High-throughput microarray detection of olfactory receptor gene expression in the mouse

Xinmin Zhang*†, Matthew Rogers*, Huikai Tian‡, Xiaohong Zhang*, Dong-Jing Zou*, Jian Liu§, Minghong Ma‡, Gordon M. Shepherd‡, and Stuart J. Firestein*¶

*Department of Biological Sciences, Columbia University, New York, NY 10027; §Section of Neurobiology, Yale University School of Medicine, New Haven, CT 06510; and ‡Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104

Communicated by Linda B. Buck, Fred Hutchinson Cancer Research Center, Seattle, WA, August 6, 2004 (received for review February 7, 2004)

The large number of olfactory receptor genes necessitates high throughput methods to analyze their expression patterns. We have therefore designed a high-density oligonucleotide array containing all known mouse olfactory receptor (OR) and V1R vomeronasal receptor genes. This custom array detected a large number of receptor genes, demonstrating specific expression in the olfactory sensory epithelium for ≈800 OR genes previously designated as ORs based solely on genomic sequences. The array also enabled us to monitor the spatial and temporal distribution of gene expression for the entire OR family. Interestingly, OR genes showing spatially segregated expression patterns were also segregated on the chromosomes. This correlation between genomic location and spatial expression provides unique insights about the regulation of this large family of genes.

Animals use their olfactory systems to monitor the chemical environment for molecules that reveal food sources or toxic substances, signal the presence of predators, and influence social and sexual behaviors. The recognition of a vast and diverse collection of compounds is accomplished initially by a large family of olfactory receptors (ORs) residing in the nasal sensory epithelium. Additionally, most mammals also have an accessory olfactory system devoted to the detection of pheromones and related odors where there are two families of receptors known as vomeronasal receptors (the V1Rs and V2Rs, each including >100 receptors; refs. 1 and 2). Additionally, OR gene expression has been reported in nonolfactory tissues, most notably in the testis (3–5). Generally, it is not known whether the OR genes expressed in nonolfactory tissues are also expressed in the olfactory epithelium.

From the mouse genome sequences, there are ≈1,100 OR genes and 170 V1R genes in the mouse genome predicted to be functional, plus ≈300 OR pseudogenes and ≈170 V1R pseudogenes (6–12). The family of V2R genes remains poorly characterized because of their complex intron–exon structure in the coding region. OR and V1R genes are distributed in tight clusters throughout the genome where the sequences of adjacent genes are generally highly related.

Although virtually all of the OR and V1R genes are known from genomic analyses, expression data are associated with only a small number of genes. A recent effort has identified ESTs from olfactory tissue for >400 OR genes (13), confirming their likely role in olfaction, but has not provided information on relative expression levels of OR genes across different tissues, ages, or spatial regions. High-density oligonucleotide arrays (14, 15) produced by using Affymetrix GeneChip technology are particularly suitable for monitoring the expression of a large number of genes simultaneously and over many conditions. However, the currently available mouse genome arrays from Affymetrix cover only a small number of OR and V1R genes, only ≈50 OR and V1R genes have been included, and the design is not optimized.

Therefore, we have designed a custom high-density specialized oligonucleotide array for mouse OR (MOR) and V1R genes. We have used the custom receptor array to address questions concerning ORs not easily investigated by using conventional one-gene-at-a-time approaches, including expression in olfactory tissue as well as age, sex, and tissue variations in OR gene expression. Also, the spatial expression of OR genes, an intriguing and little understood phenomenon, has been examined for <40 MORs.

Methods

Design of the MOR Array. We used the Celera mouse genome to design probe sets (11 pairs of perfect matches and mismatches) for 1,306 putative MORs. As detailed in *Supporting Text*, Tables 1–3, and Figs. 5 and 6, which are published as supporting information on the PNAS web site, additional consideration was given to predicted 3′ UTR regions and to potential cross-hybridization between similar genes. Complete details of the array design are given in the supporting information.

Sample Preparation. All tissues were prepared from 129/SvJ mice (The Jackson Laboratory) according to protocols described in the Expression Analysis Technical Manual (Affymetrix).

Data Analysis. Monitoring gene expression across different tissues relies crucially on having an effective method to normalize multiple arrays. When comparing different tissues, the common normalization method uses all probe sets. However, because most OR genes were selectively expressed only in OE tissue, normalization using all probe sets would lead to an overestimation of the intensity for non-OE samples. In fact, the differences between OE samples and non-OE samples on the MOR array were so large that even the invariant difference selection (IDS) algorithm, which accounts for differences in two arrays (16), did not provide satisfactory results. Normalization methods based on housekeeping genes or spike-in genes had been suggested for such situations (17), but they also have drawbacks (16). Therefore, we devised a method combining a group of stable probe sets and the IDS algorithm. The stable probe sets in the MOR array were chosen based on their low variation between samples when monitored by using the Affymetrix mouse genome U74Av2 arrays. Subsequently, when two MOR arrays were compared, the IDS algorithm was applied only to the stable probe sets to perform the normalization. This normalization method outperformed published procedures in our preliminary experiments, because it gives the best separation of OR and V1R gene expression in olfactory sensory epithelium (OE) and vomeronasal organ (VNO) tissues. All of the normalized data were analyzed by using model-based expression analysis to generate the expression values with the DCHIP software (versions 1.2 and 1.3) (18). Further details are available in the supporting information.

