bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# Modular structure of human olfactory receptor codes reflects the bases of odor perception

**Ji Hyun Bak**[1], **Seogjoo J. Jang**[2,3], **and Changbong Hyeon**[1,*]

[1]Korea Institute for Advanced Study, Seoul 02455, Korea; [2]Department of Chemistry and Biochemistry, Queens College, City University of New York, New York 11367, USA; [3]PhD programs in Chemistry and Physics, and Initiative for Theoretical Sciences, Graduate Center, City University of New York, New York 10016, USA; *Correspondence: hyeoncb@kias.re.kr

(Dated: January 19, 2019)

**The circuits of olfactory signaling are reminiscent of complex computational devices. The olfactory receptor code, which represents the responses of receptors elicited by olfactory stimuli, is effectively an input code for the neural computation of odor sensing. Here, analyzing a recent dataset of the responses of human olfactory receptors (ORs) to odorants, we show that the space of human olfactory receptor codes is partitioned into a modular structure where groups of receptors are "labeled" for key olfactory features. The existence of such independent and sizable receptor groups implies a significant dimensional reduction in the space of human odor perception. Our data-driven statistical analysis of receptor codes leads to a valuable discovery that human olfaction works by hybridizing both the combinatorial coding and labeled line strategies, even at the early stage of signal processing.**

## INTRODUCTION

Olfaction is a sensory process that captures environmental signal by detecting the molecular stimuli to a repertoire of olfactory receptors (ORs). Although the full processing of olfactory information is realized in the complicated neural circuits [1], the first step of olfactory sensing involves a selective binding of odorants to the cognate ORs, which is biochemical in nature [2]. The binding elicits an array of downstream responses in the corresponding olfactory receptor neurons (ORNs) [3]. The response pattern encoded into ORs or ORNs, termed the olfactory *receptor code* [4], provides the first neural representation of an odor and is essential in the early stage of information-processing in olfactory sensing.

Lying at a crucial juncture in the flow of olfactory information, the receptor code links between the molecular and neural spaces. An olfactory stimulus in the *molecular space*, representing the physiochemical properties of the odorants, is translated into the *receptor code space* to form an "input code". The information is then processed through the higher-order *neural spaces*, and eventually evokes the sense of smell that constitutes the *perceptual odor space* (see Fig. 1). Recent insights into the processing of olfactory information were gained mostly based on the olfactory sensing of non-human species [5–7] or through theoretical studies [8, 9]. A separate branch of research attempts to relate the molecular space directly with the perceptual odor space [10, 11]. However, the goal of understanding the principles of olfactory computation in humans faces a fundamental difficulty due to our limited access to the neural circuitry in the human brain.

Here, we show that a critical progress can still be made by a systematic analysis of the receptor code space. Whereas the OR repertoire is known to encode the input signal from odorants like a piano keyboard that combinatorially produces a variety of musical chords [4], it may be further "formatted" to facilitate the processing of relevant information, perhaps by appropriately reflecting the structure of the perceptual odor space. Specifically, we analyze a set of receptor codes for different odorants, extracted from the measurement of downstream biochemical responses of the human ORs [12–14]. Knowing that the receptor code is implemented by the many-to-many pairwise interaction between odorants and the ORs [4, 15], we treat the interaction of each odorant-receptor pair as either "on" (responding) or "off" (not responding), and focus on the binary pattern of pairwise interaction as a zeroth-order representation of the receptor code. We reveal the intrinsic structure of the human receptor code space through an analysis of the similarity patterns, which involves quantifying the extent of overlap between distinct receptor codes (receptor code redundancy).

## RESULTS

We present a quantitative analysis of the pattern of odorant-receptor interactions, employing the dataset that report the responses of 303 receptors against 89 odorants [14]. The state of all $N = 303$ receptors can be represented in terms of an $N$-dimensional binary vector $\mathbf{y}$, where $y_i = 1$ if the $i$-th receptor is "on", and $y_i = 0$ if it is "off". For a given odor $\mathbf{x}$, the corresponding *receptor code* $\mathbf{y}$ is the representation of the odor $\mathbf{x}$ in the $N$-dimensional binary receptor

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

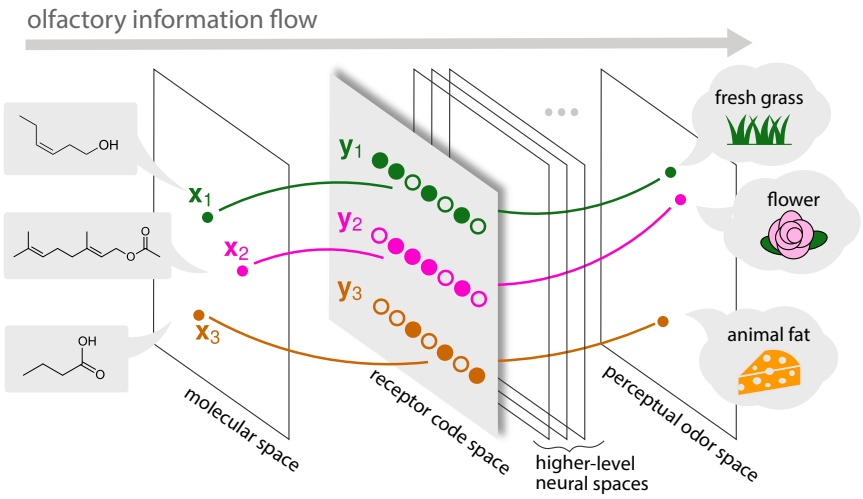


**Figure 1**: **Schematic of olfactory information flow.** The *molecular space*, representing physiochemical features of odorant molecules, is first recognized by OR repertoire and encoded into the *receptor code space*. The information stored in the receptor codes is modified while it goes through the higher-level *neural spaces* and finally projected on the *perceptual odor space*, and sensed as a smell.

code space.

We collected all 535 pairwise interactions reported in the dataset, including 60 de-activations [16], and visualized the receptor codes in the form of an interaction network (Fig. 2a). In this interaction network, each node is either an odorant or a receptor, and each edge connects an interacting odorant-receptor pair For the interaction network with a richer display with varying parameters, see Fig. S1.

## Interaction network reveals odorant hubs with non-redundant receptor codes

In the odorant-OR interaction network (Fig. 2a), the number of edges attached to an odorant node indicates the number of receptors that recognize this odorant. We call this number the *degree* of the odorant node, and denote it by $k_\lambda$, where $\lambda$ is the index for the odorant (also see Methods).

We observe two properties from the statistics of single-odorant degrees. First, the receptor-space representation of single odorants is sparse overall: On average, an odorant is recognized by only 6 ($\langle k \rangle \simeq 6$) out of $N = 303$ receptors, which amounts to 2% of the receptor space. The sparsity observed in the network is consistent with previous reports from the ORN responses [17, 18]. Second, the degrees of the odorants in the dataset are non-uniformly distributed with a heavy tail, which can be reasonably fitted to a power law $P(k) \sim k^{-\alpha}$ with $\alpha \approx 0.9$ (Fig. 2b). Such distribution allows us to identify a small number of high-degree odorant "hubs". The six highest-degree hub odorants are labeled in Fig. 2a. It is worth noting that the receptor codes for the hub odorants are highly non-redundant with one another (Fig. 2c); among the six highest-degree odorants, 10 out of all 15 pairs have completely disjoint sets of receptor codes. In particular, despite significant similarity in chemical structures, there is no co-activated OR between eugenol and eugenol acetate, which differ in a single functional group (hydroxy group versus acetate in the aromatic ring). The overlap of receptor codes between the hub odorants is significantly small compared to that between randomly sampled odorants from the dataset (Fig. S3).

We introduce an idea of *receptor code redundancy*, a quantitative measure of (dis-)similarity between the olfactory responses of two distinct odorants. If two odors have exactly the same receptor codes, they are perceived as the same olfactory signal. We hypothesize that two odors are poorly distinguishable when their receptor codes have a higher redundancy. For example, consider a discriminatory task where the goal is to detect a target odorant in the presence of a constant background odor [19, 20]. We define the receptor code redundancy in terms of the fraction of receptors that respond to the target odorant (see Fig. 2d-e). If this fraction is small, as in Fig. 2d, the target odorant is deemed unambiguously detected even in the presence of the background odorant.

Taken together, the single-odorant representation in the receptor space is sparse and non-uniform, giving rise to high-degree "hub" odorants. The receptor code is almost non-redundant among these hub odorants, whereas the odorants that have a greater receptor code redundancy with a hub odorant are grouped around it. As presented below, analysis of the receptor code redundancy enables us to decipher the structure of the receptor code space more quantitatively.

