A metric for odorant comparison

Rafi Haddad[1,2], Rehan Khan[1], Yuji K Takahashi[3], Kensaku Mori[3], David Harel[2] & Noam Sobel[1]

In studies of vision and audition, stimuli can be systematically varied by wavelength and frequency, respectively, but there is no equivalent metric for olfaction. Restricted odorant-feature metrics such as number of carbons and functional group do not account for response patterns to odorants varying along other structural dimensions. We generated a multidimensional odor metric, in which each odorant molecule was represented as a vector of 1,664 molecular descriptor values. Revisiting many studies, we found that this metric and a second optimized metric were always better at accounting for neural responses than the specific metric used in each study. These metrics were applicable across studies that differed in the animals studied, the type of olfactory neurons tested, the odorants applied and the recording methods used. We use this new metric to recommend sets of odorants that span the physicochemical space for use in olfaction experiments.

Animals can perceive many structurally diverse odorous molecules[1]. Chemical properties of odorants are represented in spatiotemporal patterns of activity across olfactory sensory neurons in the olfactory epithelium[2], and across glomeruli in the olfactory bulb[3]. Understanding the code of olfaction starts with an understanding of the receptive range of particular olfactory sensory neurons and the response patterns of a set of glomeruli. To this end, one must probe this processing with stimuli that vary along defined axes. But what axes should one use? Molecules can vary along thousands of axes, and no one of these is arguably more privileged than others. Nevertheless, olfaction research has gravitated toward probing a limited set of axes, primarily carbon atom number (CAN) and functional group type and position[4–13]. For example, a given olfactory sensory neuron will generally respond to molecules with a sequential CAN as long as these molecules share the same functional group[5,8,14–18]. Olfactory receptors that respond to odorants with 5 carbons are likely to respond to odorants with 4 or 6 carbons but are less likely to respond to odorants with 7 or more carbons. Similar tuning specificities have been observed in glomeruli[6,7,14,19–23]. In other words, odorants that have similar CAN and functional groups elicit similar glomerular response patterns.

Similarity between response patterns is usually quantified as the Pearson correlation between the two patterns. This quantification of neural response pattern suggested that similarity in CAN is correlated to the similarity in neural response patterns. This suggested correlation applies to neural responses across different species (vertebrates and invertebrates), different measurement methods (Ca^{2+} imaging, intrinsic optical imaging, [^{14}C]2-deoxyglucose, voltage-sensitive dye imaging), different olfactory neurons as well as to perceptual estimates in rodents[24] and humans[25].

Most of the studies linking odorant structure to neural response have been systematic in their choice of odorants, typically selecting saturated straight-chained aldehydes, alkanes, acids and alcohols that differed by only one or two features. When restricted to these molecules, the distance between odorants can be defined as the difference in CAN. But when using more diverse types of odorants with various molecular features, CAN is no longer an informative metric. For example, testing one rat olfactory receptor (the I7 receptor) with 90 odorants of different chemical structures suggested complicated rules linking odorant structure to olfactory sensory neuron response[13], and testing 24 *Drosophila melanogaster* olfactory sensory neurons with 110 odorants of diverse functional groups suggested that CAN alone was a poor predictor of olfactory sensory neuron response[16]. In the rat olfactory bulb, [^{14}C]2-deoxyglucose measurements after exposure to 54 odorants suggested some pattern similarity in response to odorants that share the same CAN but also yielded some unexpected results, implying additional and/or other unresolved factors governing response profiles[21]. *In vivo* extracellular recording of mice olfactory bulb mitral-tufted cells with a large panel of odorants revealed that tuning frequently extended beyond obviously defined chemical categories[26]. Thus, despite some initial efforts made in this respect[26], there remains no simple way to measure the distance between two odorants or to predict whether they will elicit a similar neural response.

To address this, we devised a multidimensional physicochemical metric that took into account not only CAN and functional group but also many other aspects of molecules. We first obtained 1,664 molecular descriptors for each molecule. Molecular descriptors are mathematical values that describe the structure or shape of molecules and can serve to predict the activity and properties thereof. Representing each odorant molecule as a vector of 1,664 descriptor values allowed us to calculate the Euclidean distance between any two odorants, defined as the square root of the sum of squares of the differences between descriptors.