When comparing two kinds of samples, we used the following criteria to classify a gene as differentially expressed. A gene was considered as clearly enriched if (*i*) the expression level showed at least 2-fold change; (*ii*) statistically, the lower bound of fold change at 90% confidence interval (as determined by the DCHIP program;

Abbreviations: OR, olfactory receptor; MOR, mouse OR; OE, olfactory epithelium; VNO, vomeronasal organ; FDR, false discovery rate; V*n*R, vomeronasal receptor type *n*.

†Present address: Peptide Biology Laboratories, The Salk Institute, La Jolla, CA 92037.

¶To whom correspondence should be addressed at: 923 Fairchild, MC2438, Columbia University, 1212 Amsterdam Avenue, New York, NY 10027. E-mail: sjf24@columbia.edu.

ref. 18) was >1.5; and (*iii*) the gene received at least one P call (the present/absent call from the Affymetrix MAS5.0 program) from the arrays used. A gene was classified as marginally enriched if it did not meet the criteria for clearly enriched but (*i*) statistically the lower bound of fold change at 90% confidence interval was >1 and (*ii*) the gene had at least one P call or had an average expression level of >300 in one of the sample sets.

K-means clustering was used to analyze the zonal data. The raw data were first processed in DCHIP, and the expression values were exported to GENESPRING 6.0 (Silicon Genetics, Redwood City, CA). Each gene was divided by the median of its measurements in all samples, and the log of the ratio was used for clustering. A list of variable genes (variation across samples: 0.25 < standard deviation/mean < 10.00, and with at least one P call from all of the samples) was obtained from DCHIP, and K-means clustering was performed on this list to obtain three clusters.

***In Situ* Hybridization.** Expression of the OR genes in the epithelium was examined by *in situ* hybridization (19). Standard protocols were used throughout, and details are provided in the supporting information.

Results

Optimization and Validation of the Custom MOR Array. A custom high-density oligonucleotide array containing probes for all of the known OR and V1R genes was designed and manufactured in consultation with Affymetrix. The custom array was named "MOR array." One important feature of the MOR array is that probe sets targeting the computationally predicted 3′ UTRs (20) were included in the design. This is necessary because the oligo(dT) priming process in mRNA labeling requires the probe design to be near the 3′ end of the transcripts (14). In addition, the 3′ UTR offers the additional advantage of higher specificity than the coding regions for large gene families such as the OR genes. In the absence of full-length mRNAs for OR genes, computational prediction was used to estimate the 3′ poly(A) positions (20). The custom arrays were produced with the same technology used extensively by Affymetrix in biological and medical applications, and subject to the same quality control. However, because this MOR array is heavily biased in that it contains mostly OR genes, optimization and validation was necessary.

Extensive tests were performed to optimize the experimental conditions and data analysis for the MOR array. Because each neuron expresses only 1 of the >1,000 OR genes, each gene is expressed on average in only ≈0.1% of the neurons, resulting in very low levels of message for any individual OR gene. To increase the signal, we increased the amount of labeled cRNA during hybridization, using ≈5 times the recommended amount (70 μg vs. 15 μg). We found that this increased the number of detected genes (based on Affymetrix MAS 5.0 present calls) by 10–20%, whereas the noise level remained low (passing standard Affymetrix quality controls). For analysis of GeneChip data, there are several options (14), and we opted to use the DCHIP software (15) because of its robustness in probe level analysis, flexibility in normalization options, user-friendly interface, and integrated high-level analysis tools.

Olfactory signal transduction genes on the MOR array were used as controls. Many of the genes involved in the signal transduction pathway in olfactory sensory neurons are highly expressed in the OE, but not in the VNO or other tissues (21). As anticipated, these genes [including Golf, adenylyl cyclase III (AC3), and CNGA2 (OCNC1)] have higher expression values in OE than in VNO, testis, or lung (Fig. 1*A*). Conversely, Trp-2 is specifically expressed in the VNO, consistent with the notion that it is likely involved in signal transduction in vomeronasal neurons (22). As another example, the olfactory marker protein gene, known to be expressed largely only in olfactory tissues (23), was clearly present in both OE and VNO, but was absent in other nonolfactory tissues. Finally, the housekeeping genes β-actin and GAPDH were expressed in all tissues that we have tested (Fig. 1*A*).