## The receptor code space is naturally partitioned to show a modular structure

The grouped structure is better manifested when the interaction network is projected to the space of receptors. Here we consider the *co-activation graph* of receptors, which inherits all receptor nodes in the original interaction network; two receptor nodes are connected by an edge if they share a common odorant in the original network (Fig. 2d-e, last column). Of particular interest in the co-activation graph are the "receptor *cliques*", or the groups of receptors that are co-activated by the same odorants. Because each pair of receptors interacting with a shared odorant $\lambda$ is connected in the

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

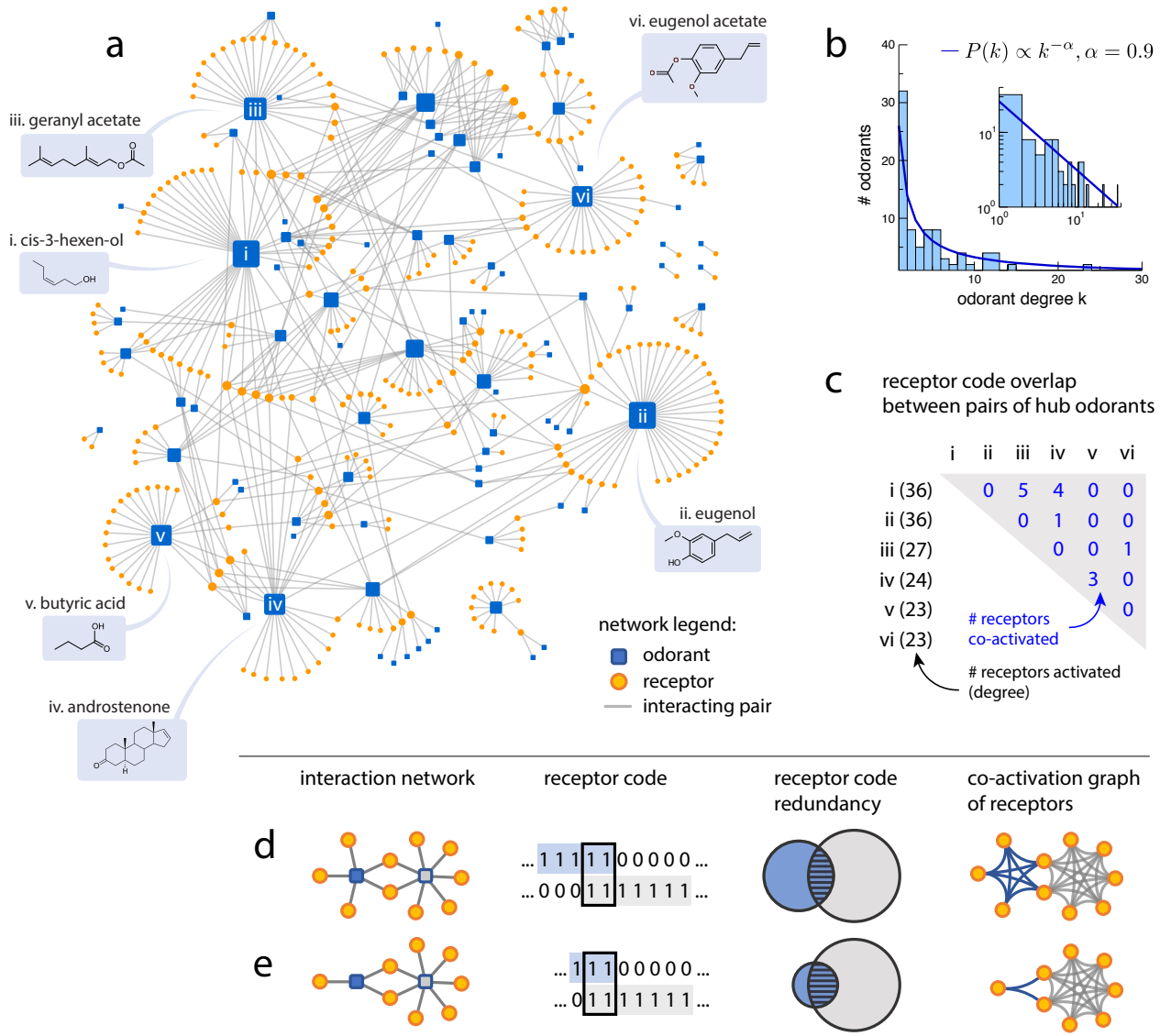


**Figure 2**: **Odorant-receptor interaction network and receptor codes. (a)** The odorant-receptor interaction network visualizes all the interactions between 89 odorants (blue squares) and 303 receptors (orange circles) from the dataset [14]. An edge is drawn when odorant and receptor is interacting. Node size reflects the degree; six highest-degree odorants are marked. Also see Fig. S1 for an extended display with all odorant labels, and with edge attributes for pairwise interaction properties. **(b)** The distribution of odorant degrees as a histogram. The degree distribution is fitted to $P(k) \sim k^{-0.9}$, and has an average value $\langle k \rangle \simeq 6$ (also see Fig. S2). **(c)** The overlaps between the receptors codes for the six hub odorants identified in **a** (also see Fig. S3). **(d-e)** Illustration of the idea of receptor code redundancy. The first two columns show two equivalent representations of the receptor code space, as an interaction network and as a set of binary vectors. We consider the receptor code redundancy of a target odorant (blue) with respect to a background odorant (gray). In the third column, the receptor code redundancy is illustrated with Venn diagrams of interacting receptors. The redundancy of adding the target odorant in the presence of the background odorant is represented by the relative size of overlap (striped area) with respect to the total size of target signal (blue-shaded area). Although the number of overlapping receptors is the same in both cases, the normalized redundancy is smaller in **d** (two out of five) than in **e** (two out of three). In the last column, the interaction network is projected to a co-activation graph of receptors.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

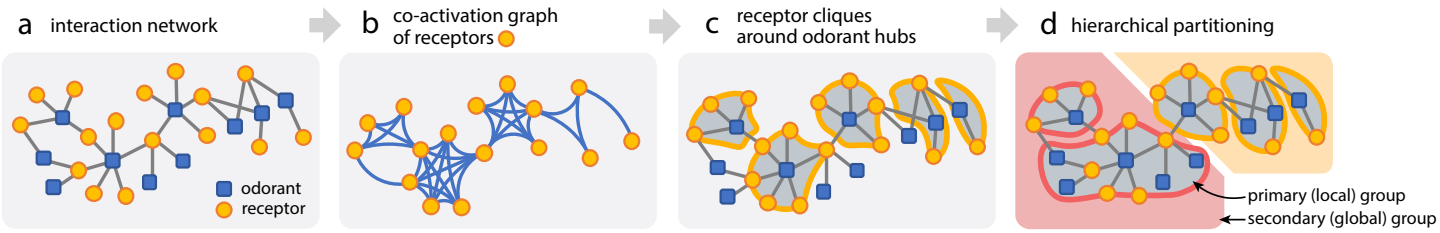


**Figure 3**: **Schematics for the partitioning of receptor code.** **(a)** Given an interaction network with particular statistics of the hub odorants, **(b)** the modular structure is revealed more clearly in the co-activation pattern of receptors. **(c)** We characterize the local structure from the receptor groups around the odorant hubs, and **(d)** merge the local groups to obtain the best partitioning of the receptor code. Also see Fig. S4 for a more detailed illustration.

co-activation graph of receptors, the set of all receptors that interact with a given odorant always form a clique. Given the particular structure of the interaction network, with hub odorants with largely non-redundant receptor codes (Fig. 3a), the co-activation graph of receptors is bound to have large and mostly non-overlapping cliques that are associated to the hub odorants (Fig. 3b). We use the receptor cliques to *partition* the receptor space into non-overlapping groups, such that each receptor group is associated to a shared odorant; when a receptor is a part of more than one receptor clique, it is assigned to the larger clique (Fig. 3c).

The receptor groups can be used to group odorants based on their receptor codes. The idea is to construct odorant groups such that an odorant $\lambda$ belongs to an odorant group $\Lambda$ if its receptor code redundancy with respect to the group $\Lambda$, $\chi_{\lambda|\Lambda} = |\mathbf{y}_\lambda \cap \mathbf{y}_\Lambda| \,/\, |\mathbf{y}_\lambda|$, is large (see Methods). We first used the receptor groups to determine the reference receptor codes for odorant groups ($\mathbf{y}_\Lambda$'s) and then assigned each odorant to the group $\Lambda$ where $\chi_{\lambda|\Lambda}$ is maximized. This results in a simultaneous partitioning of odorants and receptors (Fig. 3d, inner groups).

So far, we have demonstrated that groups emerge naturally from the particular statistics of the receptor codes. The resulting groups give us information as to which pairs of receptors (or odorants) are most strongly correlated. But in order to characterize the overall structure of the entire receptor code space, the "positive" correlations alone are not enough; for a perfect classification, receptors in the same group must respond to the same sets of odorants, and receptors in different groups to different odorants (see Methods for details). Thus, we step further to obtain an optimal classification of the co-activation patterns in the receptor code space. By appropriately *merging* the receptor groups obtained from the primary grouping, we carry out a secondary grouping of the receptors, where the threshold for pairwise merging is determined to maximize the goodness of classification (Fig. S5). Because we partition the odorants and the receptors simultaneously, a merging of receptor groups automatically leads to a merging of the corresponding odorant groups (Fig. 3d).

Application of the grouping procedure to the human olfactory receptor codes results in the sorted interaction matrix in Fig. 4, where each row represents the receptor code for a given odorant. The rows and columns of the interaction matrix are sorted according to the orders in the respective odorant/receptor groups (Fig. 4a-b). The interaction matrix has a roughly block-diagonal form (Fig. 4c), with much less significant contributions from the off-diagonals elements. Notice that we have already arranged the interaction network layout to represent a grouped structure, which is visualized more clearly in Fig. 5, where colored territories represent the best partitioning of the receptor code space. The feasibility of grouping reflects a special structure of the receptor code; such clear grouping is not obtained from a random network with the same degree statistics (Fig. S6).

