleton olfactory receptor genes (50%, versus 20% for clustered olfactory receptor genes) suggests that not all genomic locations are favorable for olfactory receptor gene survival, perhaps due to transcriptional constraints. It is also possible that evolutionary factors may be responsible for reduced pseudogene content of clustered olfactory receptors - gene conversion between neighboring olfactory receptors could rescue inactivating mutations in clustered genes, but not singletons. Before these questions about olfactory receptor gene choice can be answered, it will be important to measure expression levels of a larger number of genes, perhaps using an olfactory receptor gene microarray.

Our study provides at least partial data about the upstream transcript structures of over 300 olfactory receptor genes. These data provide tentative locations of a large set of promoter regions, allowing computational searches for shared sequence motifs that might be involved in the intriguing transcriptional regulation of olfactory receptors. However, given that not all cDNAs are full-length clones, some of these candidates will not be true promoter regions. The 5' UTR sequences we obtained will also aid in the design of experimental probes, for example, for *in situ* hybridizations or to immobilize on an olfactory receptor microarray. One of the challenges of such an array will be to design unique probes with which to represent each gene. Often, the coding region of olfactory receptors is highly similar between recently duplicated genes. Many pairs of similar olfactory receptors show more sequence divergence in the UTRs than the protein-coding region (J.Y., unpublished observations). The UTRs would therefore make a better choice of sequence from which to design unique oligonucleotides to distinguish closely related olfactory receptor genes. Locations of these regions in genomic sequence are difficult to predict - our study provides 5' UTR sequences of 343 genes and the approximate 3' UTR length for 399 olfactory receptor genes. Probe design must also account for the multiple transcriptional isoforms observed for many olfactory receptors - depending on the question being asked, probes could be designed in shared sequence to determine the total level of all isoforms, or in unique exons to measure the level of each isoform separately.

We find that the majority of the olfactory receptors, like most non-olfactory receptor genes [38,39], are transcribed as multiple isoforms, involving alternative splicing of 5' untranslated exons and alternate polyadenylation-site usage. The act of splicing itself may be important for efficient mRNA export from the nucleus [40] or to couple olfactory receptor coding regions with genomically distant promoters. The exact nature of the spliced transcript might be unimportant, such that several isoforms might be produced simply because multiple functional splice sites are available. Alternatively, the multiplicity of transcriptional isoforms might have functional significance, as UTRs may contain signals controlling mRNA stability, localization or degradation [41,42].

Our study shows that about 5% of olfactory receptor transcripts do not fit the current notion of olfactory receptor gene structure. Occasionally, an intron is spliced out of the 3' untranslated region. A number of cDNAs use splice sites within the olfactory receptor's ORF, meaning that their protein product is different from that predicted on the basis of genomic sequence alone. In two such cases, the transcript would encode a functional olfactory receptor, with the initiating methionine and first few amino acids encoded by an upstream exon, as has been observed previously for a subtelomeric human olfactory receptor gene [25]. Such within-ORF splicing might increase protein-coding diversity, although, given the small number of genes involved, splicing

is unlikely to significantly affect the functional receptor repertoire. Most of the atypical splice forms we observe appear to encode non-functional transcripts, containing frameshifts or lacking a start codon or other functional residues conserved throughout the olfactory receptor family. These nonfunctional transcripts are probably aberrant by-products of the splicing system [43] that have not yet been degraded by RNA surveillance systems [40,41]. The neurons expressing these aberrant transcripts might also make normal transcripts for the same genes and thus produce a functional olfactory receptor. Alternatively, the unusual transcriptional regulation of olfactory receptors might ensure that only one splice isoform is expressed per cell (unlikely, but possible if an RNA-based feedback mechanism operates), thus condemning cells expressing these aberrant isoforms to be dysfunctional.

