Chem. Senses 37: 639–653, 2012

doi:10.1093/chemse/bjs045
Advance Access publication March 29, 2012

# Cluster Analysis of Rat Olfactory Bulb Responses to Diverse Odorants

**Matteo Falasconi[1,*], Agustin Gutierrez-Galvez[2,3,*], Michael Leon[4], Brett A. Johnson[4] and Santiago Marco[2,3]**

[1]Department of Chemistry and Physics, University of Brescia and CNR-IDASC, Via Valotti 9, 25133 Brescia, Italy, [2]Department of Electronics, Universitat de Barcelona, Martí i Franquès 1, Planta 2, 08028 Barcelona, Spain, [3]Artificial Olfaction Group, Institute for Bioengineering of Catalonia (IBEC), Baldiri i Reixach 13, 08028-Barcelona, Spain and [4]Department of Neurobiology and Behavior, University of California Irvine, 2205 McGaugh Hall, Irvine, CA 92697-4550, USA

*Correspondence to be sent to: Agustin Gutierrez-Galvez. Department of Electronics, Universitat de Barcelona, Martí i Franquès 1, Planta 2, 08028 Barcelona, Spain. e-mail: agutierrez@el.ub.es*

*These authors contributed equally to the work.

*Accepted February 29, 2012*

## Abstract

In an effort to deepen our understanding of mammalian olfactory coding, we have used an objective method to analyze a large set of odorant-evoked activity maps collected systematically across the rat olfactory bulb to determine whether such an approach could identify specific glomerular regions that are activated by related odorants. To that end, we combined fuzzy c-means clustering methods with a novel validity approach based on cluster stability to evaluate the significance of the fuzzy partitions on a data set of glomerular layer responses to a large diverse group of odorants. Our results confirm the existence of glomerular response clusters to similar odorants. They further indicate a partial hierarchical chemotopic organization wherein larger glomerular regions can be subdivided into smaller areas that are rather specific in their responses to particular functional groups of odorants. These clusters bear many similarities to, as well as some differences from, response domains previously proposed for the glomerular layer of the bulb. These data also provide additional support for the concept of an identity code in the mammalian olfactory system.

**Key words:** cluster analysis, cluster validity, olfactory bulb, olfactory coding, 2-deoxyglucose mapping

## Introduction

Olfactory transduction occurs when odorant receptors bind to different types of odorant molecular features, such as functional groups or carbon chain length (Shepherd 1987). Each olfactory sensory neuron in mammals expresses a single type of receptor protein (Ressler et al. 1993; Vassar et al. 1993), and the axons of olfactory receptor neurons expressing the same receptor project to a small number of glomeruli on both the lateral and the medial aspect of the olfactory bulb (Ressler et al. 1994; Vassar et al. 1994; Mombaerts et al. 1996). Each glomerulus receives projections from a single type of receptor neuron (Treloar et al. 2002). As a consequence, molecular features captured by the entire complement of olfactory receptors can be seen at the glomerular level either in 3D or in a 2D activity map. Thus, the glomerular layer seems privileged in its ability to reveal information about the coding strategy used by the olfactory system.

Using uptake of [$^{14}$C]-radiolabeled 2-deoxyglucose (2DG), Johnson and Leon have systematically mapped activity across the glomerular layer of the entire rat olfactory bulb in response to more than 300 odorants (Leon and Johnson 2003; Johnson and Leon 2007; Woo et al. 2007). The 2DG technique is capable of resolving individual glomeruli and can, at the same time, describe the entire complement of glomerular responses to any particular odorant. Other techniques have been also used to obtain activity maps of the olfactory bulb (e.g., Salcedo et al. 2005; Mori et al. 2006; Costanzo and Kobayashi 2010), and together, these studies have shown that odor stimuli in the rat produce spatially

distinct patterns of activity in the glomerular layer that are overlapping but that differ for different odorants (Johnson et al. 1999). The response patterns are consistent with the concept of a chemotopic organization of the olfactory system, in which particular odorant molecular features are associated with activity in spatially distinct domains (Johnson and Leon 2000, 2007; Mori et al. 2006; Soh et al. 2011). The borders of these proposed domains have been assigned visually on the basis of experiments using odorants that differed systematically in structure, with later adjustments made as the glomerular response archive increased in size (Johnson and Leon 2007). As such, it is possible that the original notions of domain borders may have had an undue influence on the final models, and it also may be the case that overlaps in activity for dissimilar odorants have not been fully represented in the models. Given the current large size of the 2DG glomerular activity response archive in the rat, it should now be possible to examine the clustering of responses to specific groups of odorants using objective methods. Therefore, the main goal of this work was to determine whether glomerular responses to similar odorants are clustered in the glomerular layer. If so, we hoped to reach additional conclusions about the relationship of odorant chemistry to those clusters.

The analysis of glomerular activity patterns is a challenging problem because odor patterns are complex, highly irregular, noisy, and may contain missing values due to imperfections in the experimental procedures. For these reasons, sophisticated automated data analysis techniques are required to investigate the data structure systematically. In this work, we used cluster analysis methods to discover groups of pixels that have similar response patterns across the entire set of odorants. We focused on glomerular partitions obtained by fuzzy *c*-means (FCM), a method used in pattern recognition to cluster responses by allowing 1 piece of data to belong to 2 or more clusters (Dunn 1973; Bezdek 1981).

We also validated the estimates of the best number of clusters with objective criteria. Most current clustering algorithms do not provide any estimate of the significance of the calculated results; every clustering algorithm tends to produce clusters regardless of whether the data contains true clusters. Therefore, verification of clustering results is a crucial task. To this end, we have developed a validity method (Falasconi et al. 2010) for fuzzy partitions based on an evaluation of partition stability under bootstrap resampling of the data, a procedure that allows us to estimate the best partitioning of the data.

## Material and methods

### Odor exposure

We used the uptake of radiolabeled 2DG as a metabolic marker to register glomerular layer responses to odorants, as this is the only method currently able to give a quantitative measure of relative activity throughout the entire structure at a spatial resolution sufficient to detect activity in a single glomerulus. The trade-off between exposure time and the ability to obtain a complete survey of the glomerular response in a highly heterogeneous sensory system seemed reasonable for the purpose of understanding the coding strategy of the system. At the same time, it should be noted that brief exposures to odorants results in similar response patterns (Mori et al. 2006). Neat odorants and odorants with purities ranged from 90% to greater than 99.5% were dissolved in water or mineral oil and exposed to the rats at 1 L/min. All surfaces in contact with the odorant stream were made of Teflon, brass, or glass to minimize interactions with odorants. Each odorant was used with a dedicated set of tubing to prevent cross-contamination. The exposure chamber was constructed from a 2-L glass jar with holes bored in the lid for odorant entry and exhaust. Research-grade high-purity nitrogen gas was bubbled through the liquid odorant preparation in a gas-washing bottle to volatilize the odorants. The odorized nitrogen vapor then was mixed with ultra-zero grade air prior to entering the exposure chamber. For neat odorants, the flow rate into the chamber was 2 L/min.; for odorants in mineral oil, the flow rate was 1 L/min. The odor delivery system was equilibrated for at least 15 min prior to an exposure. All components except for the exposure chamber were equilibrated at the final flow rates for at least 15 min.

Odorant entry into the exposure chamber began after the rat was introduced so that the odorant concentration would rise steadily as 2DG became available to the olfactory bulb. Wistar rats 18–21 days old (about 30–50 g) were transferred with the dam into a cage containing clean bedding at least 1 h prior to the first exposure to reduce carryover of odors from home cages to the exposure chamber. Each rat was given a subcutaneous injection of [$^{14}$C]-2DG (Sigma; 54 mCi/mmol, 0.1 mCi/mL in 0.9% saline, 0.08 mL for a 50-g rat) at the back of the neck before being placed in the odorized chamber. Exposure then continued for 45 min. After the exposure period, the rat was decapitated. The brain was removed, frozen in 2-methylbutane at −45°C and stored at −80°C. Subsequently, the slides were warmed to −20°C and sectioned at 20 μm in a cryostat. Every sixth section was taken for autoradiography, and adjacent sections were used for staining with cresyl violet to be able to register 2DG uptake specifically within the glomerular layer. Odorants, sources, concentrations, and group-averaged data matrices for specific odorants can be found at http://gara.bio.uci.edu. All procedures involving rats were approved by the UC Irvine animal care committee.

### Data analysis

Mapping of 2DG uptake in the glomerular layer of the autoradiographic images was performed by taking discrete measurements dictated by a set of polar grids, the particular grid being chosen on the basis of the section's anterior–

posterior position relative to landmarks detected in the cresyl violet–stained sections. Values of film density were acquired in units of grayscale and then were transformed into nCi/g tissue by comparison to autoradiographically exposed radioactivity standards. The data from separate sections were collapsed into data matrices, with the anterior–posterior dimensions of the matrices adjusted relative to anatomical landmarks. The matrices from the left and right bulbs of the same rat were averaged. The units then were transformed into ratios of glomerular uptake to uptake occurring in a defined span of the subependymal zone in order to control for different amounts and specific activities of radiolabel reaching the bulb in each rat.

One rat in each litter was exposed to a blank condition and their littermates were each exposed to a different odorant. For a given experiment, the matrices for the blank condition rats were averaged, and the response, this average blank matrix was subtracted from each matrix of the experimental rats. This blank-subtracted matrix then was converted into a matrix of *z*-scores relative to the mean and the standard deviation of values across the original matrix. The *z*-score matrices from rats exposed to the same odorant within an experiment then were averaged and visualized as a color-coded contour chart. We prefer to display a ventral-centered orientation for displaying these charts to minimize the impact of missing tissue from the dorsal edge during sectioning. Alternative orientations, including dorsal-centered charts and rotatable 3D maps of these glomerular responses, are available at http://gara.bio.uci.edu. Prior to further analysis, all matrices for a given odorant were averaged regardless of concentration.