## Odorants in the same group tend to carry similar olfactory features

Whereas our analysis is only based on the receptor codes and their redundancy, it results in a grouping of perceptually similar odorants together. For example, the hub odorants in the largest global group (Group 1 in Fig. 4) include "plant-related" odors, such as cis-3-hexen-1-ol (the characteristic green/grassy odor); geranyl acetate, (floral); cinnamaldehyde, eugenol methyl ether (plant-derived spices). The smaller-degree odorant members in the group include phenyl acetaldehyde (green or honey-like); sandalwood (woody); coumarin (sweet, grassy, hay-like); linalool, lilial, lyral, anisaldehyde, terpineol (floral); n-amyl acetate (fruity); methyl salicylate, spearmint, cyclohexanone, carvones (minty).

In Group 2, we find more "animal-related" odors, such as androstenone, androstadienone (smell of body) and butyric acid (sweaty smell, also occurring in human body). Other members of this group includes odorants that are often associated with "bodily" or "sensual" feelings: ambrette (fragrance similar to animal musk) and jasmine. There are also several odorants that smell floral or fruity; in fact, many receptors are shared between the first (plant-related) and the second group (body-related).

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

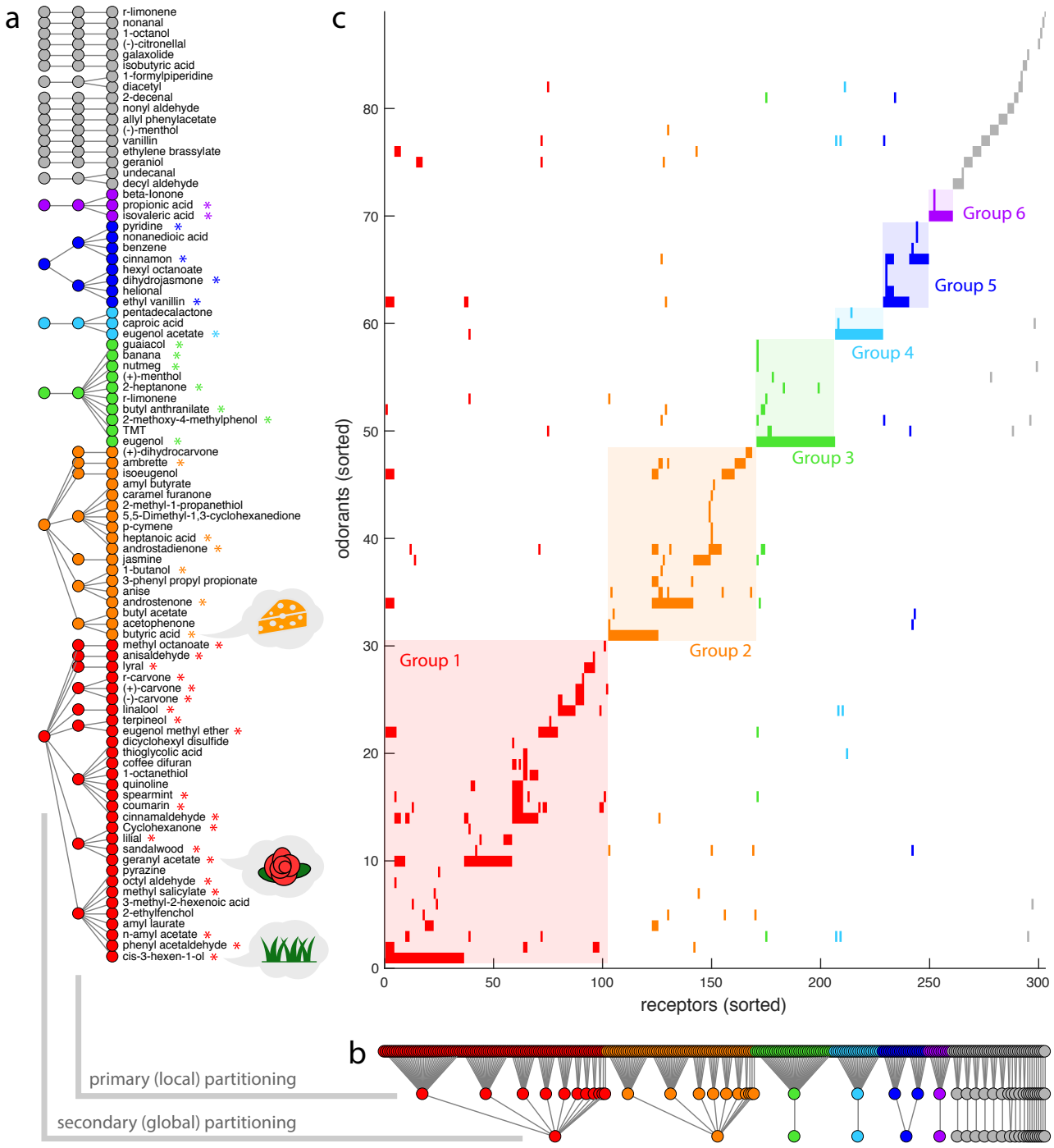


**Figure 4**: **Simultaneous partitioning of odorant and receptors, based on the co-activation pattern and the receptor code redundancy.** The primary and secondary groups are shown for **(a)** the odorants and **(b)** the receptors, where the indices are simultaneously sorted by the group rank. Also see Fig. S5 and Fig. S6 for validations. **(c)** The rows and columns of interaction matrix is consequently sorted by the odorant and receptor indices. Each row represents the receptor code for the corresponding odorant. The six largest secondary groups are colored across all panels. Diagonal-block shades in the interaction matrix are added for better visualization of the simultaneous partitioning of odorants and receptors. In **(a)**, odorants with odor qualities consistent with the characteristic olfactory feature of the corresponding secondary group (Group 1, green/floral; Group 2, body/fermenting; Groups 3,4,5, culinary spices; Group 6, cheesy/acidic) are marked with asterisks. Also see Tables S1–S2 for all odor quality descriptors, with odorants sorted in the same order as in this figure.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

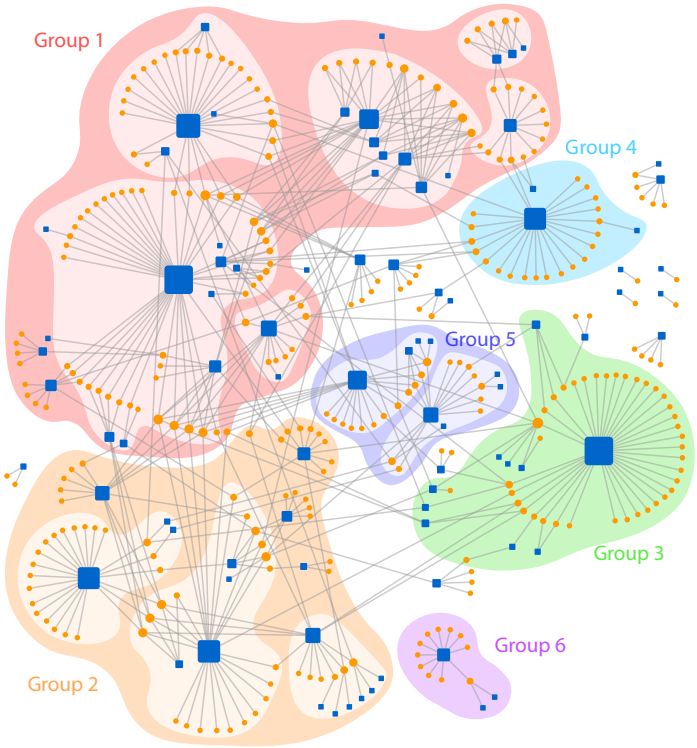


**Figure 5**: **Territories in the odorant-receptor network.** The six largest global groups, ranked by the number of receptors in the group, are shown in distinct colors. The sub-structures (local groups), if any, are also shown in lighter shades.

In Groups 3, 4 and 5, odorants are perceived as culinary spices: eugenol, eugenol acetate (the characteristic odor of clove); ethyl vanillin, cinnamon (common spices). In Group 3, which contains eugenol, we see food-related flavors: banana, nutmeg, butyl anthranilate (fruity odor); guaiacol (characteristic flavor for whiskey and roasted coffee); and 2-heptanone (characteristic flavor for gorgonzola cheese). These odorants with similar odor qualities, from which we draw the "characteristic olfactory feature" for each global group, are marked with asterisks in Fig. 4a (see also Tables S1–S2 for a full list of odor descriptors, for all odorants used in this study).

## DISCUSSION

We discovered a natural partitioning of the receptor code, by identifying groups of olfactory receptors (ORs) with correlated response patterns. Our grouping was performed without any label for the perceptual information; yet, this grouping procedure also partitioned the odorants into a few odor categories with similar odor descriptors. A correlated activity in a group of receptor codes for a certain olfactory feature is indicative of the existence of "labeled receptor groups". Below we will elaborate this idea and discuss its implications from a broader perspective of human odor perception.