[1]Department of Neurobiology and [2]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Hertzel, Rehovot 76100 Israel. [3]Department of Physiology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. Correspondence should be addressed to R.H. (rhaddad@weizmann.ac.il).

RECEIVED 3 DECEMBER 2007; ACCEPTED 13 MARCH 2008; PUBLISHED ONLINE 30 MARCH 2008; DOI:10.1038/NMETH.1197

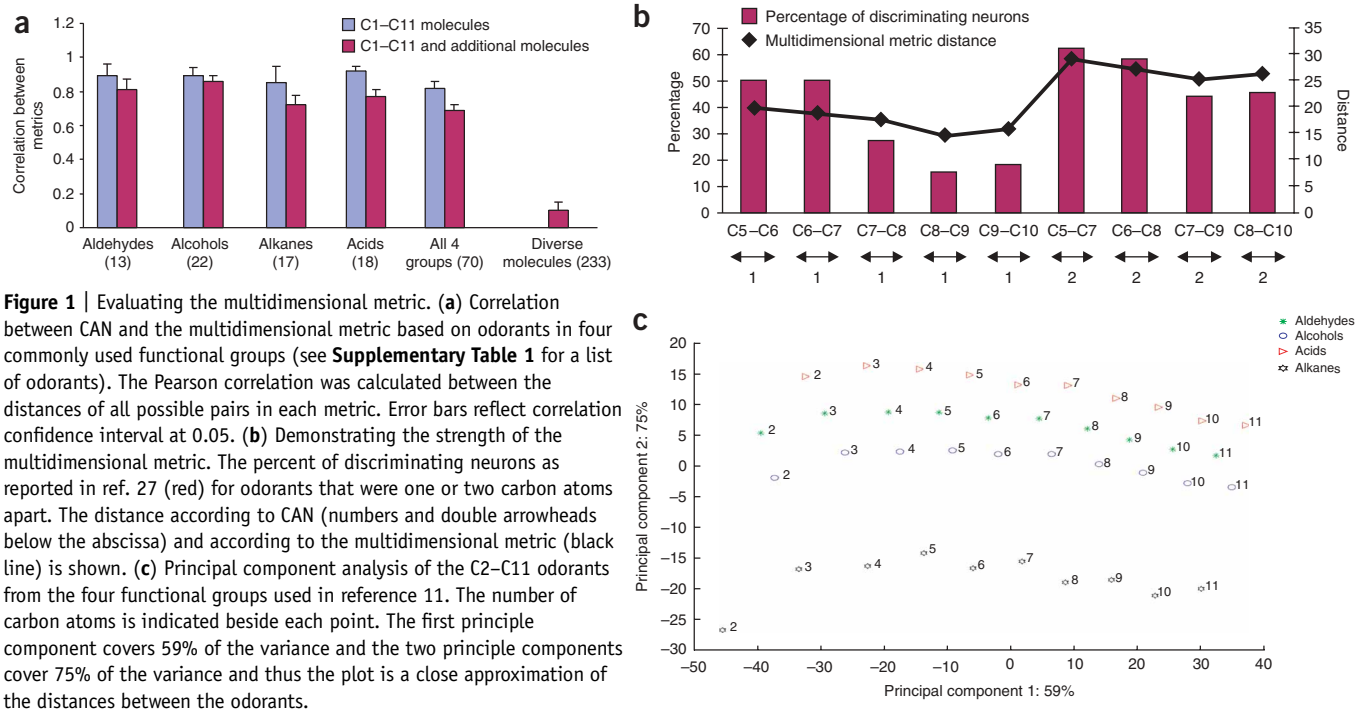

Figure 1 | Evaluating the multidimensional metric. (**a**) Correlation between CAN and the multidimensional metric based on odorants in four commonly used functional groups (see **Supplementary Table 1** for a list of odorants). The Pearson correlation was calculated between the distances of all possible pairs in each metric. Error bars reflect correlation confidence interval at 0.05. (**b**) Demonstrating the strength of the multidimensional metric. The percent of discriminating neurons as reported in ref. 27 (red) for odorants that were one or two carbon atoms apart. The distance according to CAN (numbers and double arrowheads below the abscissa) and according to the multidimensional metric (black line) is shown. (**c**) Principal component analysis of the C2–C11 odorants from the four functional groups used in reference 11. The number of carbon atoms is indicated beside each point. The first principle component covers 59% of the variance and the two principle components cover 75% of the variance and thus the plot is a close approximation of the distances between the odorants.