We also observe transcripts from a small number of olfactory receptor pseudogenes, as previously described for three human olfactory receptor pseudogenes [26,44]. Although many fewer pseudogenes than intact genes were represented in our cDNA collection, some neurons in the olfactory epithelium evidently express disrupted olfactory receptors and thus might be unable to respond to odorants or to correctly innervate the olfactory bulb. Wang, Axel and coworkers have shown that an artificial transgenic olfactory receptor gene containing two nonsense mutations can support development of an olfactory neuron, but that pseudogene-expressing neurons fail to converge on a glomerulus in the olfactory bulb [45]. By analogy with an olfactory receptor deletion mutant [45], it is likely that most pseudogene-expressing neurons die or switch to express a different olfactory receptor gene, leaving a small number of pseudogene-expressing neurons in adult mice, but at greatly reduced levels compared to neurons expressing intact olfactory receptors.

Conclusions

Our study has provided an olfactory receptor cDNA resource representing over one-third of the olfactory receptor gene family. We have thus established over 400 annotated olfactory receptor genes as having olfactory function. The sequences we generated demonstrate that the majority of the olfactory receptor gene family has multiple transcriptional isoforms. Most olfactory receptor transcripts encode functional receptor proteins, with rare exceptions. We show that individual olfactory receptor genes can have vastly different expression levels, an intriguing finding in light of the unusual one-neuron one-gene transcriptional regime of the olfactory epithelium. Our results and the sequences we provide will facilitate future global studies of the mechanisms and dynamics of olfactory receptor gene expression.

Materials and methods

Identification of olfactory receptor cDNAs

An adult mouse cDNA library made from the olfactory epithelium of a single animal was provided by Leslie Vosshall (Rockefeller University, New York, NY, USA), and an embryonic library (made from the olfactory epithelia of several E16.5-E18.5 embryos) was provided by Tyler Cutforth (Columbia University, New York, NY, USA). Both libraries were oligo-dT primed and directionally cloned into the lambdaZAP-XR vector (Stratagene, La Jolla, CA, USA). The adult library has a complexity of 6.5×10^6 primary clones, and the embryonic library has a complexity of 1.65×10^6. Libraries were amplified to give titers of 5×10^9 pfu/ml (adult) or 2×10^{10} pfu/ml (embryonic). Hybridization probes were made by degenerate PCR of mouse genomic DNA, in a fashion similar to those described previously [1], with primer pairs and annealing temperatures given in Table 2. Low-stringency hybridization conditions were as described [1]. Clonally-pure plaques were obtained through secondary screens using the same probe as the corresponding primary screen. PCR with vector primers (M13F/R) was performed to prepare sequencing templates. cDNA size estimates were obtained by agarose gel electrophoresis, and inserts were sequenced from the 5' end using the M13R primer and big-dye terminator chemistry according to ABI's protocols (Applied Biosystems, Foster City, CA, USA). In order to obtain 3' sequence, selected phage clones were converted to plasmid stocks following a scaled-down version of Stratagene's *in vivo* excision protocol. Plasmid DNA gave better 3'-end sequence than PCR products, which often suffered from polymerase stuttering through the poly(A) tract.

cDNA sequences and associated information are available through dbEST (Genbank accessions CB172832-CB174569) and our olfactory receptor database [46]. The updated olfactory receptor gene catalog is available through Genbank (accessions AY317244-AY318733). Throughout the manuscript, genes are referred to by their Genbank accession numbers.

Sequence analysis

cDNA sequences were base-called and quality-trimmed using phred (trim_cutoff = 0.05) [47], and vector sequences were removed using cross_match [48]. Any sequences of less than 50 bp after trimming were discarded. 3' UTR lengths were estimated by combining approximate insert sizes determined by PCR with 5' sequence data where possible (if the 5' sequence did not extend into the coding region we could not estimate 3' UTR size). We counted cDNAs from a given gene as showing alternative polyadenylation site usage if 3' UTR length estimates varied by at least 400 bp - smaller variation could be real, but may not be distinguishable from error in our size estimates.