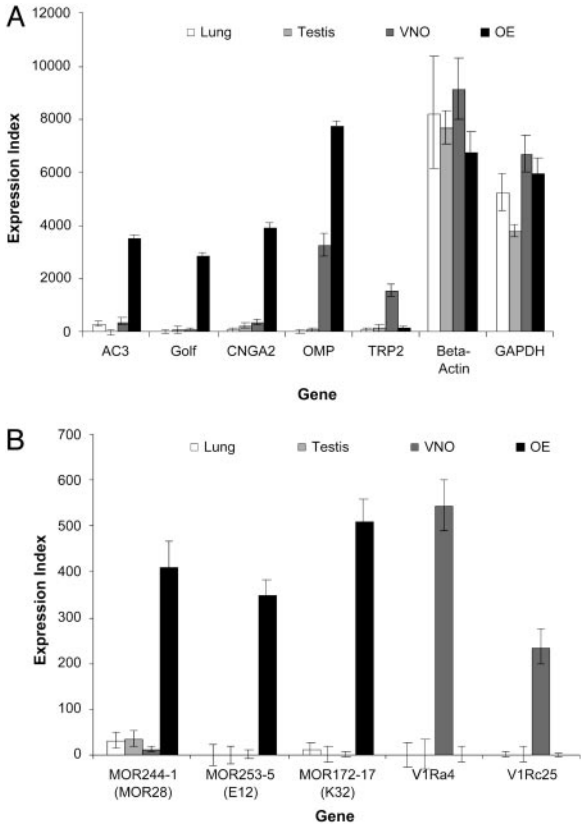

Fig. 1. Validation of the custom MOR chip by examining differential gene expression for control (*A*) and receptor (*B*) genes across tissues. The expression values and standard errors are shown with sample numbers $n = 6$ for OE and VNO, $n = 2$ for lung and testis. Control genes show expression profiles consistent with known data (*A*). For β-actin and GAPDH, the expression values were obtained when these two genes are not used during normalization, therefore avoiding artificially forcing the expression values to be similar. Receptor genes show clear differential expression in the appropriate olfactory tissue despite the relatively low signal levels (*B*).

Despite expected low levels of OR mRNAs, the array was sufficiently sensitive for us to observe a reliable signal above noise in the appropriate tissue for numerous OR and V1R genes (Fig. 1*B*; note the difference in the scales of the *y* axis in Fig. 1 *A* and *B*). Overall, these results suggested that our array design, experimental approach, and data analysis were able to reveal real biologically significant differences in gene expression, even for genes with low expression levels.

Differential Expression of Receptor Genes in OE and VNO. Because the majority of the OR and V1R gene sequences were obtained from the genome, it remains possible that some of them may be OR- or V1R-like genes that actually exert nonolfactory functions in other cells. Evidence for their specific expression in the appropriate sensory organ would confirm these genes as true chemosensory receptors. The expression levels of all of the genes on the MOR array were first compared in OE and VNO samples. Most OR genes (≈70%) have high expression levels in OE samples relative to VNO and other tissues (Fig. 2*A*). Interestingly, the VNO samples seem to have the lowest expression levels of all. One explanation for this finding is that similar mechanisms might control the restricted expression of receptor genes in neurons in OE and VNO; therefore, the expression of V1R genes suppresses the expression of OR genes in the VNO. Similarly, a moderate number (≈40) of V1R genes

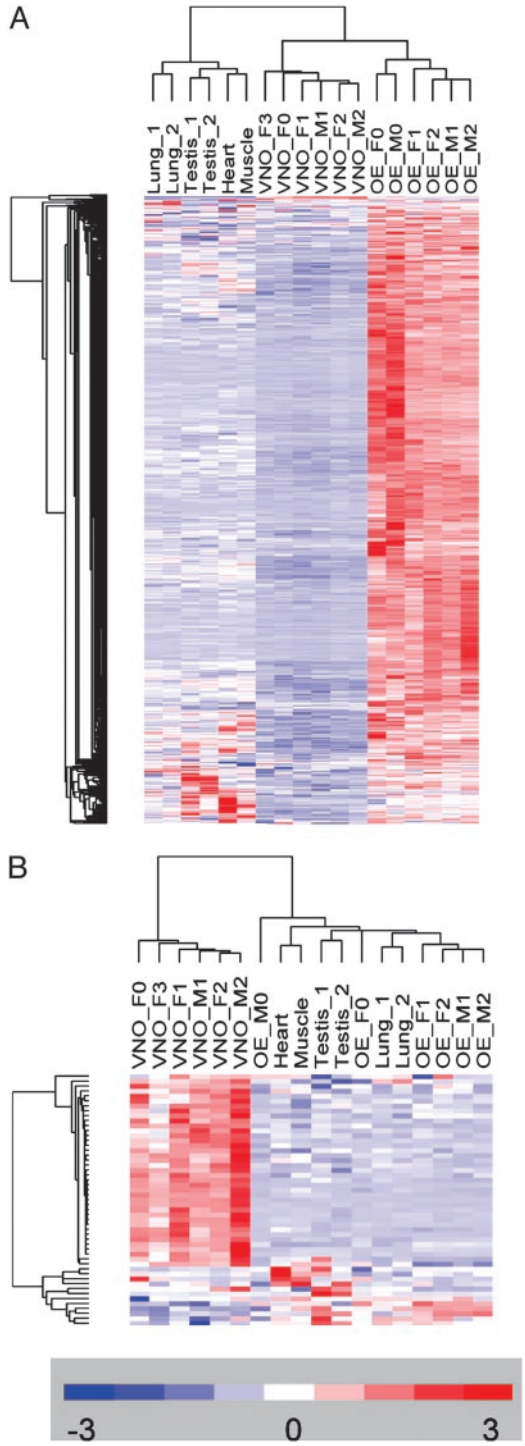

Fig. 2. Expression profiles of OR genes and V1R genes across tissues, showing OE-specific expression of OR genes (*A*) and VNO-specific expression of V1R genes (*B*). All tissues are from 2-month-old adult mice. For OE and VNO tissues, the suffix ''M'' or ''F'' denotes that the tissue was from male or female animals, respectively. The gene expression values are standardized such that the mean is 0 and standard deviation is 1 for each gene. The color represents expression values as shown in the scale bar, with red corresponding to higher-than-mean expression values and blue corresponding to lower-than-mean values. The dendrogram on the left shows clustering of genes, and the top dendrogram shows clustering of samples based on the expression data. Genes with at least two present calls and high variation across samples are chosen for the clustering analysis and are shown in the figure. In *A*, 728 probe sets representing 637 OR genes are shown; in *B*, 52 probe sets representing 51 V1R genes are shown.

show high expression in VNO samples relative to other samples (Fig. 2*B*). The percentage of V1R genes that show VNO-specific expression is lower than OE-specific OR genes, largely because of the fact the no UTR probe sets were designed for V1R genes, which may lead to false negative results for mRNAs with long 3′ UTRs.