### Receptor groups as the bases of perceptual odor space

There is a clearer association between the groups identified in the receptor code space and the major components in the perceptual odor space (the olfactory "features"). The presence of such association is supported by the following two observations. First, odorants in the same group tend to have similar perceptual descriptors, e.g., "plant-related" or "animal-related" odorant group. Second, we find that our result of odorant groups is consistent with the reported phenomenon of "olfactory white" [21] in which human subjects cannot distinguish odor mixtures if more than 30–40 odorants are randomly mixed spanning the odor space. In our analysis with odorant mixtures, the average receptor coverage by a mixture of 30–40 random odorants was equivalent to the coverage by the $\sim$6 highest-degree odorants. Moreover, a random sampling of 30 odorants was just enough to sample each of the 6 largest odor groups at least once (Fig. S3).

The presence of receptor groups implies a low-dimensional nature of the receptor code space; or eventually, of the perceptual odor space. Because the odor space is represented by the response pattern of $N \approx 400$ functional ORs, there are in principle $2^{400}$ possible OR states, even in the binary regime. But when a group of receptors respond in a correlated way, not all $2^{400}$ states are equally likely. Therefore, a significant amount of dimensionality reduction is made in the receptor code space. This aligns with the previous ideas involving the effective sparsity of odor space that, despite the apparent high-dimensionality implicated by physiochemical properties of odorants, olfaction is working in a much lower dimension in effect [22, 23]; or that the odor space is intrinsically clustered rather than uniform [21, 24].

The effective sparsity of the receptor code space, which may once again propagate to the perceptual odor space. has an analogy to vision. When one says that the natural visual scene is sparse [25], it means that the scene consists of a small number of features, such as lines or edges [26]. Receptor groups are like lines or edges, in the sense that they represent certain correlated patterns in the receptor space that are projected to the coarse-grained *olfactory features* in the perceptual odor space. Whereas feature-extracting coding was thought to be the realm of higher-level neurons [27, 28], here we found a concrete evidence that it already starts at the level of receptors.

### Combinatorial coding versus labeled lines: the hybrid strategy of human odor sensing

Our results point to a hybrid picture of the two coding hypotheses. At one extreme, one may assume that each receptor carries a specific signal, so that the glomerular array works as a set of *labeled lines*. Decoding is then unambiguous, as each OR is a unique

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

discriminator for an odor unit (odorant), and a mixture is the sum of its components. Because there is only a finite number of ORs available, however, coding with labeled lines has a clear limitation on its capacity. Alternatively, one can assume that odors are identified through the combinations of responding receptors. This latter strategy of *combinatorial coding* can accommodate a much larger number of distinct odor units than the number of ORs. In this case, let alone the biochemical cost of activating many ORs [29], decoding would be a highly non-trivial problem that requires further computation at higher-level neurons.

For the most part, the receptor code looks like a set of labeled lines (or labeled *groups* as we now prefer to call them), which maintains a set of dedicated channels for each odorant (or odor group). However, the receptor code also has signatures of combinatorial coding, in the sense that some odorants activate receptors from more than one groups.

The hybrid coding strategy is not entirely a new idea. It is known that the *sparse distributed coding* strategy, which mixes the *local coding* (labeled lines) and *dense distributed coding* (combinatorial coding), is used in visual sensing [25, 26]. There is growing evidence that insects employ a mixture of combinatorial coding and labeled-line strategies [30, 31]. Olfactory coding strategies in most organisms, including humans, are clearly more complicated than the dichotomy between labeled lines and combinatorial coding [32]. However, this is not to say that any evidence supporting either coding strategy is false. There are cases where the labeled-line structure is clearly at work [33], and there are other important cases in which the code cannot be understood without the idea of combinatorial coding [4]. Presumably, biological evolutions at both OR and neuron levels result in the hybrid strategy as a real solution. Our statistical analysis on the odorant-OR interaction data clearly demonstrates that hybridized coding strategy is already at work in the receptor code space, the very first layer of human olfaction [34, 35].

## Structures in the receptor code space, features in the perceptual odor space

Here, we further elaborate on our proposal that structures in the receptor code space represent the olfactory features in the perceptual odor space. Specifically, we put forward two hypotheses: (i) The large receptor groups found in the receptor code space through our global partitioning are "labeled" for the basic olfactory features that span the perceptual odor space. (ii) The hub odorants correspond to the salient stimuli in the natural olfactory environment.

First, we find that the olfactory features for the labeled receptor groups are in line with coarse-grained odor categories, e.g., "plant-related" and "animal-related" odors. This lends support to the idea that primary features of olfactory perception may already be "hard-coded" in the receptor code space [36], enabling innate discrimination of these features. Indeed, there is a clear advantage in keeping hard-coded labels for the basic olfactory cues, which would have been strongly conserved over the course of evolution [37] and would benefit from a faster response [38]. In fact, most examples of labeled-line receptors in other organisms are coarse-grained discriminators for certain groups of odorants [39, 40], or simply for good versus bad [33, 41], which is arguably the most important axis of odor perception for humans as well [23, 42, 43].

Second, at the core of our analysis are the hub odorants that elicit responses in the large population of receptors, which implies that degree is a measure of the functional importance of a stimulus in the receptor space. The degree of an odorant might also reflect its statistical importance in the natural odor space, the statistical weight of which may have been adjusted through the process of evolution. We hypothesize that the receptor code is designed to assign a larger number of receptors to a more *salient* stimulus, whose presence is more strongly correlated with a particular olfactory feature. The existence of salient odorants in the natural odor space is in fact plausible from the biological viewpoint. For example, because the types of odorants produced by the flower is an outcome of evolutionary selection, most flowers have a few common odorants, with their physiochemical properties attracting the pollinating insects [44].

When more salient stimulus is encoded by the response of more ORs, it has the advantage that important olfactory signals (carried by the salient odorants) are detected reliably, despite the perturbations in the receptor code space, e.g., the functional loss of certain ORs due to genetic mutation [45, 46]. If the hypothesis was correct, one would be able to predict the saliency of odorants in the natural odor space based on the number of receptors that respond. It is of great interest to explore whether the statistical salience has any relationship with the statistics of co-occurrence between odors [47].

## Concluding remarks

A set of concrete messages can be drawn from our study: (i) The existence of odorant groups, or the correlated patterns between the receptor codes, greatly reduces the dimensionality of the receptor code space and allows us to identify coarse-grained olfactory features in the perceptual odor space. (ii) The odorant and receptor groups revealed from our study clarify the balance between the labeled lines and combinatorial coding strategies. The labeled receptor groups on top of non-uniformly distributed receptor codes imply a hybrid strategy, which can be used to leverage both the fidelity of the labeled-line design and discriminative capacity of the combinatorial coding. (iii) Each large group of odorants is described by the coarse-grained olfactory feature (e.g. plant-related, animal-related odor) and the apparent hierarchy within each group. Our findings suggest a corresponding hierarchy in the perceptual odor space, which opens up a series of questions as to the natu-

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ral odor statistics and its relation to the statistics of their receptor codes. Despite several simplifying assumptions in this study, e.g., neglecting the dependence on odor concentration, competitive binding between two different types of odorants [48], the effect of inverse agonists [49], as well as any temporal effects [50], essential findings from our analysis on human receptor code space should be preserved (e.g., see Fig. S2 for the effect of odorant concentration on degree distribution).

In summary, by analyzing the dataset of OR responses, we identified a modular structure inherent in the receptor code space of human olfaction. Insights from these findings could be extended to the broader study of human odor perception and help understand exceptionally diverse data on olfaction in a more systematic manner.

# METHODS

## Modeling

**Network representation of the receptor code space.** We analyze the dose-response curves of 535 interacting odorant-receptor pairs, involving 89 odorants and 303 olfactory receptors (ORs), from the dataset of [14]. The interacting odorant-OR pairs were screened from a library of 511 human OR genes [13, 14]; the final number of 303 ORs include only those that respond to at least one odorant in the dataset. The panel of odorants was selected to span the 20-dimensional physiochemical space that explains the variance in mammalian OR responses [12]. The receptors in the dataset, which cover a near-full ($\sim 3/4$) space of known human ORs, can be viewed as the receptor-space representation of the odorant. On the other hand, the set of cognate odorants found for each receptor, which is only a small sampling of the olfactory environment (humans can detect at least $\gtrsim 10^4$ odorants), is not guaranteed to be complete. We note that the dataset also includes several natural odors that are not monomolecular odorants (for example "banana"), which do not exactly fit in our formulation of pairwise odorant-OR interactions. Nevertheless, these are only a minor fraction of the dataset, and we included all reported odors in our analysis.

A graph can be formally written as $G = (V, E)$, where $V$ is the set of all nodes ("vertices") in the graph, and $E$ is the set of all edges. In the binary regime, each edge in $E$ is an unordered pair of nodes in $V$; this gives a binary graph, which can also be represented in terms of an *adjacency matrix* $A$, whose $(i, j)$-th element $A_{ij}$ represents whether there is an edge between the two nodes $i$ and $j$ in the graph. Specifically, the odorant-receptor network is a *bipartite graph*, where each node belongs to either of two node types (either odorants or receptors) and every edge connects one odorant node to one receptor node. If $V_O$ and $V_R$ are the sets of all odorants nodes and the set of all receptors nodes, respectively, each edge in $E$ connects exactly one node in $V_O$ and the other node in $V_R$. This bipartite structure is cast into the $N \times M$ sub-matrix of the full adjacency matrix, $A_{i\lambda}$, where $N = |V_R|$ is the number of receptor nodes in the network, and $M = |V_O|$ the number of odorant nodes. Each element of the matrix is defined as $A_{i\lambda} = 1$ if and only if there is an edge between receptor $i$ and odorant $\lambda$. We call $A_{i\lambda}$ the *interaction matrix*.