Using this metric, as well as a second optimized metric, we revisited 9 previously published datasets and analyzed a new dataset, for which we knew the odorants used but did not know the neural response. We found that the new metric was always better at accounting for neural responses than the specific metric used in each study. Moreover, this single metric was applicable across studies that used different olfactory neurons, different model systems, different neuronal response measurement techniques and odorants varying along different feature types.

RESULTS

Comparing the multidimensional metric to the CAN metric

To assess the relationship between the CAN and the proposed metric, we first restricted the odorants to vary only in CAN and fixed all other features such as functional group (by considering only straight-chained aliphatic compounds from a specific chemical family; **Supplementary Table 1** online). We calculated the distances between all odorant pairs in each group according to the two metrics and computed the correlation (**Fig. 1a**). The multidimensional metric strongly correlated to the CAN metric as long as we restricted the calculation to straight-chained molecules from the same functional group ($r > 0.83$, $P < 10^{-15}$, all groups). However, when we added to each group odorants that have the same functional group but differ in other features, such as number of double bonds or methyl groups (**Supplementary Table 1**), the correlation diminished, and it diminished even further when we compared molecules from all four functional groups ($r = 0.69$, $P < 10^{-18}$). When we examined the relation between the two metrics on a large set of odorant molecules with diverse structures (**Fig. 1a**), the correlation diminished to 0.105 ($P < 10^{-67}$). In other words, the multidimensional metric and the CAN metric were very similar when restricted to odorants that differed only in the CAN dimension but differed profoundly when the set of odorants was taken from diverse molecular structures.

Advantages of the multidimensional metric over the CAN metric

The multidimensional metric provides valuable information even when restricted to odorants that differ only in the CAN. This can be illustrated in an analysis of data presented in reference 27. This study reported the response of rat olfactory sensory neurons, as measured using Ca^{2+} imaging, in response to a set of aliphatic aldehydes with 5–10 carbons (C5–C10)[27]. We compared the percentage of neurons that discriminate between pairs of odorants that are either one or two carbon atoms apart (as reported in ref. 27) to the distance between the odorants calculated using CAN or the multidimensional metric (**Fig. 1b**). Whereas the CAN metric reports that all odorants were one or two units apart, the multidimensional metric reports different distances. For example, whereas the CAN reports equidistance between C7–C8 and C8–C9, according to the multidimensional metric C8 was closer to C9 than to C7 ($\sim$17.7 distance units for C7 to C8, and $\sim$14.5 distance units for C8 to C9). Notably, this better predicted the measured fraction of discriminating neurons between C7–C8 and between C8–C9 ($\sim$27% and $\sim$15%, respectively, as reported in ref. 27). Examining the relationship between the multidimensional metric and the percentage of discriminating neurons, we found a very high correlation ($r = 0.94$, $P = 0.01$ for the group with one carbon atom difference; $r = 0.92$, $P = 0.07$ for the group with two carbon atoms difference; and $r = 0.84$, $P = 0.004$ for all differences). In contrast to the above obtained correlation value of 0.84, running the same analysis using the CAN metric we obtained a correlation value of 0.63 ($P = 0.066$), a result significantly poorer than that obtained with the multidimensional metric (*t*-test for correlated samples, $t = 4.9$, d.f. $= 6$, $P = 0.002$). In other words, the suggested multidimensional metric had wider applicability and was more accurate than the CAN metric, as defined by predicting the measured neural response, even in the case of odorants from a single chemical family.

In a study measuring the response of rabbit mitral-tufted cells using 32 odorants from four different functional groups, the

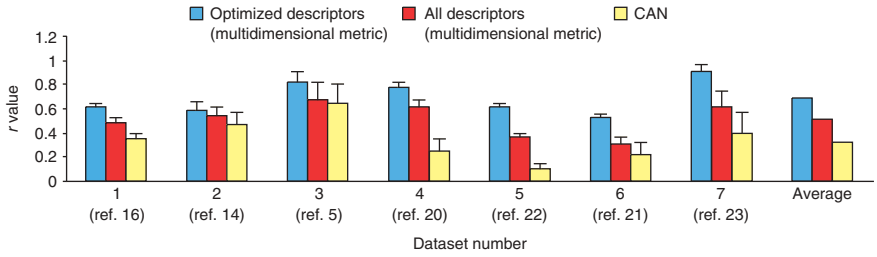

Figure 2 | Correlation between neuronal response pattern similarity and odorant distances calculated using three different metrics across 7 datasets. The average *r* value of the metric using the optimized metric was 0.69 ($P < 10^{-10}$ for all groups). Error bars reflect correlation confidence interval at 0.05.