Before correlating molecular properties with uptake in modules, we first eliminated a number of patterns that resembled blank patterns due to excessive odorant dilution, low odorant volatility, or other unknown reasons. We also eliminated certain patterns from earlier studies for which no blank condition was available for subtraction. When our database contained multiple patterns representing different concentrations of the same odorant, we used the pattern evoked by the highest concentration, and when there was more than one pattern available at the highest concentration, we used the average of values for that concentration. Almost all normalized odorant matrices retain similar glomerular response patterns over a wide range of concentrations (Johnson and Leon 2000), and therefore, the effect of averaging over such groups was minimal. The exceptions appear to be those odorants that contain a significant contaminant that can increases across a perceptual threshold with increasing concentration (Johnson et al. 1999; Johnson and Leon 2000). A total of 308 odorant-evoked activity patterns were used in these correlations. The names of the odorants, their CAS numbers, and their dilutions are provided in Supplementary Table 1.

The data set (http://gara.bio.uci.edu/) consists of group-averaged activity maps of [$^{14}$C]-2DG uptake across the entire glomerular layer in response to odorants, some of them at different concentrations, with some replications of the same exposure conditions (Johnson et al. 1999). Figure 1 shows the activity maps obtained for 6 odorants, where we can see the variability of the glomerular activity for these odorants. For the present analysis, each map was converted into a 1D data matrix, and all background pixels (with zero variance) were removed. Figure 1f shows a map with background pixels and Figure 1a–e maps with background pixels removed.

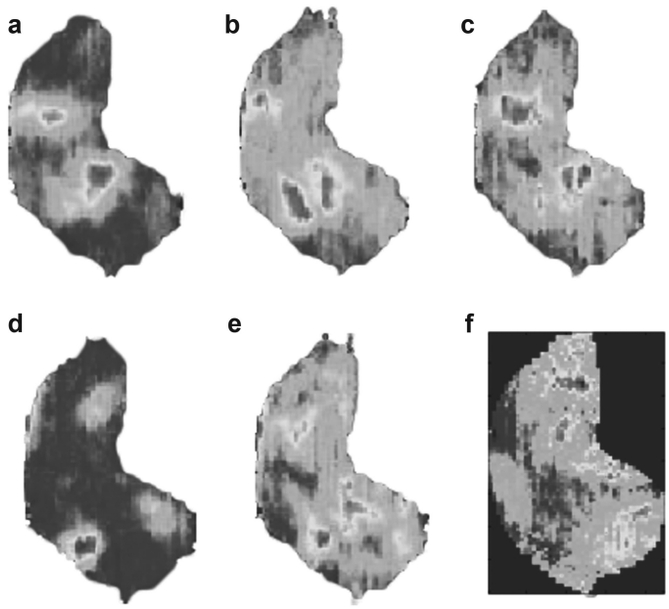


**Figure 1** Olfactory bulb activity obtained using uptake of [$^{14}$C]-radiolabeled 2DG when exposed to 6 different chemical volatiles. (**a**) Decanal. (**b**) Methyl octanoate. (**c**) Nonanal. (**d**) Valeric acid. (**e**) Ethyl eptanoate. (**f**) 1,2,3,4-Tetramethylbenzene with background pixels. This figure appears in color in the online version of *Chemical Senses*.

As a consequence of our experimental technique, most of the olfactory bulb activity maps contain some missing values due to minor tissue damage during bulb sectioning. These missing areas are mainly distributed on the border of the map and in particular in the ventrocaudal and dorsal parts of the olfactory bulb, where perfect cryostat sectioning is challenging. In addition, there is some misalignment of the maps due to the different sizes and shapes of each olfactory bulb as well as to variance in mounting of the tissue prior to sectioning. Before performing clustering, it was necessary to make all maps equal in pixels size. We restricted our analysis to the 1780 pixels that were represented in all the maps. These pixels cover almost the entire OB, except its borders. This procedure yielded 2 final data sets, 1 involving each experimental pattern (D1: $1780 \times 470$ pixels) and 1 involving each unique odorant (D2: $1780 \times 308$ pixels).

The problem addressed in this report is whether there is clustering of image pixels across odorants. By convention, the objects to cluster (the pixels) were placed in rows, whereas the odorant features were placed in the columns of the data matrix. Variance in the odorant response can be seen as a spatial spread of the activity in the matrix.

## Cluster analysis and validity

### *Fuzzy c*-means

The FCM algorithm is based on the minimization of the following objective function:

$$J_m\left(U, V\right) = \sum_{j=1}^{n}\sum_{i=1}^{C} u_{ij}^{m}\left\|\mathbf{x}_j - \mathbf{v}_i\right\|^2, \qquad (1)$$

where $U$ is the fuzzy membership matrix of elements $u_{ij}$, whereas $V$ is the cluster centroids vector of elements $\mathbf{v}_i$. The indices $i$ and $j$ run over the sum operator, where $n$ is the number of patterns and $C$ is the number of clusters. The symbol $\|\ \|$ stands for the Euclidean distance between the $j$th pattern $\mathbf{x}_j$ and the $i$th cluster center $\mathbf{v}_i$. The parameter $m$ ($m > 1$) is called the fuzzifier or weighting exponent, and it determines the degree of fuzziness of the FCM partition. For guaranteeing the proper convergence of the algorithm, $m$ is set close to one for high-dimensional data sets; we set $m = 1.1$ to ensure a proper representation of the structure of the data set.

To understand our decision regarding the value of $m$, one must take into account the counterintuitive consequences of working in a high-dimensional space, such as characterizes our data set. One of these consequences is that for any structure that our data may have, data points will lay at the periphery of this structure. This is caused by the fact that the hyper volume surrounding the boundaries of any object increases more rapidly than the inner hyper volume of the object as the dimensions of the space increase (Friedman 1994). Different values of $m$ allow the final position of the centroids to be closer to or farther from the data points, while always minimizing the objective function. Values of the $m$ parameter close to 2 tend to group centroids around the geometrical center of the data, a situation that produces a poor representation of the peripheral data points. For values of $m$ below 1.45, the centroids move close enough to the periphery of the data set to obtain a proper clustering. However, the higher the value of $m$ is, the closer to the geometrical center the centroids are. This situation translates to noisier and less-repeatable clustering of the results due to a greater distance between the centroid and data points belonging to that cluster. The proper distribution of cluster centroids for different values of $m$ was determined by visual inspection of the 4 principal components. In light of these considerations, the value chosen for m of 1.1 is close to 1, while keeping the fuzzy nature of the method (at $m = 1$, the FCM algorithm is identical to the hard $k$-means algorithm).

### *Stability-based cluster validity*

We used a novel paradigm for validating the identified fuzzy partitions that is based on the concept of partition stability under data perturbation. Data can be perturbed in several ways; here, we repeatedly apply bootstrap sampling on pixels and named the method bootstrap partition stability estimation (fBPSE, where f stands for fuzzy). The method has been described in detail in a recent report (Falasconi et al. 2010). For a given number of clusters $K$, each bootstrap iteration produces a different data set and then by applying FCM, it produces a different data partition. We performed FCM clustering and fBPSE validity testing from $K = 2$ up to $K = 20$. We thought that it was reasonable to stop the clustering at 20 because it is comparable to the number of modules proposed by the existing chemotopic model of the olfactory bulb (Leon and Johnson 2003).

The fBPSE estimates and minimizes, over a series of partitions with increasing number of clusters $K$, the partition variability given by the following relationship:

$$V\left(K\right) = \frac{1}{\frac{B(B-1)}{2}}\sum_{i=1}^{B}\sum_{j>i}^{B-1} d(U_K(Y(i)), U_K(Y(j))), \qquad (2)$$

where $B$ is the number of bootstrap samples ($B = 20$ for data analysis). $V(K)$ is formally equivalent to the empirical bootstrap estimate of the variance, where $d(.,.)$ plays the role of distance. Actually, $d(.,.)$ is a dissimilarity measure between 2 fuzzy partitions (Borgelt 2007), and $U_K(Y(i))$ is the membership matrix achieved by running the FCM algorithm on the $i$th bootstrap sample $Y(i)$. The 2 indices $i$ and $j$ run over the entire set of bootstrap partitions. In our work, we adopted the adjusted fuzzy Rand index (Campello 2007) as a similarity measure.

## Chemical meaning of the clusters

### *Ranking of the odorants inside each cluster*

The odorant scores allowed us to rank the chemicals according to their mean activity inside each cluster; the larger the score, the more important the odorant is for the activation of glomeruli in a particular cluster. As mentioned in Data analysis, our clustering results in groups of pixels that correspond to anatomical areas of the OB. Thus, the score associated to a single odorant within a given area of the OB is simply the average activity value across those pixels. This score is captured by the following expression:

$$\text{odorant\_score}_c = \frac{1}{n_c}\sum_{i=1}^{n_c}\text{Activity}_{i\in C}, \qquad (3)$$

where $C$ is the cluster and $n_c$ is the number of pixels in the selected cluster.

To judge whether the score was significant, we performed a two-sample (one sided) $t$-test, which determines whether the odorant score inside the cluster is significantly higher ($P < 0.01$) than the score for pixels outside the cluster, calculated using Equation (3). For each cluster, the ranking table of the most relevant chemical compounds can be visually inspected to determine whether the highest ranking odorants

shared a specific chemical feature. This information is provided as a table in the Supplementary Material.