**Binary vector representation of a receptor code.** The receptor code space representation of an odorant, written in a binary vector $\mathbf{y}$, is equivalent to the corresponding column of the interaction matrix. That is, $[\mathbf{y}_\lambda]_i = A_{i\lambda}$, where $[\mathbf{y}]_i$ is the $i$-th element of vector $\mathbf{y}$.

The union of two binary vectors is defined element-wise as $[\mathbf{y}_1 \cup \mathbf{y}_2]_i = [\mathbf{y}_1]_i \vee [\mathbf{y}_2]_i$, where $\vee$ is the logical "or" operation between two binary variables. Similarly, the intersection is defined as $[\mathbf{y}_1 \cap \mathbf{y}_2]_i = [\mathbf{y}_1]_i \wedge [\mathbf{y}_2]_i$, where $\wedge$ is the logical "and" operation. We also define the difference of two binary vectors as $[\mathbf{y}_1 \setminus \mathbf{y}_2]_i = [\mathbf{y}_1]_i \wedge \neg[\mathbf{y}_2]_i$, where $\neg$ is the logical "not" operation. For the union and the intersection, the definitions easily extend to more than two binary vectors.

**Degree of an odorant.** The *degree* of a node is the number of edges attached to it, or (when there is no loop) the number of other nodes that are connected to it. It is useful to define the *size* of a binary vector as the number of its non-zero elements: $|\mathbf{y}| = \sum_{i=1}^{N} [\mathbf{y}]_i$.

The degree of an odorant node is the number of receptor nodes connected to it:

$$k_\lambda = \sum_{i \in V_R} A_{i\lambda} = |\mathbf{y}_\lambda|, \quad \lambda \in V_O. \tag{1}$$

In the current study, we primarily consider the odorant degree, or how broadly (or narrowly) an odorant is "targeting" the receptors. This breadth of interaction for the odorant is well-defined, because there is a finite number of olfactory receptors. The dataset in [14] covers a majority of the known (functional) human olfactory receptor repertoire: 303 out of $\sim 400$ ORs. Our approach has advantage over other studies that considered whether a given receptor is "narrowly tuned" or "broadly tuned" across different odors, which is hard to quantify without a good metric of the odor space.

The degree distribution of odorants, $P_O(k)$, is measured by counting the number of odorants at each degree $k$. $P_O(k)$ appears approximately straight in the log-log scale, and is fitted to the curve, $P(k) \sim k^{-\alpha}$. However, the fit does not necessarily indicate an underlying distribution that is power-law in a strict sense; we are simply using the power-law dependence to describe the heavy-tailed shape of the degree distribution.

**Co-activation graph.** The *co-activation graph* is a projection of the original bipartite graph to one of the two node types. For example, the co-activation graph of receptors, $G_R = (V_R, E_{R;O})$, inherits all receptor nodes $V_R$ from the original bipartite graph. Two

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

receptor nodes in the co-activation graph $G_R$ are connected by an edge if there is a common odorant node in the original graph $G$ that is connected simultaneously to both receptors (see Fig. S4a,f). To define the edges, $E_{R;O}$, it is sufficient to define the corresponding adjacency matrix $A_R$, where $(A_R)_{ij} = 1$ indicates an edge between two receptors $i, j$ in the co-activation graph. We call $A_R$ the *co-activation matrix* of receptors. Each element of the co-activation matrix is given as $(A_R)_{ij} = \bigvee_{\lambda \in V_O} (A_{i\lambda} \wedge A_{j\lambda})$, which indicates whether there is at least one odorant $\lambda$ for which $A_{i\lambda}$ and $A_{j\lambda}$ are both 1, where $A_{i\lambda}$ is the interaction matrix for the original bipartite graph.

**Receptor code redundancy.** We define the receptor code redundancy as $\chi = (n_{\text{full}} - n_{\text{eff}})/n_{\text{full}}$, where $n_{\text{full}}$ is the full impact of a signal in the absence of redundancy, and $n_{\text{eff}}$ is the net effect of the signal under redundancy. For a discrimination task of a target odorant $\lambda$ in the presence of a background odorant $\mu$, the effect of redundancy reduces the net effect of adding $\lambda$ from $n_{\text{full}} = |\mathbf{y}_\lambda|$ to $n_{\text{eff}} = |\mathbf{y}_\lambda \setminus \mathbf{y}_\mu|$. In this case, the receptor code redundancy is $\chi_{\lambda|\mu} = (|\mathbf{y}_\lambda| - |\mathbf{y}_\lambda \setminus \mathbf{y}_\mu|)/\,|\mathbf{y}_\lambda| = |\mathbf{y}_\lambda \cap \mathbf{y}_\mu|\,/\,|\mathbf{y}_\lambda|$.

## Grouping algorithms

We will provide details of the grouping process in three steps: node ranking, primary grouping and secondary grouping. To start, we rank the odorants and receptors in the dataset.

**Ranking odorants and receptors by the degrees.** Odorants are ranked by their degree of interaction, $k$ (Eq. 1). When there are multiple odorants with the same degree, they are ranked according to the order they are indexed in the original dataset [14]. We call this ordering $h_0$, such that each odorant $\lambda$ (where $\lambda$ may be an arbitrary index) is assigned a unique rank $h_0(\lambda) \in \{1, \cdots, M\}$.

Receptors are ranked by their degree in the co-activation graph. The degree in the co-activation graph represents the number of other receptors that share at least one odorant partner with the given receptor. When there are multiple receptors with the same degree of co-activation, they are ranked according to the order they are indexed in the original dataset [14]. Similarly as for the odorants, we call this ordering $g_0$, such that each receptor $i$ is assigned a unique rank $g_0(i) \in \{1, \cdots, N\}$.

**Primary grouping.** We first partition the receptor space into a set of non-overlapping groups, and subsequently use these to classify the odorants based on the receptor code overlap.

In the primary receptor grouping ($g_1$), receptors are grouped such that each receptor "signs up" to the largest receptor clique it belongs to. Specifically, we assign to each receptor the rank $h_0$ of the highest-degree odorant it interacts with; in case of ties, we choose odorants of higher ranks (smaller $h_0$). We re-rank the receptor group index while preserving the order of the associated rank $h_0$, by assigning the primary group index $I = 1$ to the largest receptor group, and the index $I = 2$ to the next largest group, and so on ($I = 1, 2, \cdots I_{\text{max}}$). This results in a grouping $g_1 \colon i \mapsto I$ that maps each receptor $i$ to a group index $g_1(i) = I$.

In the primary odorant grouping ($h_1$), each odorant is assigned to the above-determined receptor group ($g_1$) to which its receptor code overlaps the most. The primary odorant grouping $h_1 \colon \lambda \mapsto I$ (or $h_1(\lambda) = I$) is defined by choosing the receptor group of index $I$, with which the odorant's receptor code ($\mathbf{y}_\lambda$) displays the maximum overlap, namely $I = \text{argmax}_J\, \chi_{\lambda|J}$. Here, the receptor code overlap of an odorant $\lambda$ to a receptor group index $I$ is quantified as $\chi_{\lambda|I} = |\mathbf{y}_\lambda \cap \mathbf{y}_I|\,/\,|\mathbf{y}_\lambda|$, where the binary vector $\mathbf{y}_I$ represents the receptor code for the group $I$ by having each element $(\mathbf{y}_I)_i = 1$ if $g_1(i) = I$, and $(\mathbf{y}_I)_i = 0$ otherwise. Note that the resulting odorant group indices span through $I = 1, 2, \cdots, I_{\text{max}}$, shared with the receptor group indices.

**Secondary grouping.** We merge the primary groups based on the amount of co-activation in the receptor code space.

In the secondary receptor grouping ($g_2$), pairs of primary receptor groups are merged if there is a strong co-activation. For two primary receptor groups $I, J$, we calculate the amount of co-activation between the two groups, simply in terms of the density of off-diagonal-block contribution in the co-activation matrix $A_R$:

$$v_{IJ} = \frac{\sum_{i \in I} \sum_{j \in J} (A_R)_{ij}}{\sum_{i \in I} \sum_{j \in J} 1} = \frac{1}{|I| \cdot |J|} \sum_{i \in I, j \in J, i < j} (A_R)_{ij}, \quad (2)$$

where $i \in I$ is a shorthand to indicate that the sum runs over all receptors $i$ such that $g_1(i) = I$. We merge each pair of primary groups $I$ and $J$ whenever their co-activation is stronger than a threshold, i.e., if $v_{IJ} \geq \bar{v}$. Here we used an optimal threshold value of $\bar{v} = 0.15$, where the grouping was most informative (Fig. S5; also see Statistical validation). Note that this is a greedy and transitive grouping. For example, if there are three primary groups $I_1, I_2, I_3$ such that $v_{I_1 I_2}, v_{I_2 I_3} > \bar{v}$ but $v_{I_3 I_1} < \bar{v}$, then all three are merged together in the secondary grouping; the two groups $I_1, I_2$ are first combined, then $I_3$ is also pulled into the group that $I_2$ belongs to. After performing all pairwise merges, we once again re-rank the resulting group indices while preserving the order. This defines a grouping $G_{12} \colon I \mapsto I'$ that maps each primary group $I$ to a (re-numbered) secondary group $I'$. Each receptor $i$ is then assigned a secondary group index $g_2(i) = G_{12}(g_1(i))$.