To assign cDNAs to their corresponding olfactory receptor genes, we first defined a genomic 'territory' for each gene,

with the following attributes: strand, start position (100 kb upstream of the start codon or 1 kb after the previous gene upstream on the same strand, whichever is closer) and end position (1 kb downstream of the stop codon). Trimmed sequences were compared with genomic sequences using sim4 [30] (settings P = 1 to remove polyA tails and N = 1 to perform an intensive search for small exons). The sim4 algorithm uses splice-site consensus sequences to refine alignments. Only matches of 96% or greater nucleotide identity were considered. RepeatMasked sequences [49] were also compared to genomic sequences; cDNA:genomic sequence pairings not found in both masked and unmasked alignments were rejected. Coordinates from the unmasked alignment were used for further analysis. Any cDNA sequence matching entirely within a territory was assigned to that gene. If a cDNA matched more than one gene territory, the best match was chosen (that is, the one with highest 'score', where score is the total of all exons' lengths multiplied by their respective percent identities). We found 27 cDNAs that spanned a larger genomic range than one gene territory and flagged them for more careful analysis. Of these, six cDNAs showed unusual splicing within the 3' UTR, but the remaining 'territory violators' were found to be artifacts of the analysis process which fell into three types. These included: cDNAs where the insert appeared to be cloned in the reverse orientation (six cDNAs); sequences from recently duplicated gene pairs, where sim4 assigned coding region and upstream exons to different members of the pair, although exons could equally well have been aligned closer to one another (six cDNAs); and artifacts due to use of sim4's N = 1 parameter (nine cDNAs). This parameter instructs the program to make extra effort to match small upstream exons, allowing a greater total length of EST sequence to be matched. However, occasionally the N = 1 parameter caused the program to assign very small sequences (1-4 bp) to distant upstream exons, when they probably match nearer to the corresponding coding sequence.

The expected distribution shown in Figure 2 was calculated using the equation $P(x) = e^{-\mu}\mu^{x}/x!$, where P(x) is the Poisson probability of observing x cDNAs per gene, and μ is the mean number of cDNAs observed per gene (μ = 1,176/983: 1,176 cDNAs matching olfactory receptor genes in our dataset and 983 intact class II olfactory receptors). In our analysis of expressed pseudogenes, we ignored two olfactory receptor pseudogenes found very near the ends of genomic sequences and thus likely to be error-prone. A protein sequence alignment of intact mouse olfactory receptors was generated using CLUSTALW [50], edited by hand, and used to produce the phylogenetic tree shown in Figure 1 using PAUP's neighbor-joining algorithm (v4.0b6 Version 4, Sinauer Associates, Sunderland, MA). The tree was colored using a custom script. Information content (the measure of sequence conservation shown in Figure 6) was calculated for each position in the alignment using alpro [51].

To determine the number of transcriptional isoforms for each gene, we examined the sim4 output for every matching cDNA in decreasing order of number of exons. The first cDNA was counted as one splice form, and for each subsequent cDNA, we determined whether exon structure was mutually exclusive to isoforms already counted. We were conservative in our definition of mutually exclusive, and thus our count represents the minimum number of isoforms represented in the cDNA collection.

RT-PCR

The olfactory epithelia were dissected from three adult female C57BL/6 mice, including tissues attached to the skull and septum. RNA was isolated using the Qiagen RNeasy midi kit (Qiagen, Valencia, CA, USA), including a DNase treatment step. First-strand cDNA was produced from 2.5 μg of RNA in a volume of 50 μl using random hexamers and Invitrogen's Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's recommendations. One-twenty-fifth of the resulting cDNA was used as template in subsequent PCR reactions. PCR amplification biased towards class I olfactory receptors was performed using degenerate primers P26 [17] and classI_R1 (5'-GGRTTIADIRYIGGNGG-3') with an annealing temperature of 44°C. The product was cloned (TA cloning kit, Invitrogen), and individual clones were sequenced. Specific PCR primers used to confirm expression of individual olfactory receptor genes are given in Additional data file 1, available with the online version of this article. Each PCR product was sequenced to confirm that the expected gene and no others had been amplified. Control reactions on a template made by omitting reverse transcriptase gave no product, indicating that the RNA preparation was uncontaminated by genomic DNA.