To assess the number of OR and V1R genes that show olfactory tissue-specific expression, we used a set of stringent statistical criteria to judge whether a gene was clearly enriched in one type of sample, and another set of looser criteria to judge a gene as marginally enriched (see *Methods* for details on the criteria). A summary of such comparisons between OE and VNO samples is shown in Table 1, and searchable database of enriched genes is available at http://firestein.bio.columbia.edu. In addition, to check the robustness of the statistical criteria, samples were randomly shuffled (permuted) 50 times and the false discovery rates (FDRs) were estimated (24). These results are also shown in Table 1.

As expected, the results showed that OR genes were enriched in OE, and V1R genes were enriched in VNO. A total of 817 OR genes were classified as enriched in OE with a low FDR (overall FDR 1.7%, including both clearly enriched genes and marginally enriched genes). To our knowledge, this is the first time that expression in OE has been confirmed for such a large number of OR genes. Thirty-one V1R genes were labeled as clearly enriched in VNO with a very low FDR (0%). The classification of marginally enriched in VNO is less reliable, with a high FDR (75%), and additional evidence would be required to identify those 13 V1R genes as VNO-specific. The higher FDR for VNO gene detection is mostly due to the fact that the MOR array is biased toward OE-specific genes (many more OR genes than V1R genes), and is also partially due to the fact that V1R detection is not optimized (no 3′ UTR probe sets).

To evaluate detection sensitivity, we collected 76 OR genes from the literature known to be expressed in OE and found that 56 of these were selected as enriched in OE from our array results. This finding gives a sensitivity value of 74%. More recently, >400 OR ESTs were reported (13), and 378 of these OR genes are also represented on the MOR array, with 310 of them listed as OE-specific from our array data. The estimated sensitivity from this larger data set is 82%. The array specificity was more difficult to estimate because it is not known how many of the 817 OR genes classified as enriched in OE were false positives. The overall FDR for OR genes (1.7%) can be used as an approximation of false positives, therefore setting the specificity at 98%. The real specificity is likely to be lower than this because the permutation method used in FDR computation may miss certain kinds of false positives; for example, the effect of cross-hybridization between OR genes cannot be estimated from permutation of OE and VNO samples.

Because olfaction plays an important role in social and sexual behaviors, we compared OE and VNO tissues from male and female adult animals, and found no apparent difference in either OR or V1R gene expression (Fig. 2), which suggests that differential receptor expression may not be a mechanism for different olfaction-related behaviors between sexes. However, it remains possible that there might be sexual dimorphism in genes that we cannot reliably detect with the MOR chip, particularly for a large number of V1R genes.

OR Gene Expression in Non-OE Tissues. OR gene expression has been reported in various nonolfactory tissues, principally in the testis (3, 5, 25, 26), but also in brain, heart, taste, and other tissues (27, 28). These reports were generally based on isolated genes or gene families, and only a few cases included comparisons of expression level with OE tissue. The MOR array is ideally suited to examine the extent and level of OR gene expression across multiple tissues.

Total RNAs were extracted from mouse testis, liver, heart, cerebellum, and muscle samples and processed for hybridization with MOR arrays. To minimize false positives, we used the set of stringent criteria to identify genes that were enriched in each of the

samples. The VNO was used as the background because it has the lowest level of OR gene expression. Compared to the VNO, we find 30–100 OR genes enriched in each of the various tissues tested (Table 2 and at http://firestein.bio.columbia.edu). To determine whether any of these ORs are expressed at a higher level than that observed in the OE, the background was set to OE levels and enriched genes were identified. Very few genes were classified as enriched in non-OE tissue in comparison with the OE (Table 2 and database at http://firestein.bio.columbia.edu). There remains the possibility that some level of crossreactivity with non-OR genes is responsible for false positives in other tissues. We consider that possibility unlikely, because only probe sets with sufficient specificity were included in the array. Furthermore, any crossreactivity for the probe set of any one OR gene is most likely to arise from other OR genes. These data suggest that, although there might be a small number of OR genes expressed in other tissues, very few are expressed exclusively in non-OE tissues. However, it should be noted that we cannot rule out the possibility that some OR genes are indeed highly expressed in a very small number of cells in the non-OE samples. Detailed study at a better resolution is required to address such questions.