authors pointed out the importance of both CAN and functional group[11]. After obtaining multidimensional metric values for these 32 odorants, we used principal component analysis to depict the odorants in two dimensions. Odorants that have the same functional group aligned in an almost horizontal line, but odorants that have the same CAN were roughly on the same slightly skewed vertical line (**Fig. 1c**). Thus, the major axes for these 32 odorants (the first two principal components) were CAN and functional group type. This is unsurprising, because the 32 odorants that had been selected in reference 11 varied on these two axes. However, the multidimensional metric ordered acids as closer to aldehydes than to alcohols and as farthest from alkanes. These ratios were in accordance with the experimental finding in reference 11, namely that acids and aldehydes had in common 32 responding neurons, whereas acids and alcohols had only 2 responding neurons in common, and acids and alkanes had no responding neurons in common[11].

Another example of the power of the multidimensional metric is that according to the multidimensional metric, differences in CAN have higher impact on odorant distances at low CAN values (**Fig. 1a**,**b**). Consistent with this, honeybees better generalized between pairs of long aliphatic chain odorants than between pairs of short aliphatic chained odorants[12]. We demonstrated the strength of the multidimensional metric on two additional datasets (**Supplementary Fig. 1** online).

Predicting neural response similarity

We determined whether the predictions made by our metric correlate with odorant-induced responses in datasets obtained under different experimental conditions (**Fig. 2**). Because the distance between odorants in the neuronal space is estimated using the Pearson correlation, we restricted our analysis to datasets that reported the response of large sets of neurons (that is, at least 20 neurons to at least 10 odorants). We analyzed 7 studies[5,14,16,20–23] (summarized in **Table 1**). In all cases except for dataset 3 (ref. 5), the multidimensional metric was a significantly better predictor of neural activity than CAN ($t > 2.8$, d.f. > 88, $P < 0.01$ for the 6 datasets). This improvement was evident across datasets that differed in the animal used, the method of neural activity measurement and the type of neurons measured. In dataset 3, there was no substantial improvement in the correlation when using the multidimensional metric ($r = 0.67$ for the multidimensional metric and $r = 0.64$ for the CAN metric; $P < 10^{-9}$ for both). This is because the odorants used in this experiment were all straight-chained and from only two functional groups, conditions under which the two metrics were almost identical (**Fig. 1a**).

Optimizing the multidimensional metric

Many of the 1,664 physicochemical descriptors were intercorrelated and were thus redundant. Assuming that a better metric should be able to explain more of the variance, we optimized our metric by searching for the best descriptor subset that improved our total correlation values across the 7 datasets. We obtained 32 descriptors (**Supplementary Table 2** online). We calculated correlation values between the distances predicted by the optimized metric and odorant-induced responses in the 7 studied datasets (**Fig. 2** and **Supplementary Fig. 2** online). The *r* values obtained using this optimized set (average $r = 0.69$) were substantially higher than those obtained using the full 1,664 set.

Because we obtained the optimized descriptor subset using all 7 datasets, it may be over-fitted to these datasets. To test this, we conducted a leave-one-out learning scheme, whereby we calculated the correlation values for one dataset while using the other six to obtain the optimized descriptor set (**Supplementary Fig. 3** online). The total correlation value when using the optimized descriptor set was higher than the correlation value when using all descriptors, and the improvement in the correlation values was significant in four of the datasets (*t*-test for correlated samples, $P < 0.01$ for all 4 datasets). Dataset 2 (ref. 14) was different from the others in that here the optimized metric best predicted only the absolute value of the correlation. In other words, odorants deemed close according to the multidimensional metric elicited either very similar responses (strong positive correlation) or very different responses (strong negative correlation; **Supplementary Fig. 3b**). Although we