In the secondary odorant grouping ($h_2$), pairs of primary odorant groups are merged if there is a strong co-activation. Because the primary grouping was a simultaneous partitioning of odorants and receptors, the above merging operation of receptor groups from $g_1(i)$ to $g_2(i)$ automatically propagates to the odorant groups. Specifically, each odorant $\lambda$ is assigned a secondary group index $h_2(\lambda) = G_{12}(h_1(\lambda))$.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

## Statistical validation

**Optimal threshold for secondary receptor grouping.** In the secondary receptor grouping, our goal was to obtain the best partitioning of the receptors so that the resulting structure captures the co-activation pattern. Because we start from the primary receptor groups as the "units", and work by merging them based on the overlaps, the grouping depends on the threshold $\bar{\chi}$. In the limit where the threshold is too small, most units are merged together and a giant cluster is formed (Fig. S5a). On the other hand, when the threshold is too large, most units stay un-merged, failing to capture the global structure (Fig. S5c). The optimal threshold is where the grouping is most *informative*, in the sense that more strongly co-activating primary groups are merged together in the same secondary group, and non-co-activating groups stay apart (Fig. S5b).

Here we consider our grouping as a binary classifier, where each pair of receptors are either grouped together or not. Then the informativeness of grouping can be quantified in terms of the diagnostic ability of the classifier. In particular, there are two success rates we want to maximize: the true positive rate (TPR) and the true negative rate (TNR). The TPR, also called the *sensitivity* or *recall*, is defined as the fraction of co-activated receptor pairs (non-zero elements in the co-activation matrix) that are correctly grouped together. The TNR, also called the *specificity*, is defined as the fraction of non-co-activated receptor pairs (zero elements in the co-activation matrix) that are correctly *not* grouped together. Because our "units" of grouping are the primary receptor groups, which are considered fixed, we only count the elements of the co-activation matrix that belong to different primary groups (off-block elements). It is also common to consider the false positive rate (FPR; also called the *fall-out*) instead of the TNR, with a simple relationship FPR $= 1-$TNR.

We plot the sensitivity of grouping versus the $1-$specificity, at varying threshold $\bar{\chi}$, also known as the receiver operating characteristic (ROC) curve, (Fig. S5d). A perfect classifier corresponds to the upper left corner of the ROC plane, where the sensitivity and the specificity are both 1; the results of a random guess would sit along the positive diagonal (also called the no-discrimination line). We define the optimal threshold as the point where the curve most closely approaches the top left corner of the ROC plane; more specifically, we minimize the combined loss function $((1-\text{sensitivity})^2 + (1-\text{specificity})^2)$, or the sum of squared normalized error. The best secondary grouping of the co-activation pattern of receptors is thus obtained at $\bar{\chi} = 0.15$ (Fig. S5b).

**Comparison to null models.** To confirm that the existence of receptor groups is a special property of the receptor code, we constructed null models with randomized interactions, constrained only by the marginal degree statistics of the data. First, we considered a random interaction network of 89 odorants and 303 receptors, where each odorant-receptor pair interacts with a uniform probability. The only constraint was that there are exactly 535 interacting pairs, as in the data; this is equivalent to fixing the average degree of an odorant. We observed that this random network groups poorly, reflecting the fact that it is essentially unstructured (Fig. S6a). Second, we considered a more constrained null model where the degree distribution is preserved. We obtained this by shuffling each column of the activation matrix $A_{i\lambda}$ independently, such that each odorant re-connects to the receptors randomly, while the degree is kept fixed. The primary groups are observed as large blocks in the co-activation matrix, because the high-degree odorants are preserved; however, the secondary grouping is still poor (Fig. S6b). This shows that the global structure with well-partitioned groups is a special property of the receptor code, which involves more than just the degree distribution. Note that due to randomization, the degree correlations in these null models are also bound to be neutral; the structure is not easily captured by the simple standard statistics.

## Code and data availability

List of odorants with odor descriptors are provided in Tables S1 and S2. The perceptual odor descriptors were obtained from multiple sources, including public databases, academic works and industrial reports. Data sources are listed in Table S3.

Formatted data files to reproduce the network representation of odorant-receptor interaction with full attributes for each node and edge (as .cyjs and .cys files, to open in Cytoscape), and the Matlab code for the grouping analysis along with associated documentation, are available at `https://github.com/jihyunbak/ORnetwork`.

## ACKNOWLEDGMENTS

We thank the KIAS Center for Advanced Computation for providing computing resources. This study was partly supported by the grants from the National Research Foundation of Korea (NRF-2018R1A2B3001690) (C.H.), and the National Science Foundation (CHE-1362926) (S.J.J).

## AUTHOR CONTRIBUTIONS

J.H.B., S.J.J. and C.H. designed the projects. J.H.B. and C.H. performed research and analyzed the data. J.H.B., S.J.J. and C.H. wrote the paper.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

## SUPPLEMENTARY MATERIALS

Supplementary Materials include six figures and three tables, with accompanying text to each item. All items are available at the end of this manuscript in the following order.

**Figure S1.** Interaction network with interaction properties. (Related to Fig. 2a)

**Figure S2.** Interaction network at fixed odorant concentrations. (Related to Fig. 2b)

**Figure S3.** Odorant mixture tests. (Related to Fig. 2c and Fig. 4)

**Figure S4.** Grouping procedure. (Related to Fig. 3)

**Figure S5.** Finding the optimal threshold for partitioning. (Related to Fig. 4)

**Figure S6.** No clear grouping by chance. (Related to Fig. 4)

**Tables S1-S2.** List of odorants with perceptual descriptors. (Related to Fig. 4)

**Table S3.** List of data sources.

## References


[1] Su, C. Y., Menuz, K., & Carlson, J. R. Olfactory Perception: Receptors, Cells, and Circuits. *Cell* **139**, 45–59 (2009).

[2] Block, E., Jang, S., Matsunami, H., Sekharan, S., Dethier, B., et al. Implausibility of the vibrational theory of olfaction. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2766–E2774 (2015).

[3] DeMaria, S & Ngai, J. The cell biology of smell. *J. Cell. Biol.* **191**, 443–452 (2010).

[4] Malnic, B., Hirono, J., Sato, T., & Buck, L. B. Combinatorial Receptor Codes for Odors. *Cell* **96**, 713–723 (1999).

[5] Stevens, C. F. What the fly's nose tells the fly's brain. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9460–9465 (2015).

[6] Dasgupta, S., Stevens, C. F., & Navlakha, S. A neural algorithm for a fundamental computing problem. *Science* **358**, 793–796 (2017).

[7] Roland, B., Deneux, T., Franks, K. M., Bathellier, B., & Fleischmann, A. Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex. *eLife* **6**, e26337 (2017).

[8] Zhang, Y & Sharpee, T. O. A robust feedforward model of the olfactory system. *PLoS Comp. Biol.* **12**, 1–15 (2016).

[9] Zwicker, D., Murugan, A., & Brenner, M. P. Receptor arrays optimized for natural odor statistics. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5570–5575 (2016).

[10] Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., et al. Predicting human olfactory perception from chemical features of odor molecules. *Science* **355**, 820–826 (2017).

[11] Poivet, E., Tahirova, N., Peterlin, Z., Xu, L., Zou, D.-J., et al. Functional odor classification through a medicinal chemistry approach. *Science Adv.* **4**, eaao6086 (2018).

[12] Saito, H., Chi, Q., Zhuang, H., Matsunami, H., & Mainland, J. D. Odor Coding by a Mammalian Receptor Repertoire. *Science Signaling* **2**, ra9–ra9 (2009).

[13] Mainland, J. D., Keller, A., Li, Y. R., Zhou, T., Trimmer, C., et al. The missense of smell: functional variability in the human odorant receptor repertoire. *Nat. Neurosci.* **17**, 114–120 (2013).

[14] Mainland, J. D., Li, Y. R., Zhou, T., Liu, W. L. L., & Matsunami, H. Human olfactory receptor responses to odorants. *Scientific Data* **2**, 150002 (2015).

[15] Araneda, R. C., Kini, A. D., & Firestein, S. The molecular receptive range of an odorant receptor. *Nat. Neurosci.* **3**, 1248 (2000).

[16] Bak, J. H., Jang, S. J., & Hyeon, C. Implications for human odor sensing revealed from the statistics of odorant-receptor interactions. *PLoS Comp. Biol.* **14**, 1–16 (2018).

[17] Hallem, E. A & Carlson, J. R. Coding of Odors by a Receptor Repertoire. *Cell* **125**, 143–160 (2006).