Temporal Pattern of OR Gene Expression. The developmental course of OR gene expression remains largely undocumented (29, 30). Mouse OE tissues were collected from animals at various ages and the levels of OR expression were compared (Fig. 3 and numeric values and database at http://firestein.bio.columbia.edu). At embryonic day 13 (E13), OR gene expression in OE did not appear different from non-OE samples. There has been observation of OR gene expression reported as early as E12 (30); however, the number of OR-positive cells at that time is very low, presumably making it difficult to detect these by microarray. Between E15 and E16, a time when OR gene expression is thought to begin in relatively large numbers of cells (30), there is a jump from background to a low level of OR expression. After this initial period, OR gene expression remains at a relatively low level until birth, after which it increases steadily. A large number of OR genes appear to be detected only after birth. In <2 weeks postnatal time, the number of detected OR genes reached a peak and remained high until the animals were ≈2 months old (Fig. 3). After 2 months, we see a slow trend toward a declining diversity of OR expression. It should be noted that the detection threshold for each OR gene might be different, depending on the probe design, level of expression level in the cell, and how many cells are expressing it.

Even during the period from 2 weeks to 2 months of age, when the total number of OR genes expressed is nearly constant, each time point is characterized by distinct gene expression patterns. These patterns did not appear to be caused by random variation between animals, because duplicates of the same time point usually shared similar gene expression profiles (data not shown).

OR Genes Expressed in Different Zones also Segregate in the Genome. One of the earliest findings regarding OR genes was their expression in one of four parallel zones that run rostral to caudal across the turbinates within the OE. However, this spatial expression pattern has been determined for fewer than 40 OR genes by *in situ* investigation (30–35). This approach is far too labor intensive (and costly) to perform for each of the >1,000 genes, but the MOR array provides an effective tool for investigating this zonal pattern globally.

To obtain OE tissues with defined zones from wild-type animals, we first microdissected olfactory epithelium in the dorsal recess above the turbinates, a region that includes only zone 1; the ventral portions of the turbinates were then dissected such that only zone 4 and parts of zones 3 and 2 were included (the dissection is shown schematically in Fig. 4). This dissection method left out the dorsal portion of the turbinate to ensure a clean separation of zone 1 and the other zones. Five replicates, each from the dissection of 10 or more adult mice, were made for dorsal (zone 1) and ventral (zones 2–4) tissues and tested on the MOR array.

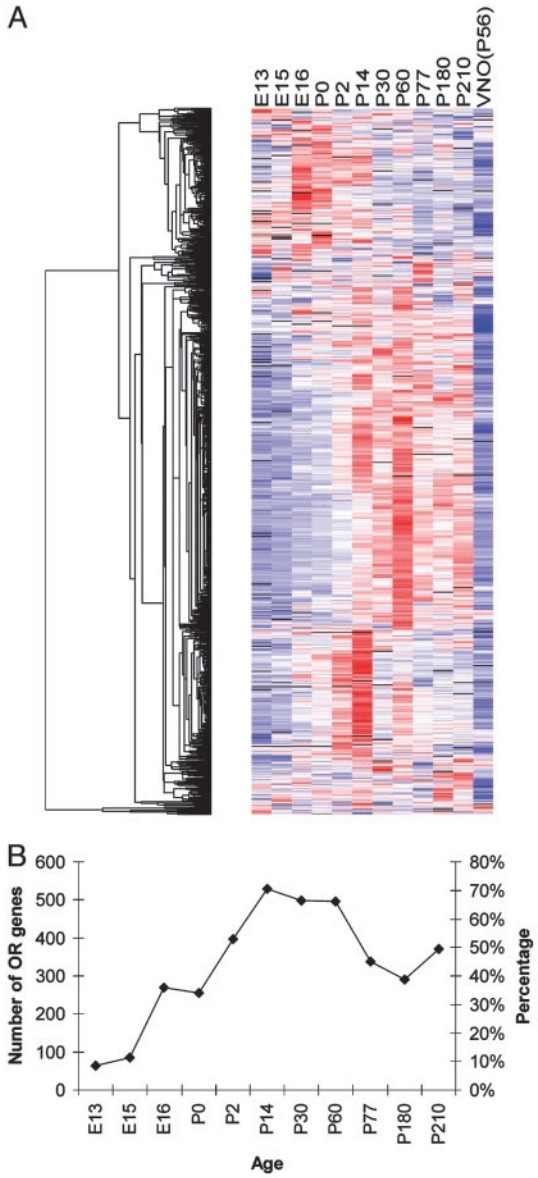

Fig. 3. Temporal expression of OR genes. (*A*) Expression profiles of OR genes from different ages are plotted in the same manner as in Fig. 2. Samples are listed according to the age, and genes are shown according to the clustering result. One sample per time point is used in this analysis. All OR genes that show enrichment compared to VNO in at least one time point are chosen in the clustering. In total, 876 probe sets representing 760 OR genes are shown. (*B*) The numbers and percentages of expressed OR genes (using VNO as background) during different stages. Percentages are calculated by using the same 760 genes shown in *A* as total number of OR genes.