Table 1 | Experimental conditions and methods in the 7 studies analyzed

Dataset number	Reference number	Animal used	Neurons measured	Measurement method	Number of odorants	Number of odorants removed	Number of neurons
1	16	Fruit fly	Olfactory sensory neurons	Extracellular single unit	110	63	24
2	14	Rat	Glomeruli	Intrinsic signal	36	7	30
3	5	Mouse	Olfactory sensory neurons	Ca^{2+} imaging	14	1	30
4	20	Honeybee	Glomeruli	Ca^{2+} imaging	36	0	37
5	22	Rat	Glomeruli	Intrinsic signal	72	9	115
6	21	Rat	Glomeruli	[^{14}C]2-deoxyglucose	44	0	Entire bulb
7	23	Tadpole	Glomeruli	Ca^{2+} imaging	15 amino acids	0	67

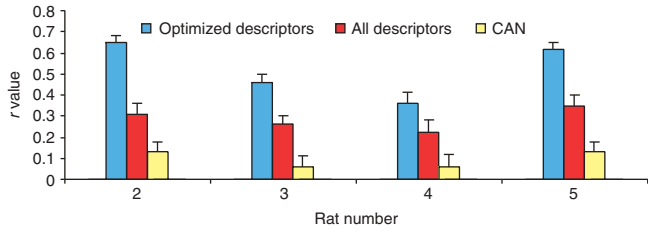

Figure 3 | Testing the predictive power of the multidimensional metric using new data. The datasets represent data collected from 4 individual rats (marked as rats number 2–5). Error bars reflect correlation confidence interval at 0.05.

do not know what underlies this unique response pattern, it may be related to the selection criteria applied by the authors, who reported data for 30 olfactory neurons out of 90 measured.

Optimizing using only datasets that share a particular attribute, such as all datasets that report glomerulus data or datasets that use similar neural response measurement methods (for example, Ca^{2+} imaging versus intrinsic optical imaging), did not substantially change the results. For example, when optimizing using the 5 datasets reporting glomerulus responses, the average correlation value for the datasets reporting olfactory receptor neuron response was 0.61, which was the same as when using the all-descriptors metric. We obtained similar values when we optimized the metric using datasets reporting only olfactory receptor neuron responses and tested the prediction on the datasets reporting glomerulus responses. Other parameters had similar results (taking into account the number of datasets used for the learning process; **Supplementary Fig. 4** online).

Blind test of the multidimensional metric

We tested the multidimensional metrics against data that were unavailable when we were developing the computational rule. We obtained a list of odorants (**Supplementary Table 3** online) for which glomerulus responses have been measured with optical imaging in rats (K.M.; unpublished data). We then calculated the distances between all pairs of odorants using the suggested multidimensional metric and the optimized metric, based on which we could predict differences in response (see **Supplementary Data 1** online for the predicted values using the optimized metric). Then we tested the correlation between the metric-predicted distances and the measured response pattern in 4 rats (**Fig. 3** and **Supplementary Table 3**). Whereas the CAN metric did not predict response (average $r = 0.12$), the optimized multidimensional metric average correlation value was 0.5 ($P < 10^{-30}$ for all four experiments). Furthermore, the optimized metric was significantly better than the non optimized metric (d.f. > 987, $t > 5.4$, $P < 0.0001$; for all 4 experiments), thus validating that the optimization was universal and not restricted to the data from which it was developed. In sum, our metric served to predict differences in neuronal responses that were unknown to us at the time of prediction.

DISCUSSION

Using the multidimensional physicochemical metric that we developed, we found that when working with straight-chain odorants from a few functional groups, the first axis of the resultant olfactory space correlated to the number of carbon atoms and the second axis correlated to the functional group. Thus, previously suggested axes are special cases of our multidimensional metric, yet unlike these axes the proposed metric imposed distances between odorants across different functional groups, and these distances accurately predicted neural response. Furthermore, aliphatic odorants with similar high CAN values were closer to each other, according to the multidimensional metric, than aliphatic odorants with similar low CAN values. This is consistent with previous results[11,27] and supports observations made by others[12,14].