[18] Lin, D. Y., Shea, S. D., & Katz, L. C. Representation of natural stimuli in the rodent main olfactory bulb. *Neuron* **50**, 937–949 (2006).

[19] Mathis, A., Rokni, D., Kapoor, V., Bethge, M., & Murthy, V. N. Reading out olfactory receptors: Feedforward circuits detect odors in mixtures without demixing. *Neuron* **91**, 1–19 (2016).

[20] Rokni, D., Hemmelder, V., Kapoor, V., & Murthy, V. N. An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**, 1225–1232 (2014).

[21] Weiss, T., Snitz, K., Yablonka, A., Khan, R. M., Gafsou, D., et al. Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19959–19964 (2012).

[22] Secundo, L., Snitz, K., & Sobel, N. The perceptual logic of smell. *Curr. Opin. Neurobiol.* **25**, 107–115 (2014).

[23] Koulakov, A. A., Kolterman, B. E., Enikolopov, A. G., & Rinberg, D. In search of the structure of human olfactory space. *Front. Syst. Neurosci.* **5**, 65 (2011).

[24] Castro, J. B., Ramanathan, A., & Chennubhotla, C. S. Categorical Dimensions of Human Odor Descriptor Space Revealed by Non-Negative Matrix Factorization. *PLoS ONE* **8** (2013).

[25] Field, D. J. What Is the Goal of Sensory Coding? *Neural Computation* **6**, 559–601 (1994).

[26] Olshausen, B. A & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.* **37**, 3311–3325 (1997).

[27] Jeanne, J. M., Fişek, M., & Wilson, R. I. The Organization of Projections from Olfactory Glomeruli onto Higher-Order Neurons. *Neuron* **98**, 1198–1213.e6 (2018).

[28] Yoshida, I & Mori, K. Odorant Category Profile Selectivity of Olfactory Cortex Neurons. *J. Neurosci.* **27**, 9105–9114 (2007).

[29] Mehta, P & Schwab, D. J. Energetic Costs of Cellular Computation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17978–17982 (2012).

[30] Andersson, M. N., Löfstedt, C., & Newcomb, R. D. Insect olfaction and the evolution of receptor tuning. *Front. Ecol. Evol.* **3**, 53 (2015).

[31] Haverkamp, A., Hansson, B. S., & Knaden, M. Combinatorial codes and labeled lines: How insects use olfactory cues to find and judge food, mates, and oviposition sites in complex environments. *Front. Physiol.* **9**, 49 (2018).

[32] Howard, J. D & Gottfried, J. A. Configural and elemental coding of natural odor mixture components in the human brain. *Neuron* **84**, 857–869 (2014).

[33] Li, Q & Liberles, S. D. Aversion and Attraction through Olfaction. *Curr. Biol.* **25**, R120–R129 (2015).

[34] Berck, M. E., Khandelwal, A., Claus, L., Hernandez-Nunez, L., Si, G., et al. The wiring diagram of a glomerular olfactory system. *eLife* **5** (2016).

[35] Zwicker, D. Normalized neural representations of complex odors. *PLOS ONE* **11**, 1–16 (2016).

[36] Lapid, H., Shushan, S., Plotkin, A., Voet, H., Roth, Y., et al. Neural activity at the human olfactory epithelium reflects olfactory perception. *Nat. Neurosci.* **14**, 1455 (2011).

[37] Hughes, G. M., Teeling, E. C., & Higgins, D. G. Loss of olfactory receptor function in hominin evolution. *PLoS ONE* **9**, 1–6 (2014).

[38] Boesveldt, S., Frasnelli, J., Gordon, A. R., & Lundström, J. N. The fish is bad:

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Negative food odors elicit faster and more accurate reactions than other odors. *Biol. Psychol.* **84**, 313–317 (2010).

[39] Ai, M., Min, S., Grosjean, Y., Leblanc, C., Bell, R., et al. Acid sensing by the Drosophila olfactory system. *Nature* **468**, 691 (2010).

[40] Kurtovic, A., Widmer, A., & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a Drosophila sex pheromone. *Nature* **446**, 542 (2007).

[41] Semmelhack, J. L & Wang, J. W. Select Drosophila glomeruli mediate innate olfactory attraction and aversion. *Nature* **459**, 218 (2009).

[42] Khan, R. M., Luk, C.-H., Flinker, A., Aggarwal, A., Lapid, H., et al. Predicting Odor Pleasantness from Odorant Structure: Pleasantness as a Reflection of the Physical World. *J. Neurosci.* **27**, 10015–10023 (2007).

[43] Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B., Matsunami, H., et al. Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–72 (2007).

[44] Wright, G. A & Thomson, M. G. Odor Perception and the Variability in Natural Odor Scenes. In Romeo, J. T, editor, *Recent Advances in Phytochemistry* volume 39 pages 191–226. Elsevier (2005).

[45] Fleischmann, A., Shykind, B. M., Sosulski, D. L., Franks, K. M., Glinka, M. E., et al. Mice with a "Monoclonal Nose": Perturbations in an Olfactory Map Impair Odor Discrimination. *Neuron* **60**, 1068–1081 (2008).

[46] Keller, A & Vosshall, L. B. Influence of odorant receptor repertoire on odor perception in humans and fruit flies. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5614–9 (2007).

[47] Zhou, Y., Smith, B. H., & Sharpee, T. O. Hyperbolic geometry of the olfactory space. *Sci. Adv.* **4** (2018).

[48] Singh, V., Murphy, N. R., Balasubramanian, V., & Mainland, J. D. A competitive binding model predicts nonlinear responses of olfactory receptors to complex mixtures. *bioRxiv* (2018).

[49] Reddy, G., Zak, J. D., Vergassola, M., & Murthy, V. N. Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures. *eLife* **7**, e34958 (2018).

[50] Su, C.-Y., Martelli, C., Emonet, T., & Carlson, J. R. Temporal coding of odor mixtures in an olfactory receptor neuron. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5075–5080 (2011).

[51] Jang, S & Hyeon, C. Kinetic Model for the Activation of Mammalian Olfactory Receptor. *J. Phys. Chem. B* **121**, 1304–1311 (2017).

[52] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504 (2003).

[53] Koulakov, A., Gelperin, A., & Rinberg, D. Olfactory Coding With All-or-Nothing Glomeruli. *J. Neurophysiol.* **98**, 3134–3142 (2007).

[54] Zeng, X.-N., Leyden, J. J., Lawley, H. J., Sawano, K., Nohara, I., et al. Analysis of characteristic odors from human male axillae. *J. Chem. Ecol.* **17**, 1469–1492 (1991).

[55] Jacob, T. J., Wang, L., Jaffer, S., & McPhee, S. Changes in the odor quality of androstadienone during exposure-induced sensitization. *Chem. Senses* **31**, 3–8 (2006).

[56] Arctander, S. *Perfume and Flavor Materials of Natural Origin*. Self Publication (1960).