### *Expression of molecular features inside each cluster*

Each chemical compound is associated with a set of chemical features. To describe the present odorant set, we considered the using 50 different descriptors involving functional groups, hydrocarbon chain structures, and the presence of benzene rings. However, the final set of 21 descriptors was selected after eliminating chemical features that are correlated, to avoid the introduction of undesired statistical relationships, and including only those odorants in the database with 15 or more examples to maximize statistical consistency. The final list includes the following properties: carboxylic acid, alcohol (not phenol), primary alcohol, aliphatic (not alicyclic) secondary alcohol, ester (not lactone), aliphatic ester (not alicyclic), aromatic ester, aldehyde, aliphatic aldehyde, aromatic aldehyde, ether, ketone, aliphatic ketone (not alicyclic), aliphatic or alicyclic with multiple O-containing functional groups, aliphatic or alicyclic hydrocarbon, alkane, aromatic, aromatic with O-containing substituent, alicyclic, polycyclic, and heterocyclic.

From the computational point of view, each odorant in the database was mapped into a binary vector of molecular features to indicate whether a certain molecular feature was present in its chemical structure. We assigned a score to every chemical feature in each cluster by taking the average score value across all the odorants sharing that feature:

$$\text{molfeature\_score}_c = \frac{1}{M}\sum_{j=1}^{M} \text{odorant\_score}_{C,j}. \qquad (4)$$

A large score value for a chemical feature in the given cluster *c*-means that the related odorants are among the highest ranking compounds in that cluster. To judge significance ($P < 0.01$), we used a two-sample (1 sided) *t*-test comparing feature scores inside the cluster with those calculated outside the cluster using Equation (4).

## Results

### Initial examination of data

The mean and the variance of activity values for each pixel were calculated across the 308 different odorants of data set D2 (very similar results were obtained for data set D1). The mean activity (Figure 2a) was higher in the posterior domain of the bulb. Indeed, we observed that many odorants with diverse molecular features have their predominant activity in this area of the bulb. Activity of glomeruli in the posterior domain is more correlated with water solubility and the presence of multiple oxygen-containing functional groups rather than with the presence of any particular single functional group, as previously reported (Johnson Arguello, et al. 2007). Conversely, the ventral aspect of the bulb had a very low average activity. Very few odorants in the database activated glomeruli in this region, which responds to odorants with long aliphatic hydrocarbon chains, odorants without oxygen-containing functional groups and methyl-substituted bicyclic compounds (Johnson and Leon 2007; Johnson Ong, et al. 2007). These compounds constitute a minority of the odorants in the database, explaining their small contribution to the mean activity value in this area.

The distribution of the average variance of pixels across the odorant set is shown in Figure 2b. Groups of pixels with higher variance (meaning higher information content) were found to be located in the dorsomedial and dorsolateral regions of the olfactory bulb, whereas the ventral aspect of the bulb was characterized by very little variance. Focal areas of lower variance also were observed in the midlateral

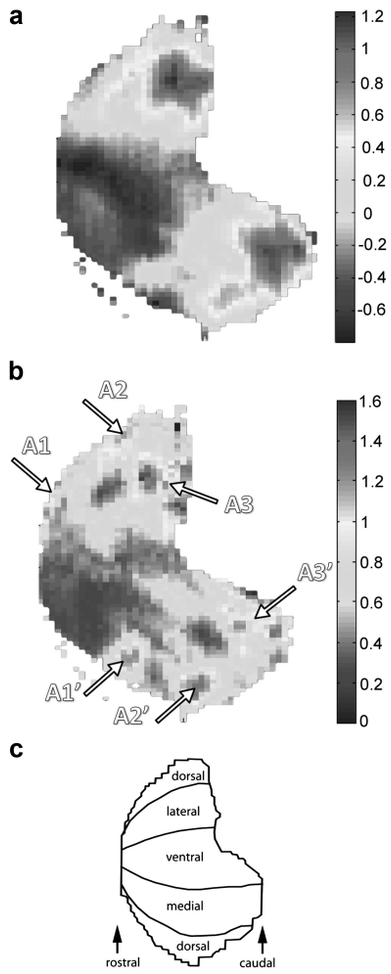


**Figure 2** Initial examination of the activity images show the distribution maps of the mean (**a**) and variance (**b**) activity values across the entire set of odorants in D2. Areas of high activity in the dorsolateral aspect (A1, A2, and A3) are paired with corresponding high activity areas at the dorsomedial aspect (A1′, A2′, and A3′). (**c**) Anatomical correspondence of the activity images. This figure appears in color in the online version of *Chemical Senses*.

and midmedial parts of the bulb. Focal areas of both high and low variance occurred in lateral–medial pairs, a finding that is reminiscent of paired projection of sensory neurons expressing the same receptor into lateral and medial glomeruli and the paired responses that have been reported (Ressler et al. 1994; Vassar et al. 1994; Mombaerts et al. 1996; Johnson and Leon 2007). The existence of regions of high variance, as well as the generally patchy distribution apparent in Figure 2b, strongly supports the hypothesis that there is a finer clustering structure inherent in the data that might be recovered by using fuzzy clustering partitions with high $K$ values.

## Data clustering and validity

### *FCM clustering*

Principal component analysis confirmed that the difference in activity between the posterior aspect and the rest of the bulb was the main source of variance in the data set because the distribution map of PC1 was very similar to the distribution of the mean activity (data not shown). Likewise, when $K = 2$, the FCM method revealed 2 clusters separated along the first PC (Figure 3a); one cluster coincided with the posterior aspects of the OB and the other cluster that involved the remainder of the bulb (Figure 4a). When the number $K$ clusters increase, the multidimensional data are subsequently split along higher PC dimension, as shown in Figure 3b.

This result is expected because the FCM objective function Equation (1) with Euclidean distance is just a generalization of variance minimization method (hard $c$-means or $k$-means) and hence tends to recover hyperspherical fuzzy clusters in the feature (odorants) space. Figure 4a–d illustrates the partitions achieved by FCM at different $K$ values. All of the partitions produced clusters of pixels paired in the medial and lateral aspects of the olfactory bulb as indicated by the color coding in Figure 4. In order to distinguish the 2 aspects of the same module, apices were used to indicate the lateral module, 2a′ is paired with 2a, etc.

This finding is consistent with the existence of lateral–medial paired regions of high variance (Figure 2b) and with the fact that glomeruli are coactivated by individual odorants in both the medial and the lateral aspects of the olfactory bulb (Johnson et al. 1998; Johnson and Leon 2007). This result is not obvious because pixels were clustered in the odorant space, disregarding any topological information related to the OB map. Therefore, pixels that are related to a different aspect the OB—hence relatively far apart on the OB map—but that are ''close'' according to their chemical response were correctly assigned by the FCM to lateral and medial domains of the same cluster.

It is worthwhile to note that by implementing different clustering algorithms, using different distances in Equation (1) or using other clustering methods (i.e., hierarchical approaches), it was not possible to reproduce this result (data not shown). Hence, this type of subdivision seems to be intrinsically related to the variance minimization criterion of FCM clustering algorithm.

In the next sections, we will discuss the intrinsic validity of the achieved FCM partitions (across different $K$ values) as well as the possible hierarchical relationship between clusters within selected partitions.

### *Cluster validity*

As described above, we used a bootstrap method (fBPSE) to estimate cluster validity through objective mathematical criteria applied at every partition of the data within a certain range of $K$ values. We found that the best partition (the most stable, according to fBPSE) was at $K = 3$, which corresponds to a global minimum in each of the variability curves for both data sets D1 (Figure 5a) and D2 (Figure 5b). The partition $K = 2$ yielded a rather similar variability value; the difference between these 2 partitions lies in the splitting of the pixels belonging to the posterior clusters leading to 2 smaller clusters 2b–b′ and 3b–b′ (compare Figure 4a with Figure 4b) while the remaining (ventral) area is similar for both $K = 2$ and $K = 3$ partitions.

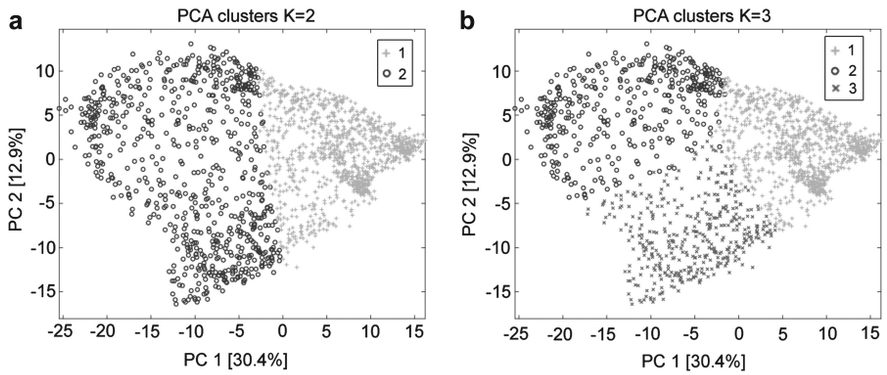


**Figure 3** Principal component analysis (PCA) score plots of partitions obtained with FCM. Feature space is formed by all odorants in the database, and each point in this space corresponds to a pixel of the olfactory bulb activity images. (**a**) $K = 2$. (**b**) $K = 3$. This figure appears in color in the online version of *Chemical Senses*.