Relative transcript levels were estimated using real-time PCR according to Applied Biosystems' protocols, with magnesium concentration, primer pair and fluorescent probe given in Additional data file 2, available with the online version of this article. The increase in fluorescence during thermocycling was measured on an ABI PRISM 7900HT. Standard curves were constructed for each primer pair using triplicate samples of mouse genomic DNA of nine known concentrations (range 0.01-100 ng, or about 3-30,000 copies of the haploid genome). Relative expression level of each gene was determined by comparing the mean Ct (cycle where fluorescence reaches an arbitrary threshold value) obtained with six replicate samples of reverse-transcribed RNA to the standard curve for the corresponding primers. Relative RNA levels of a housekeeping gene, ribosomal S16, were measured as previously described [52]. Control reactions on template prepared by omitting reverse transcriptase amplified at a relative level of 0.03 ± 0.01 ng or less in each case. Expression measurements of the seven genes were normalized for each mouse so that S16 levels were equal to 1 (arbitrary units).

In situ hybridization

Coronal sections were cut from the olfactory epithelia of an adult mouse (Figure 4) and a young (P6) C57BL/6 mouse. RNA *in situ* hybridization was carried out as described previously [15,53] with digoxigenin-labeled antisense riboprobes specific for the 3' UTRs of genes AY318555 (0.5 kb) and AY317365 (0.5 kb). Riboprobe sequences were generated by PCR using primer pairs 5'-TCTTCCAAACCTGGACCCCCC-3' and 5'-ATCTCTCCAGCACCTTACTTG-3' for AY318555 and primer pairs 5'-TAAGATGTAAGTGATAATTTAGATTA-CAGG-3' and 5'-TTTCTGCCTCAGCTATGACAG-3' for AY317365. Hybridization was carried out in 50% formamide at 65°C, and slides were washed at high stringency (65°C, 0.2 × SSC). The probes each hybridize to only one band on a Southern blot, indicating that each probe only detects one olfactory receptor gene. BLAST analyses show that the AY318555 probe is unique in Celera's mouse genome assembly (Release 13), and that the AY317365 probe is similar to only one other genomic region. This potential cross-hybridizing region is over 10 Mb from the nearest olfactory receptor coding region and is thus highly unlikely to be included in any olfactory receptor transcript. Low-power images are composed of three overlapping micrographs (40×) assembled in Adobe Photoshop 7.0.

Additional data files

A list of the primers used to confirm the expression of olfactory receptor genes by RT-PCR and PCR from cDNA library templates can be found in Additional data file 1, available with the online version of this article. The experimental conditions used for real-time PCR can be found in Additional data file 2, available with the online version of this article.

Acknowledgements

We thank Leslie Vosshall and Tyler Cutforth for providing cDNA libraries, staff of the core facilities at the Fred Hutchinson Cancer Research Center and the University of Washington's former Department of Molecular Biotechnology for sequencing assistance, Linda Buck and Michael Schlador for comments on the manuscript and Colin Pritchard for S16 primers. The data in this paper were analyzed in part through use of the Celera Discovery System™ and Celera Genomics' associated databases. This work was supported by NIH grant R01 DC04209.