Although our metric was applicable across methods of recording, across species, and across levels within the olfactory system (olfactory sensory neurons and glomeruli), this nevertheless does not imply that olfactory coding is identical across these domains, or that different methods record the same underlying mechanisms. For example, different animals are probably better tuned to different portions of olfactory space. This will modify their acuity within portions of this space, but will not reorder the space. The same is true across levels of processing and methods of recording. Thus, given three odorants using our metric, two that are very close to each other (*a* and *b*) and one that is distant (*c*), our metric cannot predict whether a given animal or detection method will respond to any of these odorants, or whether a given odorant will be more or less salient (odor *a* may be overwhelming for one animal and barely detectable for another). What our method can predict, however, is that if a response of a particular pattern was elicited by odorant *a*, then there is a higher chance for odorant *b* to elicit a similar response than there is for odorant *c* to elicit a similar response.

The proposed metric has a few notable weaknesses. For one, it explains only about 50% of the variance, and a considerable portion of the variance in response remains unexplained. A consequence of this is that our metric is ill-suited to dealing with sparse data, that is, cases with only few olfactory sensory neurons or glomeruli. For example, in datasets reporting response patterns of 14 neurons or less, the correlation between odorant distance and pattern similarity in odorant-induced olfactory sensory neuron and glomerulus activity was not significant[8,19,28] ($r < 0.3$, $P > 0.05$ for all datasets). This is probably because the power of the metric is not sufficient to overcome the inherently poor power of comparing to only a small number of response patterns. In the 7 published datasets we analyzed, the number of neurons ranged from 24 to 115.

Finally, the suggested metric provides a practical solution to odorant selection, a critical component in olfaction experiments. To select odorants, one needs to decide what part of the olfactory space one wishes to cover and what the covering resolution is. Once these parameters have been decided upon, one can use the multidimensional metric to decide which odorants are appropriate. Although the calculations we have used and described can be implemented with ease, to increase the practical contribution of this manuscript we provide an interactive tool (**Supplementary Data 2** online) and also suggest several possible odorant selection groups (**Supplementary Data 3** online).

METHODS

Data sets. We mined data from published papers[5,11,14,16,20–23,27]. In almost all cases, the response values had been reported as the response strength category of each olfactory sensory neuron or glomerulus. In dataset 6 (ref. 21), the raw data had not been reported; instead, the correlation values between response activity

maps had been reported in a grayscale image. To extract these correlation values, we analyzed this image and extracted the gray values and then estimated the r values from it. This is a noisy process and might explain why this dataset received the lowest correlation values.

Odorants. Some odorants did not elicit any response, or elicited a very weak one. Calculating correlations required that each variable have a nonzero (or close to zero) variance. We therefore removed from each dataset all the odorants that had less than 4 responding neurons. This amounted to removal of odorants from the datasets we analyzed (**Table 1**). The large number of non-responding odorants in dataset 1 is probably due to the relatively low number of neurons measured.

Descriptors. We generated physicochemical descriptors by obtaining the molecular structure for each odorant from PubChem (http://pubchem.ncbi.nlm.nih.gov/search/) and inputting this into Dragon (http://www.talete.mi.it/download.htm), a commonly used and well-described[29] software that provides more than 1,600 molecular descriptors. We normalized the descriptors.

Calculating correlations. For each 2 odorants out of the $n(n-1)/2$ possible pairs, we calculated the Pearson correlation between their response pattern and the physicochemical distance using the Euclidean distance. We then calculated the correlation between the above 2 sets of $n(n-1)/2$ numbers.

Searching for the best descriptors subset. Searching for the smallest subset of descriptors that maximize the total correlation value was an intractable task. Thus, we used approximation greedy algorithms (**Supplementary Methods**).

Additional methods. Details of the search algorithm, additional examples of the strength of the metric and further analysis, with tools for calculating distances between odorants and recommendations on sets of odorants that are best in terms of spanning the physicochemical space, are available in the **Supplementary Methods**.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank A. Elite.