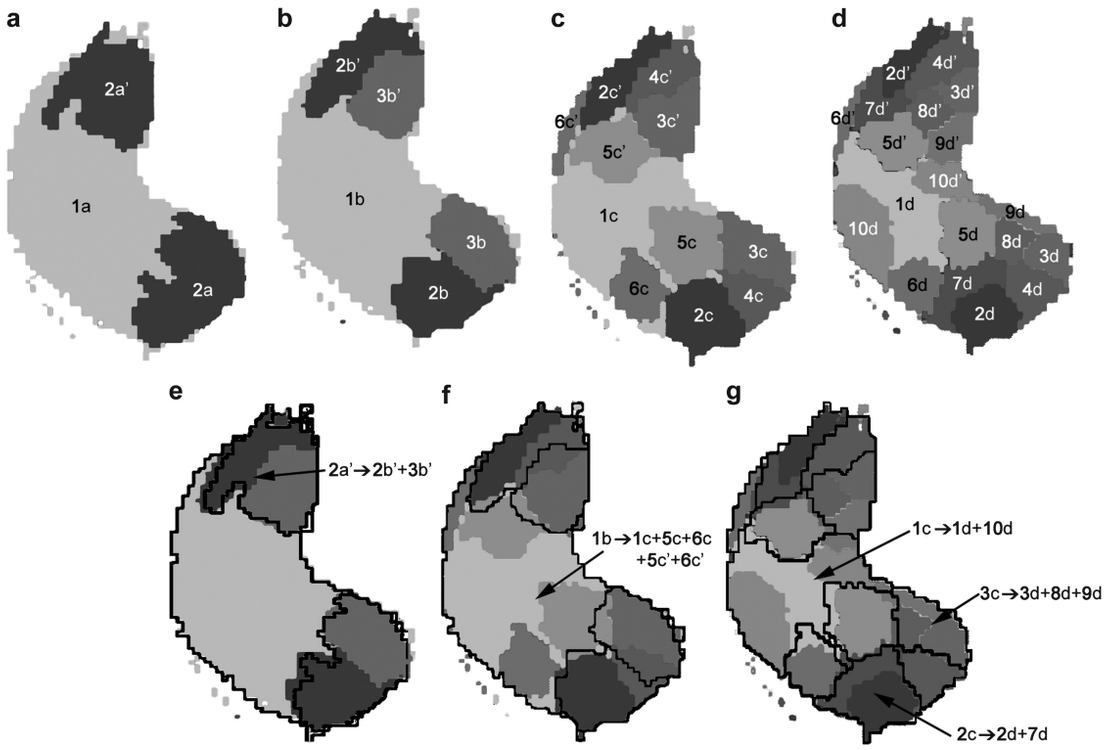


**Figure 4** Clustering and hierarchical structure of the olfactory bulb images. (**a**) Clusters obtained with FCM for $K = 2$, (**b**) $K = 3$, (**c**) $K = 6$, and (**d**) $K = 10$. We compare the clusters obtained at certain value of $K$ with those obtained at the previous $K$ value. (**e**) Clusters for $K = 3$ with boundaries of $K = 2$ clusters. Cluster 1a becomes 1b and cluster 2a is split up into 2b and 3b. (**f**) Clusters for $K = 6$ with boundaries of $K = 3$ clusters. Cluster 1b becomes 1c + 5c + 6c + 5c′ + 6c′. Clusters 2b + 3b become 2c + 3c + 4c. This is the only case of a nonhierarchical subdivision of clusters in this study. (e) Clusters for $K = 10$ with boundaries of $K = 6$ clusters. Three clusters keep most of their area: 4c, 5c, and 6c that becomes 4d, 5d, and 6d, respectively. Cluster 1c becomes 1d + 10d + 10d′, cluster 3c becomes 3d + 8d + 9d, and cluster 2c becomes 2d + 7d. These results show the existence of an underlying hierarchical structure in the data. This figure appears in color in the online version of *Chemical Senses*.

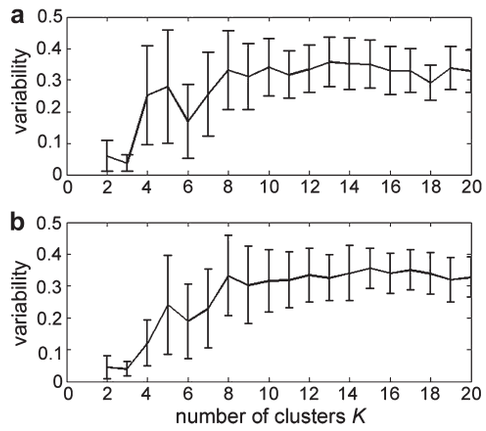


**Figure 5** Cluster validity results by fBPSE on the data sets D1 (**a**) and D2 (**b**). The error bars report the standard deviation of $V(K)$ calculated value over the 190 pairs of compared partitions.

Therefore, the most reliable overall clustering seems associated with the main difference in the average activity between the posterior aspect and the remainder of the OB. However, because the partition $K = 3$ shows a further split between the anterior dorsal and the posterior ventral parts of the original posterior cluster (cluster 2 in Figure 4a), there may be an important difference between those areas in terms of the odorant response. In terms of variance minimization, the transition from $K = 2$ to $K = 3$ produces a gain that comes from a further subdivision of the objects along the second principal component (Figure 3b).

Although the most stable partition is associated with the global minimum of fBPSE variability curve ($K = 3$), we have pointed out in a former theoretical work (Falasconi et al. 2010) that local minima of variability can also represent significant "local best" solutions, cluster solutions, which are more stable and hence more reliable when compared with nearest neighbor partitions. Partition $K = 6$ (Figure 4c) fits this condition. This partition has lower variability value as compared with lower order ($K = 4, 5$) and higher order ($K = 7, 8$) partitions (see Figure 5a). The absolute value of the local minimum is different for the 2 data sets, being shallow for data set D2, but the 2 trends are perfectly consistent. These data further support the partition significance at local level.

Moreover, detailed investigations indicate that one of the major factors delivering higher partition variability arises from pixels belonging to cluster borders. When the number of clusters increases, the number of border pixels increases significantly while the size of individual clusters diminishes.

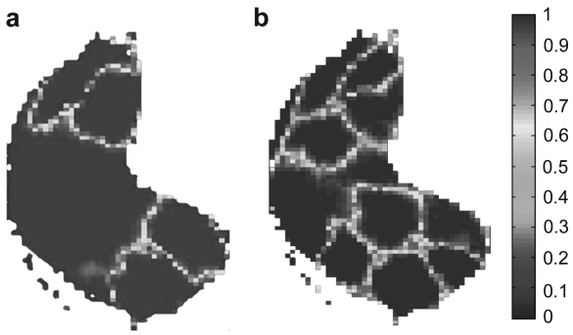


**Figure 6** Best partition of data (both for D1 and D2) in $K = 3$ clusters according to the fBPSE criterion. (**a**) Hard partition obtained from the fuzzy clustering. (**b**) Maximum value of the fuzzy membership of each pixel where values equal to one stand for core pixels. This figure appears in color in the online version of *Chemical Senses*.

This is shown in Figure 6, where the only fuzzy partitions corresponding to 3 and 6 clusters are projected onto the olfactory bulb map. These fuzzy partitions can be contrasted with corresponding hard partitions (respectively, Figure 4b and c) obtained by converting the fuzzy outcome according to the maximum membership rule, where each pixel is assigned to the cluster having its maximum membership. To give an insight into the "fuzziness" of the cluster borders, we show the maximum membership value of each pixel in Figure 6a and b. Dark regions represent the core pixels of every cluster, for which the membership is equal to one. Observe that the number of cluster pixels belonging to borders is larger for $K = 6$ partition than for $K = 3$, which could be a reason for smoothing the depth of the local minimum.

Relevant arguments in favor of the hypothesis that OB glomeruli are clustered according to their chemical response to different odorants also come from the absolute values of fBPSE variability. Though the fBPSE method is not suited to assess the absence of clusters in the data set ($K = 1$ value), previous theoretical analysis (Falasconi et al. 2010) supported the hypothesis that low (absolute) variability values are consistent with the existence of a clustering structure underlying the data. Conversely, variability values close to 1 would argue against the clustering assumption.

For this reason, while global and local minima of fBPSE variability give indications of more relevant partitions, taking into account the dimensionality of the data set, variability around 0.3 (as shown in Figure 5a and b) may indicate the existence of a cluster structure even for high $K$ values ($K > 8$), despite not being fully stable to bootstrap perturbations of the pixel sampling. The present analysis shows that from a statistical point of view, it is not possible to infer a large number of modules (i.e., 15 clusters have no more information than 14 or 16 clusters) due to the stabilization of the bootstrap variability beyond 10 clusters.

Finally, we can get additional information from the comparison of validity results obtained from the 2 data sets D1 and D2 because both the trend of the curves and absolute variability values were very similar (Figure 5). Indeed, the results also were very similar for the individual FCM outcomes. This result suggests that the presence of multiple maps involving the same odorant had negligible effects on the FCM clustering, and hence, on the stability of the partitions. This conclusion also was valid for other clustering parameters, such as different fuzzifier values (data not shown). These results are consistent with previous observations that most odorants evoke replicable glomerular patterns in different experiments and at different concentrations in the same experiment (Johnson and Leon 2007).

### *Investigation of hierarchical structure of data*

The FCM algorithm is a partitional clustering algorithm, which does not allow, in principle, the production of cluster hierarchies (nested sets of nonoverlapping clusters). However, it is possible to determine whether there is any inclusion relationship within response clusters (nested sets even with partial overlap of cluster borders) between partitions obtained at subsequent $K$ values, which denotes a possible hierarchical structure of the data.