Examination of the 10 zonal samples by hierarchical clustering showed that OR genes can be roughly separated into three categories: those enriched in dorsal (zone 1) samples, those enriched in ventral (zones 2–4) samples, and those without apparent enrichment in either (Fig. 5). This observation prompted us to use an unsupervised clustering method, K-means clustering, to separate OR genes into groups with a high degree of similarity within each group and a low degree of similarity between groups (36). A total of 680 OR genes with reliable detection and differential expression across samples were separated into three groups by K-means

GENETICS

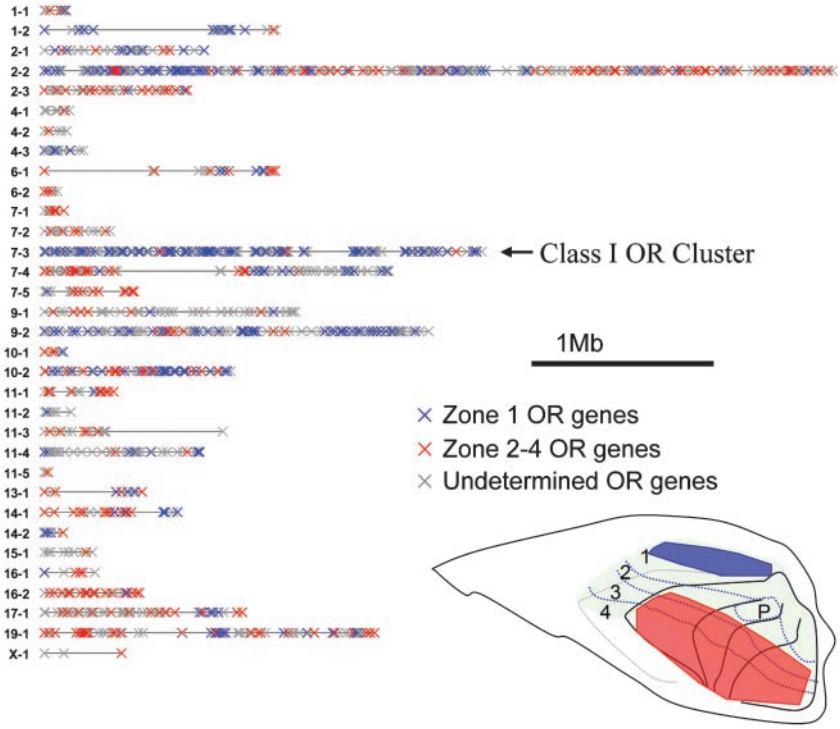

Fig. 4. OR genes from different zones are segregated in the genome. All OR genes located in clusters represented by three or more OR genes on the MOR chip are plotted according to their chromosomal locations. Each cluster is plotted separately and labeled in the format of [chromosome]-[index]. OR genes identified as zone 1 (dorsal) genes are shown in blue, those identified as zone 2–4 (ventral) genes are shown in red; genes with undetermined zone expression are shown in gray. (*Inset*) Diagram of mouse olfactory sensory epithelium, in which the zone 1 and zone 2–4 regions dissected out during the experiments are labeled.

clustering. The expression profiles of the three groups confirmed that group 1 (308 OR genes) is dorsal (zone 1)-specific, group 2 (312 OR genes) is ventral (zones 2–4)-specific, and group 3 (60 genes) contains genes without clear differences. Therefore, our array experiment and data analysis determined the dorsal/ventral expression for a total of 620 OR genes (see http://firestein.bio.columbia.edu).

To validate the zonal results from the array experiments, we collected 31 OR genes represented on the array whose zonal distribution had been previously studied (30–35). We were able to classify zones for 17 of these, and all but one were consistent with previous results. With our dissection method, zone 1 genes are well covered (10 of 13), as are zone 4 genes (five of six); but OR genes in the middle zones (zones 2 and 3 and "Patch") receive relatively low coverage (1 of 12 and one misclassification).

To further confirm the MOR chip data independently, we examined the zonal distribution of 17 OR genes by using *in situ* hybridization (Fig. 5 and Table 3). We were able to determine the zones (dorsal/zone 1 vs. ventral/zones 2–4) for 14 of these genes based on microarray data, and the *in situ* data confirmed all of them. For the remaining three genes whose zones cannot be determined from the array data, *in situ* hybridization showed that one is expressed in zone 1 and the other two are expressed in zone 4 (Table 3). Overall, although there are some genes whose zonal expression cannot be determined by array data, presumably reflecting false negatives, we did not observe any false positives from the 17 genes tested. These validation experiments suggest that the microarray data and subsequent analysis generated reliable results.

Notably, OR genes that are segregated into different zones often appear segregated on the chromosomes. If the OR genes are coded by color for their zonal expression (zone 1 in blue and zones 2–4 in red) and then plotted according to genomic location, zone 1-specific genes form blue patches on the chromosomes that are separate from the red patches formed by genes specific to zones 2–4 (Fig. 4). In cluster 7-3, which contains only the phylogenetically separate cluster of class I ORs (37), 82 of the 85 classified genes were zone 1-specific. The three outliers may be within the normal error rate of array experiments. This result strongly suggests that all class I OR genes are expressed solely in the most dorsal regions (zone 1) of the OE, consistent with *in situ* data from a small number of mouse and rat class I OR genes that all showed specific expression in zone 1 (38). Patches of OR genes from the same zone are observed for other OR gene clusters, albeit to a lesser extent than cluster 7-3. Some clusters are dominated by either zone 1 (e.g., cluster 9-2) or zones 2–4 (cluster 7-5, cluster 16-2, cluster 17-1) OR genes, others can be separated into zone-specific subclusters (e.g., cluster 2-2, cluster 7-4). It should be noted that the clusters we observed are unlikely to be artifacts of cross-hybridization among similar genes, which are known to locate near each other (6, 8, 9, 11, 12, 37). One line of evidence arguing against that possibility is our use of probes targeting 3′ UTRs, which are generally dissimilar between ORs with similar coding regions.