References

1. Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ: **Different evolutionary processes shaped the mouse and human olfactory receptor gene families.** *Hum Mol Genet* 2002, **11:**535-546.
2. Zhang X, Firestein S: **The olfactory receptor gene superfamily of the mouse.** *Nat Neurosci* 2002, **5:**124-133.
3. Buck L, Axel R: **A novel multigene family may encode odorant receptors: a molecular basis for odor recognition.** *Cell* 1991, **65:**175-187.
4. Krautwurst D, Yau KW, Reed RR: **Identification of ligands for olfactory receptors by functional expression of a receptor library.** *Cell* 1998, **95:**917-926.
5. Glusman G, Yanai I, Rubin I, Lancet D: **The complete human olfactory subgenome.** *Genome Res* 2001, **11:**685-702.
6. Zozulya S, Echeverri F, Nguyen T: **The human olfactory receptor repertoire.** *Genome Biol* 2001, **2:**research0018.1-0018.12.
7. Walmsley AR, Barrett MP, Bringaud F, Gould GW: **Sugar transporters from bacteria, parasites and mammals: structure - activity relationships.** *Trends Biochem Sci* 1998, **23:**476-481.
8. Parmentier M, Libert F, Schurmans S, Schiffmann S, Lefort A, Eggerickx D, Ledent C, Mollereau C, Gerard C, Perret J, *et al.*: **Expression of members of the putative olfactory receptor gene family in mammalian germ cells.** *Nature* 1992, **355:**453-455.
9. Abe K, Kusakabe Y, Tanemura K, Emori Y, Arai S: **Primary structure and cell-type specific expression of a gustatory G protein-coupled receptor related to olfactory receptors.** *J Biol Chem* 1993, **268:**12033-12039.
10. Yuan TT, Toy P, McClary JA, Lin RJ, Miyamoto NG, Kretschmer PJ: **Cloning and genetic characterization of an evolutionarily conserved human olfactory receptor that is differentially expressed across species.** *Gene* 2001, **278:**41-51.
11. Feingold EA, Penny LA, Nienhuis AW, Forget BG: **An olfactory receptor gene is located in the extended human beta-globin gene cluster and is expressed in erythroid cells.** *Genomics* 1999, **61:**15-23.
12. Nef S, Nef P: **Olfaction: transient expression of a putative odorant receptor in the avian notochord.** *Proc Natl Acad Sci USA* 1997, **94:**4766-4771.
13. Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H: **Identification of a testicular odorant receptor mediating human sperm chemotaxis.** *Science* 2003, **299:**2054-2058.
14. Ressler KJ, Sullivan SL, Buck LB: **A zonal organization of odorant receptor gene expression in the olfactory epithelium.** *Cell* 1993, **73:**597-609.
15. Vassar R, Ngai J, Axel R: **Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium.** *Cell* 1993, **74:**309-318.
16. Chess A, Simon I, Cedar H, Axel R: **Allelic inactivation regulates olfactory receptor gene expression.** *Cell* 1994, **78:**823-834.
17. Malnic B, Hirono J, Sato T, Buck LB: **Combinatorial receptor codes for odors.** *Cell* 1999, **96:**713-723.
18. Serizawa S, Ishii T, Nakatani H, Tsuboi A, Nagawa F, Asano M, Sudo K, Sakagami J, Sakano H, Ijiri T, *et al.*: **Mutually exclusive expression of odorant receptor transgenes.** *Nat Neurosci* 2000, **3:**687-693.
19. Lane RP, Cutforth T, Young J, Athanasiou M, Friedman C, Rowen L, Evans G, Axel R, Hood L, Trask BJ: **Genomic analysis of orthologous mouse and human olfactory receptor loci.** *Proc Natl Acad Sci USA* 2001, **98:**7390-7395.
20. Cross F, Hartwell LH, Jackson C, Konopka JB: **Conjugation in *Saccharomyces cerevisiae*.** *Annu Rev Cell Biol* 1988, **4:**429-457.
21. Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302:**575-581.
22. Borst P: **Antigenic variation and allelic exclusion.** *Cell* 2002, **109:**5-8.
23. Bulger M, Bender MA, van Doorninck JH, Wertman B, Farrell CM, Felsenfeld G, Groudine M, Hardison R: **Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters.** *Proc Natl Acad Sci USA* 2000, **97:**14560-14565.
24. Sosinsky A, Glusman G, Lancet D: **The genomic structure of human olfactory receptor genes.** *Genomics* 2000, **70:**49-61.
25. Linardopoulou E, Mefford HC, Nguyen OT, Friedman C, van den Engh G, Farwell DG, Coltrera M, Trask BJ: **Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location.** *Hum Mol Genet* 2001, **10:**2373-2383.
26. Younger RM, Amadou C, Bethel G, Ehlers A, Lindahl KF, Forbes S, Horton R, Milne S, Mungall AJ, Trowsdale J, *et al.*: **Characterization of clustered MHC-linked olfactory receptor genes in human and mouse.** *Genome Res* 2001, **11:**519-530.
27. Volz A, Ehlers A, Younger R, Forbes S, Trowsdale J, Schnorr D, Beck S, Ziegler A: **Complex transcription and splicing of odorant receptor genes.** *J Biol Chem* 2003, **278:**19691-19701.
28. Glusman G, Sosinsky A, Ben-Asher E, Avidan N, Sonkin D, Bahar A, Rosenthal A, Clifton S, Roe B, Ferraz C, *et al.*: **Sequence, structure, and evolution of a complete human olfactory receptor gene cluster.** *Genomics* 2000, **63:**227-245.
29. **Celera Genomics** [http://www.celera.com]
30. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W: **A computer program for aligning a cDNA sequence with a genomic DNA sequence.** *Genome Res* 1998, **8:**967-974.
31. Lane RP, Roach JC, Lee IY, Boysen C, Smit A, Trask BJ, Hood L:

Genomic analysis of the olfactory receptor region of the mouse and human T-cell receptor alpha/delta loci. *Genome Res* 2002, **12:**81-87.
32. Sullivan SL, Adamson MC, Ressler KJ, Kozak CA, Buck LB: **The chromosomal distribution of mouse odorant receptor genes.** *Proc Natl Acad Sci USA* 1996, **93:**884-888.
33. Kubick S, Strotmann J, Andreini I, Breer H: **Subfamily of olfactory receptors characterized by unique structural features and expression patterns.** *J Neurochem* 1997, **69:**465-475.
34. Pyrski M, Xu Z, Walters E, Gilbert DJ, Jenkins NA, Copeland NG, Margolis FL: **The OMP-lacZ transgene mimics the unusual expression pattern of OR-Z6, a new odorant receptor gene on mouse chromosome 6: implication for locus-dependent gene expression.** *J Neurosci* 2001, **21:**4637-4648.
35. Vassalli A, Rothman A, Feinstein P, Zapotocky M, Mombaerts P: **Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb.** *Neuron* 2002, **35:**681-696.
36. Tsuboi A, Yoshihara S, Yamazaki N, Kasai H, Asai-Tsuboi H, Komatsu M, Serizawa S, Ishii T, Matsuda Y, Nagawa F, *et al.*: **Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb.** *J Neurosci* 1999, **19:**8409-8418.
37. Zhao H, Reed RR: **X inactivation of the OCNC1 channel gene reveals a role for activity-dependent competition in the olfactory system.** *Cell* 2001, **104:**651-660.
38. Kan Z, Rouchka EC, Gish WR, States DJ: **Gene structure prediction and alternative splicing analysis using genomically aligned ESTs.** *Genome Res* 2001, **11:**889-900.
39. Iseli C, Stevenson BJ, de Souza SJ, Samaia HB, Camargo AA, Buetow KH, Strausberg RL, Simpson AJ, Bucher P, Jongeneel CV: **Long-range heterogeneity at the 3' ends of human mRNAs.** *Genome Res* 2002, **12:**1068-1074.
40. Maquat LE, Carmichael GG: **Quality control of mRNA function.** *Cell* 2001, **104:**173-176.
41. Wilusz CJ, Wormington M, Peltz SW: **The cap-to-tail guide to mRNA turnover.** *Nat Rev Mol Cell Biol* 2001, **2:**237-246.
42. Jansen RP: **mRNA localization: message on the move.** *Nat Rev Mol Cell Biol* 2001, **2:**247-256.
43. Kan Z, States D, Gish W: **Selecting for functional alternative splices in ESTs.** *Genome Res* 2002, **12:**1837-1845.
44. Crowe ML, Perry BN, Connerton IF: **Olfactory receptor-encoding genes and pseudogenes are expressed in humans.** *Gene* 1996, **169:**247-249.
45. Wang F, Nemes A, Mendelsohn M, Axel R: **Odorant receptors govern the formation of a precise topographic map.** *Cell* 1998, **93:**47-60.
46. **Database of mouse olfactory receptor genes and cDNAs** [http://www.fhcrc.org/labs/trask/OR/]
47. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8:**175-185.
48. **Documentation for phrap and cross_match (version 0.990319)** [http://www.phrap.org/phrap.docs/phrap.html]
49. **RepeatMasker documentation** [http://ftp.genome.washington.edu/RM/RepeatMasker.html]
50. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
51. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences.** *Nucleic Acids Res* 1990, **18:**6097-6100.
52. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: defining normal variance in mouse gene expression.** *Proc Natl Acad Sci USA* 2001, **98:**13266-13271.
53. Vassar R, Chao SK, Sitcheran R, Nunez JM, Vosshall LB, Axel R: **Topographic organization of sensory projections to the olfactory bulb.** *Cell* 1994, **79:**981-991.