Comparing the clusters obtained as $K$ increases (Figure 3a–d), we can observe a tendency of clusters that appear for higher $K$ values to subdivide the area of clusters maintaining their original boundaries. Figure 3e–g shows the overlay of the resulting clusters for $K = 3, 6, 10$ with the boundaries of the clusters obtained with the lower value of $K = 2, 3, 6$. According to previous analysis, we selected $K = 3$ and $K = 6$ as best global and local partitions, respectively, while $K = 10$ was selected to be the point at which the variability starts to remain constant across the set of partitions. In Figure 3e, cluster 2a is subdivided into clusters 2b and 3b. Notice that we have also the same subdivision for the clusters in the lateral aspect ($2a' \rightarrow 2b' + 3b'$). We report below only subdivisions in the medial region, even though each of them has an equivalent subdivision in the lateral region.

Figure 3f shows that

i. Cluster 1b, mapping the entire ventral region for $K = 3$, becomes clusters 1c, 5c–c′, 6c–c′ for $K = 6$.
ii. Other clusters are nested inside larger ones, for example, 2c within 2b and 3c within 3b.
iii. Only cluster 4c–c′ emerges on the border of the previous partition.

In the comparison of clusters for $K = 6$ and $K = 10$ (Figure 3g), we find the following hierarchical relationships:

i. $1c \rightarrow 1d + 10d$,
ii. $3c \rightarrow 3d + 8d + 9d$, and
iii. $2c \rightarrow 2d + 7d$.

Certain clusters remained almost unchanged, such as clusters 4, 5, and 6. In this respect, we have observed that some clusters emerge at partitions involving low $K$ values and have

a long ''lifetime'' in terms of $K$, they survive almost unaltered (either for centroid position, shape, and size) up to relatively high $K$ values before splitting into smaller clusters.

This is a further concept of ''stability'' related to individual clusters that can be visually ascertained. The cluster ''invariance'' in terms of cluster shape and position can be taken as an indicator of particular validity of that cluster, albeit an objective, though not quantitative one. For example, the cluster that corresponds to modules 4c–4c′ in Figure 4c originates for the first time at $K = 5$ and remains almost identical up to $K = 10$ (module 4d–d′ in Figure 4d). Similar arguments hold for both modules 5c–c′ and 6c–c′ that come first at $K = 6$ and gets unaltered until $K = 10$ also.

This finding may offer new perspectives for further validity studies of single clusters that have not yet been addressed. In this report, however, we have focused on a different process of validation involving the entire partition; hence, the fBPSE approach cannot be used to assess individual cluster validity quantitatively.

These results strongly suggest

a) the existence of a hierarchical organization in the data, which leads to certain clusters fragmenting in smaller ones across subsequent partitions;
b) there could exist a relationship—perhaps connected to the chemical meaning of clusters—between certain modules that combined together permit one to recover larger clusters found in lower order partitions; and
c) the existence of certain individual clusters that are relatively more stable across the set of partitions and that can probably contribute more than others to the overall partition stability.

## Response clusters and olfactory coding

To evaluate the relationships between odorant chemistry and each of the clusters that we have identified, our first approach was to rank each odorant in terms of the average activity value evoked within each cluster. Because each odorant was characterized by various molecular feature descriptors, we then could evaluate the correspondence between these features and activity within the clusters.

In the following subsections, we detail our consideration of the clusters found in the most stable partition $K = 3$ (clusters 1b, 2b, and 3b at Figure 3b) and for the local best partition $K = 6$ (clusters 1c, 2c, 3c, 4c, 5c, and 6c at Figure 3c). Lateral and medial aspects of the same cluster (b–b′ and c–c′) were considered together for our analysis, but we only labeled the cluster with a nonindexed letter for the sake of simplicity.

### *Clusters 1b and 1c, 5c, 6c (ventral and anterior aspects)*

Only 25 of 308 odorants evoked significantly greater ($P < 0.01$) activity values inside cluster 1b compared with outside the cluster (Supplementary Figure S1). Nineteen of these 25 odorants were aliphatic odorants with 8 or more carbons, independent of the particular functional group. Indeed, when analyzed with respect to their functional groups, none of the classes evoked significantly greater activity inside the cluster compared with outside it (Table 1). Part of the reason for the low level of significance might be the large size of the cluster compared with the focal activity occurring within it for many odorant stimuli (Supplementary Figure S2).

We then explore the response clusters for the finer partition $K = 6$ nested within this region and found out that cluster 5c was rather specific and mostly associated with aldehydes and alkanes (Table 2, yellow cells); these compounds constitute the majority of the top-ranked odorants activated within this. Conversely, cluster 6c was most associated with carboxylic acids (Table 2, green cells). Indeed, cluster 6c is specifically activated by only those compounds: 12 of 15 of top-scoring odorants were carboxylic acids.

Cluster 1c, mapping the inner most region of ventral aspect, was activated by only few odorants (table provided as Supplementary Material). The top ranking compounds are related to aliphatic compounds with more than 8 carbons, yet, there are not chemical features emerging to be significant within this cluster (Table 2). These data are consistent with our preliminary analysis of variance, which shows nearly zero variance across the entire set of odorants in this region of the OB (Figure 2b).

### *Clusters 2b and 2c (posterior, dorsal aspect)*

One hundred and sixty-five top-ranked odorants showed significantly higher activity inside the cluster 2b than outside cluster ($P < 0.01$; Supplementary Figure S3). The top 50 odorants in that cluster included mainly aromatic compounds (Table 1, yellow cells) and ketones, among which were the 2 aromatic ketones 2-acetylpyridine and acetophenone (Table 1, green cells).

Significant results for aldehydes and esters probably reflect the presence of these functional groups in aromatic compounds because aliphatic odorants bearing these groups did not show a significant association within the cluster (Table 1, gray cells). Results for ethers probably also reflect aromatic compounds, because almost all of the ethers in the database also were aromatic. Heterocyclic odorants, alicyclic odorants, secondary alcohols, polycyclic odorants, and hydrocarbons also showed significantly higher activity inside the cluster than outside, although their average scores were considerably lower than those for aromatics and ketones (cf. Table 1).

Cluster 2c was activated by about 140 odorants (Supplementary Figure S4). Aromatic odorants, and in particular, aromatic with oxygen-containing functional groups, aromatic aldehydes, and aromatic esters showed the highest activity values within cluster 2c (Table 2, blue cells), and activity evoked by such odorants significantly exceeded the activity they evoked outside the cluster (data not shown). Consistent with his hierarchical organization within cluster

**Table 1** Score of selected molecular features within the 3 main valid clusters of partition $K$= 3

| Loading (molecular features) | Cluster 1b | | | Cluster 2b | | | Cluster 3b | | |
|---|---|---|---|---|---|---|---|---|---|
| | Score IN | Score OUT | *P* value | Score IN | Score OUT | *P* value | Score IN | Score OUT | *P* value |
| Carboxylic acid | −0.3766 | −0.0328 | 1 | −0.2323 | −0.2267 | 1 | **0.6774** | **−0.3159** | **0** |
| Alcohol (not phenol) | −0.3100 | 0.4142 | 1 | **0.1633** | **−0.1252** | **0** | **0.777** | **−0.1600** | **0** |
| Primary alcohol | −0.2507 | 0.4005 | 1 | 0.0378 | 0.0349 | 1 | **0.7193** | **−0.1575** | **0** |
| Aliphatic (not alicyclic) secondary alcohol | −0.4503 | 0.5447 | 1 | **0.3105** | **−0.2579** | **0** | **0.9505** | **−0.2536** | **0** |
| Ester (not lactone) | −0.2530 | 0.3812 | 1 | **0.1358** | **−0.0605** | **0.0002** | **0.7678** | **−0.1142** | **0** |
| Aliphatic ester (not alicyclic) | −0.2682 | 0.4183 | 1 | −0.0339 | −0.0579 | 1 | **1.0334** | **−0.1895** | **0** |
| Aromatic ester | −0.2576 | 0.3755 | 1 | **0.5968** | **−0.1425** | **0** | **0.2162** | **−0.0959** | **0** |
| Aldehyde | −0.1344 | 0.2788 | 1 | **0.2928** | **0.0477** | **0** | **0.2678** | **0.0439** | **0** |
| Aliphatic aldehyde | −0.1029 | 0.1615 | 1 | −0.1951 | 0.2017 | 1 | **0.5946** | **−0.1497** | **0** |
| Aromatic aldehyde | −0.3265 | 0.2964 | 1 | **0.7120** | **−0.256** | **0** | −0.0331 | −0.138 | 1 |
| Ether | −0.1922 | 0.3470 | 1 | **0.4483** | **−0.1253** | **0** | **0.1861** | **−0.0566** | **0** |
| Ketone | −0.3450 | 0.5342 | 1 | **0.504** | **−0.2405** | **0** | **0.5788** | **−0.2302** | **0** |
| Aliphatic ketone (not alicyclic) | −0.3591 | 0.5355 | 1 | **0.4156** | **−0.2411** | **0** | **0.7135** | **−0.2300** | **0** |
| Aliphatic or alicyclic with multiple O-containing functional groups | −0.3263 | 0.3030 | 1 | **0.0901** | **−0.1996** | **0.007** | **0.9010** | **−0.1997** | **0** |
| Aliphatic or alicyclic hydrocarbon | −0.3925 | 0.5431 | 1 | **0.221** | **−0.0023** | **0** | **0.8450** | **−0.1386** | **0** |
| Alkane | −0.3161 | 0.4830 | 1 | −0.0185 | 0.083 | 1 | **0.9353** | **−0.2166** | **0** |
| Aromatic | −0.3181 | 0.4707 | 1 | **0.6721** | **−0.2364** | **0** | **0.1807** | **−0.199** | **0** |
| Aromatic with O-containing substituent | −0.2999 | 0.3105 | 1 | **0.6224** | **−0.1837** | **0** | **0.1126** | **−0.1336** | **0** |
| Alicyclic | −0.2721 | 0.4154 | 1 | **0.3048** | **−0.1487** | **0** | **0.5599** | **−0.1394** | **0** |
| Polycyclic | −0.2286 | 0.3263 | 1 | **0.1784** | **−0.0949** | **0** | **0.4627** | **−0.1061** | **0** |
| Heterocyclic | −0.2890 | 0.4320 | 1 | **0.3866** | **−0.1853** | **0** | **0.4775** | **−0.1667** | **0** |

For every cluster, the score values inside (score IN) can be compared with scores outside (score OUT). Bold values mean significantly higher activity of glomerular inside the cluster than outside with $P < 0.01$ (zeros mean $P < 0.0001$ while above such threshold exact values are reported). Colored cells are referred in the paper text. This table appears in color in the online version of *Chemical Senses*.