AUTHOR CONTRIBUTIONS

R.H., R.K., D.H. and N.S. are authors of the concept. R.H. performed the analysis; R.H., N.S. and D.H. wrote the manuscript; Y.K.T. and K.M. provided data post-hoc for the blind test.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211–218 (2001).
2. Kent, P.F. & Mozell, M. The recording of odorant-induced mucosal activity patterns with a voltage-sensitive dye. *J. Neurophysiol.* **68**, 1804–1819 (1992).
3. Johnson, B.A., Woo, C.C. & Leon, M. Spatial coding of odorant features in the glomerular layer of the rat olfactory bulb. *J. Comp. Neurol.* **393**, 457–471 (1998).
4. Zhao, H. *et al.* Functional expression of a mammalian odorant receptor. *Science* **279**, 237–242 (1998).
5. Sato, T., Hirono, J., Tonoike, M. & Takebayashi, M. Tuning specificities to aliphatic odorants in mouse olfactory receptor neurons and their local distribution. *J. Neurophysiol.* **72**, 2980–2989 (1994).
6. Rubin, B.D. & Katz, L.C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
7. Meister, M. & Bonhoeffer, T. Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.* **21**, 1351–1360 (2001).
8. Malnic, B., Hirono, J., Sato, T. & Buck, L.B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
9. Laska, M. & Teubner, P. Olfactory discrimination ability for homologous series of aliphatic alcohols and aldehydes. *Chem. Senses* **24**, 263–270 (1999).
10. Johnson, B.A., Woo, C.C., Hingco, E.E., Pham, K.L. & Leon, M. Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. *J. Comp. Neurol.* **409**, 529–548 (1999).
11. Imamura, K., Mataga, N. & Mori, K. Coding of odor molecules by mitral/tufted cells in rabbit olfactory bulb. I. Aliphatic compounds. *J. Neurophysiol.* **68**, 1986–2002 (1992).
12. Guerrieri, F., Schubert, M., Sandoz, J.C. & Giurfa, M. Perceptual and neural olfactory similarity in honeybees. *PLoS Biol.* **3**, e60 (2005).
13. Araneda, R.C., Kini, A.D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nat. Neurosci.* **3**, 1248–1255 (2000).
14. Uchida, N., Takahashi, Y.K., Tanifuji, M. & Mori, K. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nat. Neurosci.* **3**, 1035–1043 (2000).
15. Hallem, E.A. & Carlson, J.R. The spatial code for odors is changed by conditioning. *Neuron* **42**, 359–361 (2004).
16. Hallem, E.A. & Carlson, J.R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
17. de Bruyne, M., Foster, K. & Carlson, J.R. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).
18. de Bruyne, M., Clyne, P.J. & Carlson, J.R. Odor coding in a model olfactory organ: the *Drosophila* maxillary palp. *J. Neurosci.* **19**, 4520–4532 (1999).
19. Wachowiak, M. & Cohen, L.B. Representation of odorants by receptor neuron input to the mouse olfactory bulb. *Neuron* **32**, 723–735 (2001).
20. Sachse, S., Rappert, A. & Galizia, C.G. The spatial representation of chemical structures in the antennal lobe of honeybees: steps towards the olfactory code. *Eur. J. Neurosci.* **11**, 3970–3982 (1999).
21. Johnson, B.A. *et al.* Functional mapping of the rat olfactory bulb using diverse odorants reveals modular responses to functional groups and hydrocarbon structural features. *J. Comp. Neurol.* **449**, 180–194 (2002).
22. Takahashi, Y.K., Kurosaki, M., Hirono, S. & Mori, K. Topographic representation of odorant molecular features in the rat olfactory bulb. *J. Neurophysiol.* **92**, 2413–2427 (2004).
23. Manzini, I., Brase, C., Chen, T. & Schild, D. Response profiles to amino acid odorants of olfactory glomeruli in larval *Xenopus laevis*. *J. Physiol. (Lond.)* **581**, 567–579 (2007).
24. Cleland, T., Morse, A., Yue, E. & Linster, C. Behavioral models of odor similarity. *Behav. Neurosci.* **116**, 222–231 (2002).
25. Laska, M., Galizia, C.G., Giurfa, M. & Menzel, R. Olfactory discrimination ability and odor structure-activity relationships in honeybees. *Chem. Senses* **24**, 429–438 (1999).
26. Davison, I.G. & Katz, L.C. Sparse and selective odor coding by mitral/tufted neurons in the main olfactory bulb. *J. Neurosci.* **27**, 2091–2101 (2007).
27. Kaluza, J.F. & Breer, H. Responsiveness of olfactory neurons to distinct aliphatic aldehydes. *J. Exp. Biol.* **203**, 927–933 (2000).
28. Friedrich, R.W. & Korsching, S.I. Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron* **18**, 737–752 (1997).
29. Tetko, I.V. *et al.* Virtual computational chemistry laboratory–design and description. *J. Comput. Aided Mol. Des.* **19**, 453–463 (2005).