Interestingly, zone 1 appears to contain more than one-third of all of the OR genes that are expressed in OE. Of the 620 OR genes for which zonal expression was determined, 308 (50%) of these are classified as zone 1 genes. Because these genes cover 77% (10 of 13) of all known mouse zone 1 genes, the total zone 1 genes can be estimated as 400 (317, 77%), more than one-third of all ≈1,100 MOR genes. This was somewhat surprising, as it has been commonly thought that each zone expressed about one-quarter of the OR repertoire. Together with the zone 1-specific expression of class I OR genes, these data suggest that the separation between zone 1 (dorsal) and other zones (ventral) might be especially significant.

Discussion

To meet the daunting numerical challenges of understanding the very large OR gene family, we have made use of a custom-designed high-density oligonucleotide array for MOR genes. Our main

findings were that the custom array reliably detects a large number of OR transcripts, that most OR genes are preferentially expressed in the olfactory epithelium, that OR genes undergo developmental regulation, that spatial expression patterns in the OE are reflected in chromosomal organization, and that OR genes distribute unequally between zones.

Designing Arrays from Genomic Data. One important issue that often receives insufficient attention is the 3′ UTR of the gene. Most gene-finding projects focus on the coding regions. However, when designing an array, the 3′ UTRs become important because probe sets in the coding region are ineffective if the 3′ UTR is too long. For the MOR genes, we used computationally predicted 3′ UTR sequences to design UTR probe sets for genes with long 3′ UTRs (>1,000 bp), significantly increasing the detection of OR genes. For 434 OR genes with long 3′ UTRs, coding region probe sets alone resulted in only 47% present calls. However, when UTR probe sets were included, 74% of these OR genes obtained present calls. Because of limited space on the array, no UTR probe sets were designed for genes with short 3′ UTRs, and the P call ratio for these genes was actually lower, only 58% (431 of 739). Similarly, no UTR probe sets were designed for V1Rs in the MOR array, and only 49% received present calls in VNO tissue. One caveat is that the current computational prediction method for 3′ UTR is not perfect (20), and may lead to false negatives. Better prediction algorithms and using probe sets covering multiple predicted 3′ poly(A) sites for each gene may increase microarray detection sensitivity.

Genes with Negative Calls. There were a number of OR genes with absent calls in the OE sample. Similarly, when comparing OE with VNO, or zone 1 with zones 2–4, there were also a number of genes that we could not classify as enriched in either of the two samples. Were these genes with negative calls (not present or no change) real negatives, and what biological information can we obtain from these genes?

We believe that a large number of the absent calls for OR genes are false negatives. No UTR probe sets were designed for 739 of the OR genes, and we might have expected to obtain a P call rate similar to or higher than that obtained for those OR genes currently with UTR probe sets on the MOR array. That would constitute at least a 16% increase, or 118 additional OR genes that would receive present calls. Additionally, because of the low expression of OR genes in the OE samples, many genes that were actually expressed may have been labeled as absent because of weak signals on the array.

Insights from the Zonal Expression of OR Genes. Here we examined the expression of OR genes in zone 1 (dorsal) versus other zones (ventral). Although *in situ* data indicate at least four zones running laterally across the OE, our results indicate that a more fundamental division might be between the most dorsal zone 1 and the rest of the OE. There are several lines of evidence that support this distinction. First, there is a different molecular environment in zone 1 than the other zones. Olfactory cell adhesion molecule is expressed in all olfactory sensory neurons except those in zone 1 (39, 40), as are a few transcription factors (41), and certain carbohydrate epitopes are present only in zone 1 (42). Second, based on our results, zone 1 expresses ≈40% of all OR genes, making it the largest zone. Third, the array results suggest that all class I OR genes are expressed in zone 1, suggesting a functional difference between zone 1 and the other zones. Based on limited functional data, we had proposed that class I ORs might recognize different types of ligands from class II ORs (8). It remains to be tested whether those class II ORs in zone 1 recognize chemically different ligands from class II ORs in other zones.

Class I ORs are related to the family of ORs in fish, where there is no clear zonal expression pattern; i.e., fish have only "zone 1" (43). In frog, class I and II ORs are expressed in separate compartments in the nasal cavity. The class I OR region is accessible in the water, whereas class II OR regions are accessible in the air (44). In terrestrial species, there are no separate compartments for OR expression, but the separation between zone 1, containing all of the class I ORs, and other zones is reminiscent of the two compartments in amphibians.

Previously, no developmental or functional attributes have been associated with the zonal expression of OR genes. Our observation that genes expressed in different zones also segregate in the genome suggests a possible relation to the regulation of OR gene expression. Our current results provide a foundation for future investigations involving promoter analysis to distinguish the two possibilities and gain insight into the mechanism of zonal expression. Overall, the OR custom microarray provides a powerful tool for assessing global regulation and function of this unusually large gene family.

This research was supported by National Institute on Deafness and Other Communication Disorders, Human Brain Project (G.M.S.), Multidisciplinary University Research Initiative (G.M.S.), and Whitehall Foundation (M.M.).