2b, ketones and heterocyclic compounds also evoked more activity inside cluster 2c, but the average activity was lower than for the aromatic compounds.

### *Clusters 3b and 3c, 4c (posterior domain)*

A total of 258 odorants of 308 showed significantly greater activity inside cluster 3b compared with outside the cluster ($P < 0.01$) (Supplementary Figure S5). The 50 odorants evoking the greatest average activity in the cluster included 21 esters. Indeed, as a class, esters scored highest in their activation of the cluster (Table 1, red cells), although nearly every molecular class (except aliphatic aldehydes) evoked significantly greater activity inside the cluster compared with outside it (Table 1). This result is entirely consistent with our previous finding that the highest mean activity across all odorants occurs in this area (Figure 2a).

The aromatic odorants that were the best stimuli for cluster 2b were among the weakest stimuli for cluster 3b (Table 1, orange cells), although in most cases, they showed significantly greater average activity inside the cluster compared with outside. Similar observations hold for clusters 3c and 4c, which are hierarchically nested within cluster 3b. Both clusters 3c and 4c respond to more than 220 odorants and both clusters seem rather nonspecific in their response to specific chemical features; all features considered in Table 2 are significantly activated for both clusters ($P < 0.01$). Despite this finding, the most posterior cluster, 3c, shows higher activity values for aliphatic odorants than for aromatic compounds (cf. orange and gray cells in Table 2). On the contrary, cluster 4c mirrors cluster 3b in its partial specificity toward aromatics, consistent with the fact that this cluster emerges on the boundary between 2b and 3b superclusters. In fact, we note that 8 of the 13 aromatic hydrocarbons in the database were among the top 50 stimuli for cluster 4c.

### *Comparisons to previous models*

The objective partitions of bulb responses revealed in our data show some similarities and some differences compared with previously proposed domain organizations. One recognized subdivision of the bulb involves the presence of dorsal and ventral zones defined by the projections of olfactory sensory neurons expressing different types of odorant receptor genes (Mori et al. 2006). Although the ventral projection zone(s) would be contained mostly in our cluster 2a at $K = 2$ (Figure 3a) and our cluster 1b at $K = 3$ (Figure 3b), our partitions are different from the ventral zone(s) in that they appear to wrap around the dorsal surface of the anterior part of the bulb. Any further comparison of the relationship between these boundaries probably will require mapping the markers of these expression zones into the Johnson and Leon odorant response matrix.

Figure 7 shows alignments between clusters identified in the current study compared with those odorant response modules identified visually (Johnson and Leon 2007). Cluster 1b was found to contain 3 nested subclusters for $K = 6$ (Figure 3f). Analyses of these clusters confirm a clear

**Table 2** Score of selected molecular features within the 3 main valid clusters of partition $K = 3$

| Loading (molecular features) | Cluster 1c | | Cluster 2c | | Cluster 3c | | Cluster 4c | | Cluster 5c | | Cluster 6c | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Score | $P < 0.01$ | Score | $P < 0.01$ | Score | $P < 0.01$ | Score | $P < 0.01$ | Score | $P < 0.01$ | Score | $P < 0.01$ |
| Carboxylic acid | −0.5001 | 1 | −0.2650 | 1 | **0.8107** | **0** | **−0.0636** | **0.006** | −0.3310 | 1 | **0.6914** | **0** |
| Alcohol (not phenol) | −0.4024 | 1 | 0.0640 | 0.5 | **0.7501** | **0** | **0.5119** | **0** | −0.0029 | 1 | −0.2892 | 1 |
| Primary alcohol | −0.3760 | 1 | −0.0466 | 1 | **0.7446** | **0** | **0.3538** | **0** | 0.1084 | 0.12 | −0.0859 | 1 |
| Aliphatic (not alicyclic) secondary alcohol | −0.4745 | 1 | **0.2338** | **0** | **0.8174** | **0** | **0.7418** | **0** | −0.1730 | 1 | −0.6127 | 1 |
| Ester (not lactone) | −0.4102 | 1 | 0.0516 | 1 | **0.7394** | **0** | **0.5061** | **0** | 0.0306 | 1 | −0.1351 | 1 |
| Aliphatic ester (not alicyclic) | −0.4664 | 1 | −0.1472 | 1 | **0.9933** | **0** | **0.5653** | **0** | 0.0962 | 0.9 | −0.1999 | 1 |
| Aromatic ester | −0.3647 | 1 | **0.7547** | **0** | **0.1725** | **0.0001** | **0.3226** | **0** | −0.1812 | 1 | 0.1253 | 0.025 |
| Aldehyde | −0.3950 | 1 | **0.3541** | **0** | **0.2088** | **0** | **0.2214** | **0** | **0.1723** | **0** | **0.2490** | **0** |
| Aliphatic aldehyde | −0.4332 | 1 | −0.2508 | 1 | **0.6074** | **0** | 0.0939 | 0.2 | **0.6179** | **0** | 0.1625 | 0.1 |
| Aromatic aldehyde | −0.3986 | 1 | **0.9526** | **0** | −0.1325 | 1 | **0.2442** | **0** | −0.3234 | 1 | **0.3358** | **0** |
| Ether | −0.2479 | 1 | **0.5134** | **0** | **0.1266** | **0.0045** | **0.2925** | **0** | −0.1687 | 1 | **0.1224** | **0.007** |
| Ketone | −0.3607 | 1 | **0.5034** | **0** | **0.4886** | **0** | **0.6829** | **0** | −0.2517 | 1 | −0.4225 | 1 |
| Aliphatic ketone (not alicyclic) | −0.4181 | 1 | **0.3965** | **0** | **0.6419** | **0** | **0.6177** | **0** | −0.0966 | 1 | −0.4809 | 1 |
| Aliphatic or alicyclic with multiple O-containing functional groups | −0.3980 | 1 | 0.0550 | 0.4 | **1.0166** | **0** | **0.3640** | **0** | −0.2542 | 1 | −0.1340 | 1 |
| Aliphatic or alicyclic hydrocarbon | −0.4868 | 1 | 0.0633 | 1 | **0.6785** | **0** | **0.8740** | **0** | **0.2792** | **0** | −0.8239 | 1 |
| Alkane | −0.4117 | 1 | −0.2361 | 1 | **0.8578** | **0** | **0.6650** | **0** | **0.4700** | **0** | −0.7308 | 1 |
| Aromatic | −0.3602 | 1 | **0.7110** | **0** | **0.0564** | **0.002** | **0.5243** | **0** | −0.3418 | 1 | −0.0570 | 1 |
| Aromatic with O-containing substituent | −0.3699 | 1 | **0.7942** | **0** | 0.0793 | 0.3 | **0.2706** | **0** | −0.2865 | 1 | **0.1727** | **0** |
| Alicyclic | −0.2585 | 1 | **0.2312** | **0** | **0.4003** | **0** | **0.7312** | **0** | −0.2905 | 1 | −0.2353 | 1 |
| Polycyclic | −0.2132 | 1 | **0.0951** | **0** | **0.3963** | **0** | **0.6695** | **0** | −0.1612 | 1 | −0.3360 | 1 |
| Heterocyclic | −0.3341 | 1 | **0.4006** | **0** | **0.4003** | **0** | **0.4913** | **0** | −0.3164 | 1 | −0.1068 | 1 |

For every cluster, the score values inside the cluster are reported (score in) together with its significance value (zeros mean $P < 0.0001$). Bold values are significant at $P < 0.01$. Colored cells are referred in the paper text. This table appears in color in the online version of *Chemical Senses*.

correspondence with the results obtained by Johnson and Leon through visual association of glomerular modules to chemical properties. Cluster 1c (Figure 7e) overlapped with the ventral-most part of the module responding to hydrocarbon chains in a variety of aliphatic compounds (Johnson and Leon 2007), modules H and I in Figure 7a. This region was best activated by aliphatic compounds of more than 8 carbons in the current analysis, a finding that is consistent with the observed ventral progressions of activity with increasing carbon number seen in this area in several systematic studies involving straight-chained compounds of differing molecular length (Johnson et al. 1999, 2004; Ho et al. 2006).

Cluster 5c partially overlaps with a module previously identified as responding to primary alcohols, aldehydes, and phenols as well as with the dorsal part of the module previously reported to respond to hydrocarbon chains in many aliphatic compounds (cf. Figure 7a and e). Indeed, we found with our objective analysis that cluster 5c responded best to aldehydes, alkanes, and alicyclic compounds. Cluster 6c showed a very good correspondence with the carboxylic acid-preferring module (cf. Figure 7a and e) described by Johnson and Leon (2007), and we also confirmed the specificity of cluster 6c for carboxylic acids.

Cluster 2b corresponded quite well to a module previously proposed to respond robustly to ketones and aromatic compounds with oxygenic functional groups (Johnson and Leon 2007), modules Cc in Figure 7a, showed a specificity that we also confirmed with this objective analysis. In addition, we found that this cluster also responds to heterocyclic odorants, secondary alcohols, and hydrocarbons, a finding that had not been appreciated previously. This conclusion may be due to the fact that cluster 2b also overlaps with module Dd (cf. Figure 7a and e).

Cluster 4c was found to correspond in location to modules proposed to respond to aromatic hydrocarbons (Johnson and Leon 2007), modules Dd in Figure 7a, a specificity we have confirmed here. Cluster 3b overlapped with 2 of the visually defined modules of Johnson and Leon (2007), one of which was associated with esters, modules Ee in Figure 7a and the other that was reported to respond to water-soluble odorants, modules Ff in Figure 7a, independent of functional group (Johnson and Leon 2007). Interestingly, even though subclusters were not resolved into stable partitions by FCM, the 2 separate modules of Johnson and Leon (2007) did correspond to distinct areas of high variance (high information content) in the data (Figure 7c), and there is a suggestion of a corresponding subdivision in the overall mean activity (Figure 7b), suggesting that the modules likely have a true foundation in the data.

Consistent with the proposal of Johnson and Leon (2007) that the determinants of activity in the posterior bulb are not specific to a particular functional group, we found that odorants of many classes showed significantly higher activity inside cluster 3b than outside (Table 1). As shown in Table 1 (red cells), the highest ranking odorants for cluster 3b were (aliphatic) esters, which is consistent with the fact that the cluster contains the ester module reported by Johnson and Leon (2007). The most posterior module G that was associated with water solubility in the analysis of Johnson Ong, et al. (2007) was not entirely included in our cluster 3b due to

missing values at the edge of the maps for individual odorants, and the cluster included a large area that was not previously included in the module (Johnson Ong, et al. 2007; Figure 7).

## Discussion

Perhaps the first point to be made regarding this analysis is that it addresses the importance of dealing with imaged data just like any other kind of data. Images of glomerular events are often shown as conclusive evidence of a valid olfactory response. However, there are many reasons why a transient increase in activity might not actually be a coded olfactory response. For example, there may be an intense glomerular response distant to the recorded glomerulus and the reported response may not be the one that the animal actually uses to perceive that odor. That is, one may be looking at a secondary or tertiary odorant response. Similarly, one may be recording a response to a contaminant that is not perceived concurrently with the dominant odor to which the individual is being exposed. Also, background odorants may be recorded that are unrelated to the perceived odor. It is similarly possible that the activity of the glomeruli could be driven by centrifugal activity, rather than only by olfactory signals. Moreover, it is not possible to know with many reported images whether the glomerular response has been artificially trimmed to resemble the size of a glomerulus, without actually matching it to a specific glomerulus. The variance associated with these alternative factors for evoking responses either in spatial terms or the variability associated with individual differences also are not known if the data is simply shown as a picture. Our technique has single glomerulus resolution and the spread in activity that we see is due both to the variability across individuals, just as one would expect in any other kind of data as well as the broad responsiveness that many odorants evoked across the glomerular layer (Johnson and Leon 2007). This variability across individuals has as a consequence a local spread and averaging of glomerular activity in our images because we are averaging across different individuals. However, the impact that this

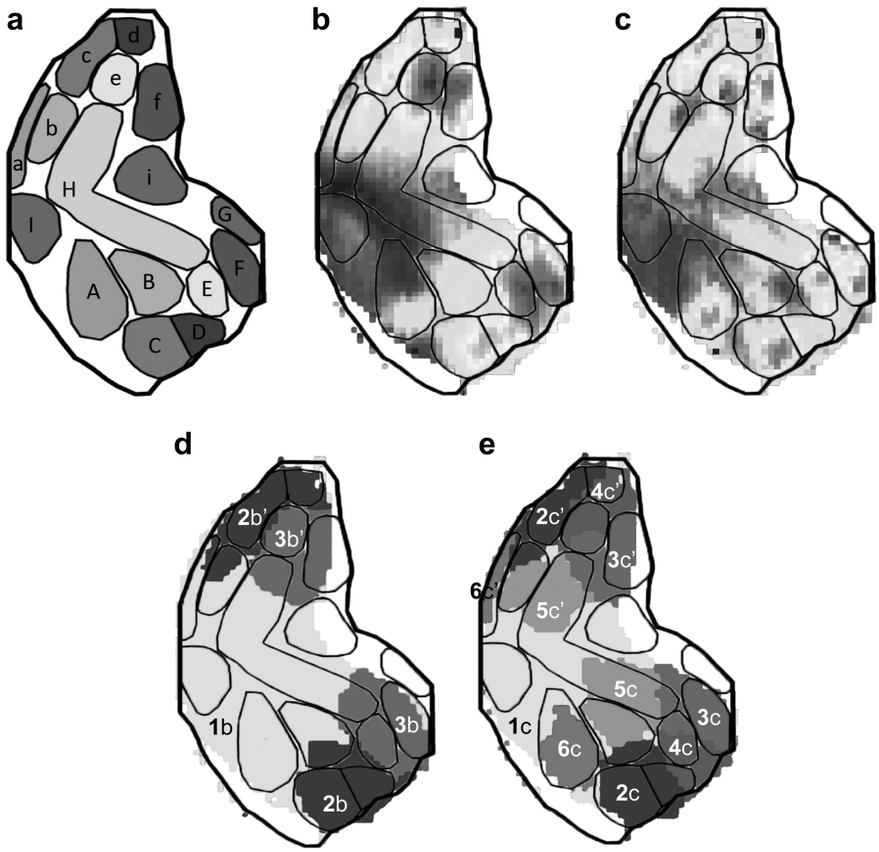


**Figure 7** Comparison of previously proposed modules of the olfactory bulb with results obtained in this paper. (**a**) Modules proposed by Johnson and Leon (2007) with the following correspondence between modules and odorants: Aa—carboxylic acids, methyl and ethyl esters; Bb—primary alcohols, aldehydes, phenols; Cc—aromatics with O groups, high concentration of ketones; Dd—aromatic hydrocarbons; Ee—aliphatic esters; Ff—highly water-soluble compounds; Gg—septal organ projection, broadly responsive; Ii—methyl-substituted bicyclic compounds; H—aliphatic hydrocarbon chain; the following figures show overlays with this modules; (**b**) modules over mean activity of the image data set; (**c**) modules over variance of the images data set; (**d**) modules over stable clusters found with FCM clustering and $K = 3$; and (**e**) modules over clusters found with FCM clustering for $K = 6$. This figure appears in color in the online version of *Chemical Senses*.

variability may have on the results obtained in this study is limited. First, because the loss of resolution that the averaging produces is much smaller than the size of the clusters reported. Second, methods based on fuzzy logic, such as the one used in this work, are more robust than other classical methods against this loss of resolution.

The use of a data matrix rather than activity-related images of the bulb allows one to treat imaged data with the same statistical approaches as one might use to describe similarities and differences across individuals on any metric. In our previous work, we have clustered the focal glomerular responses in these matrices using visual cues, along with statistical comparisons of their clustering across odorants. The objective computational approach that we report here reinforces the conclusions that we have made regarding the organization of odorant responses in the glomerular layer of the rat olfactory bulb.

The second point that should be made concerns the importance of using a wide range of odorants and measuring the response of the entire glomerular layer to such stimulation to be able to tease apart the complexities of a system that relies on a vast number of odorants that vary in multiple dimensions, along with a remarkable number of olfactory receptors that underlie the first step in olfactory coding. Given the broad response pattern across much of the bulb to many odorants (Woo et al. 2007), it would be easy to come to the conclusion that secondary or background responses distant to the primary coding response were critically involved in a perceptual event. Considering responses across the entire bulb and stimulating the olfactory system with hundreds of odorants allows one to consider the responses that are critical for perception in relation to the secondary or background responses. Indeed, we have made the case previously that the critical perceptual signal evolves from a relational response wherein the primary responses are perceptually significant only with respect to other responses in the glomerular layer (Cleland et al. 2007). However, the absolute glomerular response to odorants is typically reported rather than the relative response, increasing the possibility of reporting responses that are unrelated to olfactory perception.

In addition, much of what is reported in the literature regarding glomerular responses is restricted to a small area on the dorsal aspect of the olfactory bulb. Although any responses that are seen in that area must be understood relative to the responses seen throughout bulb, these responses are typically regarded as critically related to the perception without any behavioral evidence of their importance in that regard. The use of the relative responses in this report has been validated by effectively correlating such activity to differences in perceptual behavior (Youngentob et al. 2006).