1. Buck, L. B. (2000) *Cell* **100,** 611–618.
2. Dulac, C. & Torello, A. T. (2003) *Nat. Rev. Neurosci.* **4,** 551–562.
3. Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J., *et al.* (1992) *Nature* **355,** 453–455.
4. Walensky, L. D., Ruat, M., Bakin, R. E., Blackshaw, S., Ronnett, G. V. & Snyder, S. H. (1998) *J. Biol. Chem.* **273,** 9378–9387.
5. Vanderhaeghen, P., Schurmans, S., Vassart, G. & Parmentier, M. (1997) *Genomics* **39,** 239–246.
6. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. (2001) *Genome Res.* **11,** 685–702.
7. Zozulya, S., Echeverri, F. & Nguyen, T. (2001) *Genome Biol.* **6,** research0018.1–0018.12.
8. Zhang, X. & Firestein, S. (2002) *Nat. Neurosci.* **5,** 124–133.
9. Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L. & Trask, B. J. (2002) *Hum. Mol. Genet.* **11,** 535–546.
10. Rodriguez, I., Del Punta, K., Rothman, A., Ishii, T. & Mombaerts, P. (2002) *Nat. Neurosci.* **5,** 134–140.
11. Malnic, B., Godfrey, P. A. & Buck, L. B. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2584–2589.
12. Godfrey, P. A., Malnic, B. & Buck, L. B. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2156–2161.
13. Young, J., Shykind, B., Lane, R., Tonnes-Priddy, L., Ross, J., Walker, M., Williams, E. & Trask, B. (2003) *Genome Biol.* **4,** R71.
14. Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. & Brown, E. L. (1996) *Nat. Biotechnol.* **14,** 1675–1680.
15. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. (1999) *Nat. Genet.* **21,** 20–24.
16. Schadt, E. E., Li, C., Ellis, B. & Wong, W. H. (2001) *J. Cell Biochem. Suppl.* **37,** 120–125.
17. Ermolaeva, O., Rastogi, M., Pruitt, K. D., Schuler, G. D., Bittner, M. L., Chen, Y., Simon, R., Meltzer, P., Trent, J. M. & Boguski, M. S. (1998) *Nat. Genet.* **20,** 19–23.
18. Li, C. & Wong, W. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 31–36.
19. Hirota, S., Ito, A., Morii, E., Wanaka, A., Tohyama, M., Kitamura, Y. & Nomura, S. (1992) *Brain Res. Mol. Brain Res.* **15,** 47–54.
20. Tabaska, J. E. & Zhang, M. Q. (1999) *Gene* **231,** 77–86.
21. Firestein, S. (2001) *Nature* **413,** 211–218.
22. Liman, E. R., Corey, D. P. & Dulac, C. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5791–5796.
23. Rogers, K. E., Dasgupta, P., Gubler, U., Grillo, M., Khew-Goodall, Y. S. & Margolis, F. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1704–1708.
24. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., *et al.* (1999) *Science* **286,** 531–537.
25. Asai, H., Kasai, H., Matsuda, Y., Yamazaki, N., Nagawa, F., Sakano, H. & Tsuboi, A. (1996) *Biochem. Biophys. Res. Commun.* **221,** 240–247.
26. Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J. A., Wetzel, C. H., Zimmer, R. K. & Hatt, H. (2003) *Science* **299,** 2054–2058.
27. Mombaerts, P. (1999) *Annu. Rev. Neurosci.* **22,** 487–509.
28. Young, J. M. & Trask, B. J. (2002) *Hum. Mol. Genet.* **11,** 1153–1160.
29. Margalit, T. & Lancet, D. (1993) *Brain Res. Dev. Brain Res.* **73,** 7–16.
30. Sullivan, S. L., Bohm, S., Ressler, K. J., Horowitz, L. F. & Buck, L. B. (1995) *Neuron* **15,** 779–789.
31. Ressler, K. J., Sullivan, S. L. & Buck, L. B. (1993) *Cell* **73,** 597–609.
32. Strotmann, J., Beck, A., Kubick, S. & Breer, H. (1995) *J. Comp. Physiol. A* **177,** 659–666.
33. Sullivan, S. L., Adamson, M. C., Ressler, K. J., Kozak, C. A. & Buck, L. B. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 884–888.
34. Qasba, P. & Reed, R. R. (1998) *J. Neurosci.* **18,** 227–236.
35. Pyrski, M., Xu, Z., Walters, E., Gilbert, D. J., Jenkins, N. A., Copeland, N. G. & Margolis, F. L. (2001) *J. Neurosci.* **21,** 4637–4648.
36. Hartigan, J. A. (1975) *Clustering Algorithms* (Wiley, New York).
37. Zhang, X., Rodriguez, I., Mombaerts, P. & Firestein, S. (2004) *Genomics* **83,** 802–811.
38. Mombaerts, P. (2001) *Annu. Rev. Genomics Hum. Genet.* **2,** 493–510.
39. Schwob, J. E. & Gottlieb, D. I. (1986) *J. Neurosci.* **6,** 3393–3404.
40. Alenius, M. & Bohm, S. (1997) *J. Biol. Chem.* **272,** 26083–26086.
41. Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L. & Dulac, C. (2003) *Neuron* **38,** 161–175.
42. Schwarting, G. A. & Crandall, J. E. (1991) *Brain Res.* **547,** 239–248.
43. Ngai, J., Chess, A., Dowling, M. M., Necles, N., Macagno, E. R. & Axel, R. (1993) *Cell* **72,** 667–680.
44. Freitag, J., Krieger, J., Strotmann, J. & Breer, H. (1995) *Neuron* **15,** 1383–1392.