Bozza et al. (2004), along with Soucy et al. (2009), looked at that relatively small area on the dorsal bulb and reported that they found no chemotopy in that area. It is possible that this area of the bulb has no chemotopy or that there is no chemotopy in the entire bulb or that these studies were done with insufficient rigor to identify clusters of similar odorants that reliably activate that area. Other reports of activity in that part of the bulb have reported chemotopy (Rubin and Katz 1999; Wachowiak and Cohen 2001), which could have been due to the existence of chemotopy for the odorants that were tested or that these responses were actually artifacts. Indeed, much more careful work has shown that this area has highly specific clustered responses for odorants associated with predators and spoiled foods that were untested in the other reports (Kobayakawa et al. 2007; Matsumoto et al. 2010). Moreover, the importance of this area for coding these odorants was verified with destruction of the area and the consequent inability of these mice to avoid those odors. It is clear that similar odorants are processed in similar places in the bulb (chemotopy) and that these responses play a critical role in odor perception (Mori et al. 2006; Johnson and Leon 2007; Kobayakawa et al. 2007; Leon and Johnson 2009; Matsumoto et al. 2010).

Our data also support the notion of an identity code operating in the olfactory system, which underlies the first step in odorant perception. An identity code, in which specific olfactory receptor neurons play a critical role in mediating odor perception, would have to have differential responses among different olfactory receptor neurons. Because those neurons expressing the same olfactory receptor cluster reliably in particular glomeruli, one would expect relatively reliable responses to be seen in spatially distinct areas in the glomerular layer. The data that we present in this report reinforce such a conclusion.

It is also worth pointing out again that such conclusions are reinforced by similar experiments that monitor rapid glomerular responses (Mori et al. 2006). Electrical stimulation of the glomerular layer similarly supports the notion of an identity code because rats can distinguish between widely spaced electrical microstimuli and have more difficulty discriminating between closely placed electrodes (Mouly et al. 1985; Mouly and Holley 1986). The conclusions of the electrical stimulation experiments closely reinforce the conclusions that we have reached in this report because closely related odorants, which evoke spatially similar responses, are more difficult to discriminate than distantly related odorants, which evoke a similar responses in the glomerular layer of the bulb (Youngentob et al. 2006). It is therefore possible to predict the odorant from the neuronal response, the neural response from the odorant stimulating the system, as well as predicting the behavioral response from the neural response to an odorant. Elimination of areas of high activity in the bulb eliminates the responsiveness to the odorant and the relationships between the perceptions of the odors can be mimicked by electrical stimulation. Together, these data provide strong support for the notion of an identity code in the mammalian olfactory system.

In summary, we attempted to further our understanding of the functional organization of the olfactory bulb by applying fuzzy cluster analysis techniques to investigate glomerular

response patterns to diverse odorants. Toward that end, we also solved the challenging problem of validating the achieved partitions by using an original criterion based on fuzzy partition stability. We found that the organization of responses in the glomerular layer is indeed chemotopic, in that, groups of odorants sharing aspects of their chemical structure are associated with glomerular response subclusters using an objective means of clustering their responses. We further obtained evidence that the organization of the glomerular response is partially hierarchical, in that, there are some clusters that can be further subdivided into smaller regions showing higher specificity toward narrower subgroups of odorants. These data are consistent with the conclusion that there is an identity code at work in the mammalian olfactory system.

## Supplementary material

Supplementary material can be found at http://www.chemse.oxfordjournals.org/.

## Funding

This work was partially supported by the European Network of Excellence GOSPEL General Olfaction and Sensing Projects on a European Level (FP6-IST-2002-507610). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216916. Partial support has been given by CNR under the Short-Term Mobility Program 2007. M.L. and B.A.J. have been supported by United States Public Health Service Grants DC03545, DC006391, and DC006516.

## References

Bezdek JC. 1981. Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press.

Borgelt C. 2007. Resampling for fuzzy clustering. Int J Uncertain Fuzz Knowl Based Syst. 15:595–614.

Bozza T, McGann JP, Mombaerts P, Wachowiak M. 2004. In vivo imaging of neuronal activity by targeted expression of a genetically encoded probe by the mouse. Neuron. 42:9–21.

Campello RJGB. 2007. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. Pattern Recogn Lett. 28:833–841.

Cleland TA, Johnson BA, Leon M, Linster C. 2007. Relational representation in the olfactory system. Proc Natl Acad Sci U S A. 104:1953–1958.

Costanzo RM, Kobayashi M. 2010. Age-related changes in P2 odorant receptor mapping in the olfactory bulb. Chem Senses. 35:417–426.

Dunn JC. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern. 3:32–57.

Falasconi M, Gutiérrez A, Pardo M, Sberbegliery G, Marco S. 2010. A stability based validity method for fuzzy clustering. Pattern Recogn. 43:1292–1305.

Friedman JH. 1994. An overview of predictive learning and function approximation. In: Cherkassky V, Friedman JH, Wechsler H, editors. From statistics to neural networks. Theory and pattern recognition applications. Berlin: Springer Verlag. p. 1–61.

Ho SL, Johnson BA, Leon M. 2006. Long hydrocarbon chains serve as unique molecular features recognized by ventral glomeruli of the rat olfactory bulb. J Comp Neurol. 498:16–30.

Johnson BA, Arguello S, Leon M. 2007. Odorants with multiple oxygen-containing functional groups and other odorants with high water solubility preferentially activate posterior olfactory bulb glomeruli. J Comp Neurol. 503:1–34.

Johnson BA, Farahbod H, Xu Z, Saber S, Leon M. 2004. Local and global chemotopic organization: general features of the glomerular representations of aliphatic odorants differing in carbon number. J Comp Neurol. 480:234–249.

Johnson BA, Leon M. 2000. Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. J Comp Neurol. 422:496–509.

Johnson BA, Leon M. 2007. Chemotopic odorant coding in a mammalian olfactory system. J Comp Neurol. 503:1–34.

Johnson BA, Ong J, Lee K, Ho SL, Arguello S, Leon M. 2007. Effects of double and triple bonds on the spatial representations of odorants in the rat olfactory bulb. J Comp Neurol. 500:720–733.

Johnson BA, Woo CC, Hingco EE, Pham KL, Leon M. 1999. Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. J Comp Neurol. 409:529–548.

Johnson BA, Woo CC, Leon M. 1998. Spatial coding of odorant features in the glomerular layer of the rat olfactory bulb. J Comp Neurol. 393:457–471.

Kobayakawa K, Kobayakawa R, Matsumoto H, Oka Y, Imai T, Ikawa M, Okabe M, Ikeda T, Itohara S, Kikusui T, et al. 2007. Innate versus learned odour processing in the mouse olfactory bulb. Nature. 450:503–508.

Leon M, Johnson BA. 2003. Olfactory coding in the mammalian olfactory bulb. Brain Res Rev. 42:23–32.

Leon M, Johnson BA. 2009. Spatial representations of odorants in olfactory bulbs of rats and mice: Similarities and differences in chemotopic organization. J. Comp. Neurol. 514:658–673.

Matsumoto KK, Kobayakawa R, Tashiro T, Mori K, Sakano H, Mori K. 2010. Spatial arrangement of glomerular molecular-feature clusters in the odorant-receptor class domains of the mouse olfactory bulb. J Neurophysiol. 103:3490–3500.

Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmonson J, Axel R. 1996. Visualizing an olfactory sensory map. Cell. 87:675–686.

Mori M, Takahashi YK, Igarashi KM, Yamaguchi M. 2006. Maps of odorant molecular features in the mammalian olfactory bulb. Physiol Rev. 86:409–433.

Mouly AM, Holley A. 1986. Perceptive properties of the multisite electrical microstimulation of the olfactory bulb in the rat. Behav Brain Res. 21:1–12.

Mouly AM, Vigouroux M, Holley A. 1985. On the ability of rats to discriminate between microstimulations of the olfactory bulb in different locations. Behav Brain Res. 17:45–58.

Ressler KJ, Sullivan SL, Buck LB. 1993. A zonal organization of odorant receptor gene expression in the olfactory epithelium. Cell. 73:597–609.

Ressler KJ, Sullivan SL, Buck LB. 1994. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. Cell. 79:1245–1255.

Rubin RD, Katz LC. 1999. Optical imaging of odorant representations in the mammalian olfactory bulb. Neuron. 23:499–511.

Salcedo E, Zhang C, Kronberg E, Restrepo D. 2005. Analysis of training-induced changes in ethyl acetate odor maps using a new computational tool to map the glomerular layer of the olfactory bulb. Chem Senses. 30:615–626.

Shepherd GM. 1987. A molecular vocabulary for olfaction. Ann N Y Acad Sci. 510:98–103.

Soh Z, Tsuji T, Takiguchi N, Ohtake N. 2011. An artificial neural network approach for glomerular activity pattern prediction using the graph kernel method and the Gaussian mixture functions. Chem Senses. 36:413–424.

Soucy ER, Albeanu DF, Fantana AL, Murthy VN, Meister M. 2009. Precision and diversity in an odor map on the olfactory bulb. Nat Neurosci. 12:210–220.

Treloar HB, Feinstein P, Mombaerts P, Greer CA. 2002. Specificity of glomerular targeting by olfactory sensory axons. J Neurosci. 22:2469–2477.

Vassar R, Chao SK, Sitcheran R, Nuñez JM, Vosshall LB, Axel R. 1994. Topographic organization of sensory projections to the olfactory bulb. Cell. 79:981–991.

Vassar R, Ngai J, Axel R. 1993. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. Cell. 74:309–318.

Wachowiak M, Cohen L. 2001. Representation of odorants by receptor neuron input to the mouse olfactory bulb. Neuron. 32:723–735.

Woo CC, Hingco EE, Johnson BA, Leon M. 2007. Broad activation of the glomerular layer enhances subsequent olfactory responses. Chem Senses. 32:51–55.

Youngentob SL, Johnson BA, M, Sheehe PR, Kent PF. 2006. Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. Behav Neurosci. 120: 1337–1345.