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

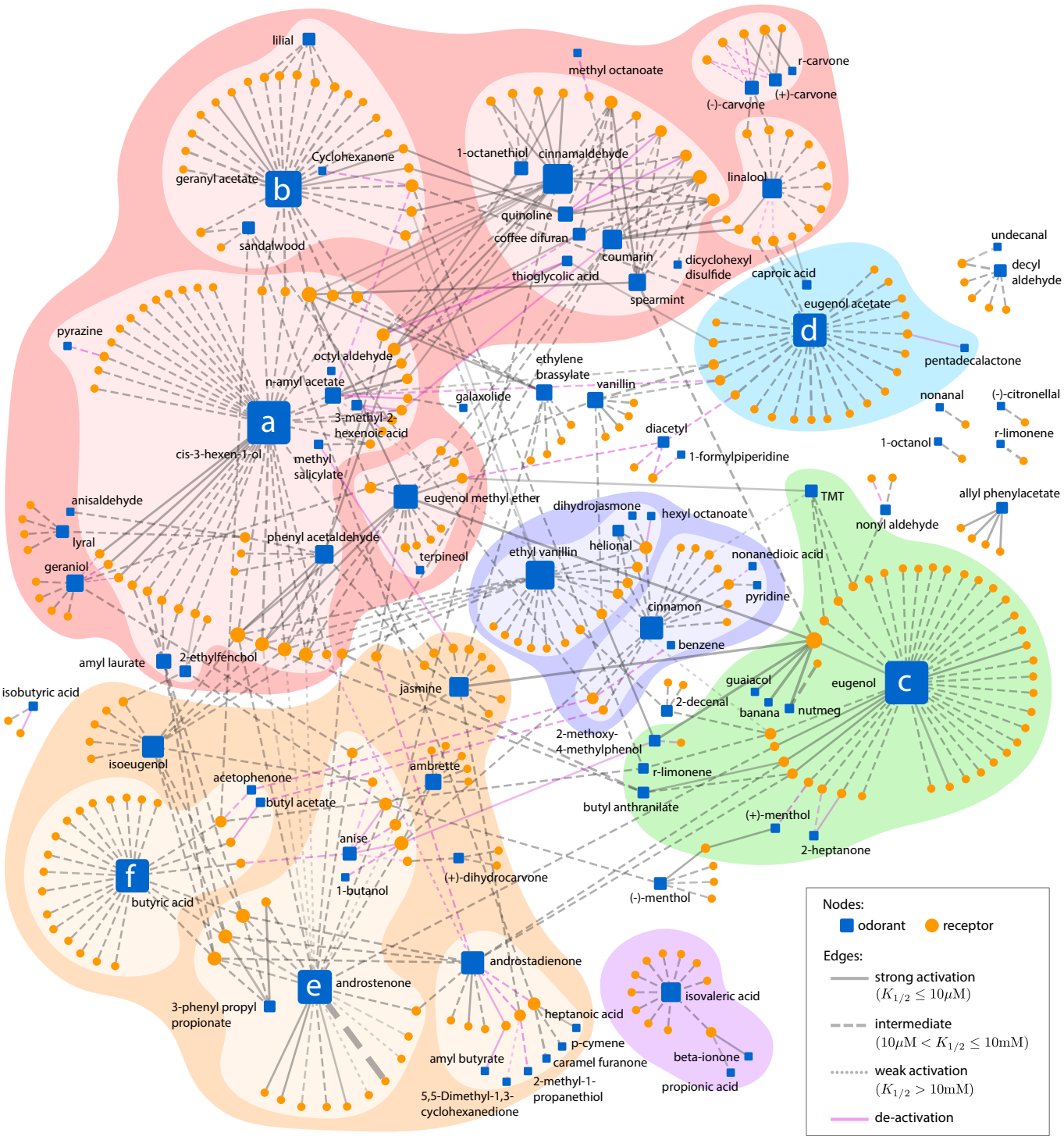


**Figure S1**: **Interaction network with full edge attributes.** Edge thickness indicates the strength of activation, and the transparency of an edge is used to quantify the correlation coefficient of the fit (more transparent means a smaller correlation coefficient). Edge color indicates the sign of interaction, such that gray for activating pairs, magenta for de-activating pairs. The line style of an edge indicates the potency of the corresponding interaction, as shown in the legends. The distribution of $K_{1/2}$ has a peak around the average value $K^*_{1/2} \approx \langle K_{1/2} \rangle \sim 100\,\mu$M [16]. A network file with full information on all nodes and edges, prepared for easy interactive access, is also available online (`https://github.com/jihyunbak/ORnetwork`).

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

### Interaction network at fixed odorant concentrations

Here we show how the analysis used in our work can be extended beyond the binary regime by taking into account the effect of odorant concentration.

We start from previous characterization of the pairwise interactions in the dataset, based on a minimal kinetic model [51]. For each interacting pair of an odorant and a receptor, the dose-response curve $r(c|\theta)$, or the activity of the receptor in response to an odorant concentration $c$, is characterized with four parameters $\theta = \{B, E, K_{1/2}, H\}$ as $r(c|\theta) = B + (E - B) \cdot c^H / (c^H + K_{1/2}^H)$. Here $B$ is the basal activity of the receptor; $E$ is the efficacy, or the maximum response from the interaction; $K_{1/2}$ is the $\mathrm{EC}_{50}$, the effective concentration of the odorant required to elicit half-maximum response; and $H$ is the Hill coefficient. The interaction is either activating ($E > B$), or deactivating ($E < B$). In the previous study [16], we determined the four parameters $\{B, E, K_{1/2}, H\}$ for each of the 535 interacting odorant-receptor pairs from the dataset of [14]. The interaction parameters are displayed as the edge attributes in Fig. S1. A network file with full information on all nodes and edges, ready to import in an interactive platform (Cytoscape [52]), is also available online (`https://github.com/jihyunbak/ORnetwork`).

Now we construct a binary interaction network at a fixed odorant concentration, $c$, assumed to be uniform for all odorants considered. In this case, an edge between each odorant-receptor pair indicates that the receptor is *activated* when the odorant is present at this concentration $c$; this selects all activations (positive responses) with $c > K_{1/2}$, and excludes the de-activations (negative responses). This binarization scheme is consistent with the prediction that the olfactory response is effectively binary at the level of receptor neurons [16]; it is also related to the ON/OFF model considered in [53]. Varying the odorant concentration $c$ uniformly for all odorants, we obtain a family of binary (unweighted and undirected) interaction networks (see Fig. S2a). As expected, the average degree of an odorant decreases as $c$ decreases (Fig. S2b). Specifically, the average degree is $\langle k \rangle \simeq 6$ at high odorant concentration ($c \gg K_{1/2}^* \sim 100$ $\mu$M [16]), but is reduced to $\langle k \rangle \simeq 3.6$ at $c \approx K_{1/2}^*$. The degree distributions still look qualitatively similar (consistent with a power law), while the decay becomes steeper at lower concentrations (Fig. S2c).

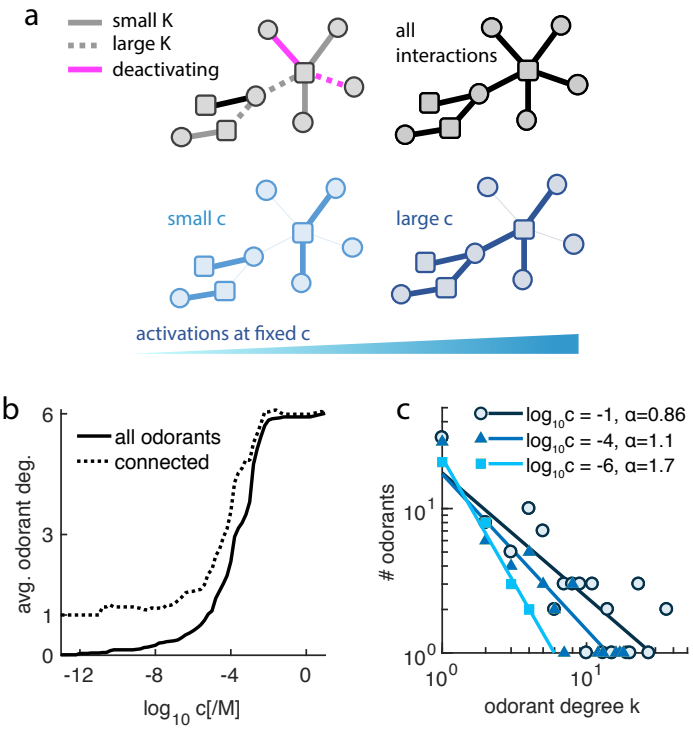


**Figure S2**: **Interaction network at fixed odorant concentrations.** **a.** Construction of the binary interaction network at fixed odorant concentration, $c$, assumed uniform for all odorants. **b.** The average degree of odorants at varying $c$. Average was taken either over all the odorants in the network (solid line) or over the odorants connected to at least one receptor at the respective concentration level (dotted line). **c.** Degree distributions obtained by counting activations at three fixed odorant concentrations ($c$ =0.1 M, 100 $\mu$M, 1 $\mu$M), each showing a heavy tail but a different rate of decay.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

### Odorant mixture tests

We selected mixtures of odorants from the dataset of [14]. Varying the number of odorants in the mixture, $m$, we counted the number of receptors that respond to at least one odorant in the mixture, $n_{\text{union}} = |\mathbf{y}_1 \cup \cdots \cup \mathbf{y}_m|$, and the sum of individual odorant degrees, $n_{\text{sum}} = |\mathbf{y}_1| + \cdots + |\mathbf{y}_m|$ (Fig. S3a).

Odorants were selected either from the highest-degree odorants (Fig. S3a-b, red), or randomly (gray line; average over 500 samplings each). To confirm that the receptor code redundancy is significantly small for the hub odorants, we measured how the receptor code redundancy grows with the number of odorants in a mixture, where the mixture consists of either the highest-degree odorants or a random sample of odorants. Analogous to the case of two-odorant discrimination, the receptor code redundancy in a mixture is quantified in terms of a fraction of overlapping receptors, $\chi_{\text{mix}} = (n_{\text{sum}} - n_{\text{union}})/n_{\text{sum}}$. When we drew $m$ odorants from the database at random, the redundancy increased almost linearly with the total number of receptors involved in the representation of the mixture (Fig. S3b, gray). On the other hand, when we selected odorants in the descending order of their degrees, the redundancy curve for the high-degree-odorant mixture (Fig. S3b, red) stayed below the chance curve for random selection (gray). Maximum deviation from the chance curve was observed when the six hub odorants identified in Fig. 2 were selected. The mixture of the six hub odorants elicited responses in $n_{\text{union}} = 155$ receptors (about half of the receptor space), with a small redundancy of $\chi_{\text{mix}} \lesssim 0.1$; whereas in a random-odorant mixture, the same level of redundancy ($\chi_{\text{mix}} \approx 0.1$) was reached already from a population of $n_{\text{union}} \approx 70$ receptors, with $m \approx 13$ odorants in the mixture on average.

We also tested how many odor groups, as defined by the secondary grouping in our analysis in this paper, are sampled by a random mixture of $m$ odorants (Fig. S3c). The number of top 6 odor groups sampled by a random selection of $m$ odorants from the dataset was closer to 6 than to 5 at $m \gtrsim 30$ (see blue line and shade). Result is based on 500 independent sampling for each $m$. The standard errors are plotted but are smaller than the marker size at all points. If we instead consider the top 5 groups only, the cutoff is $m \gtrsim 25$ (red line and shade).

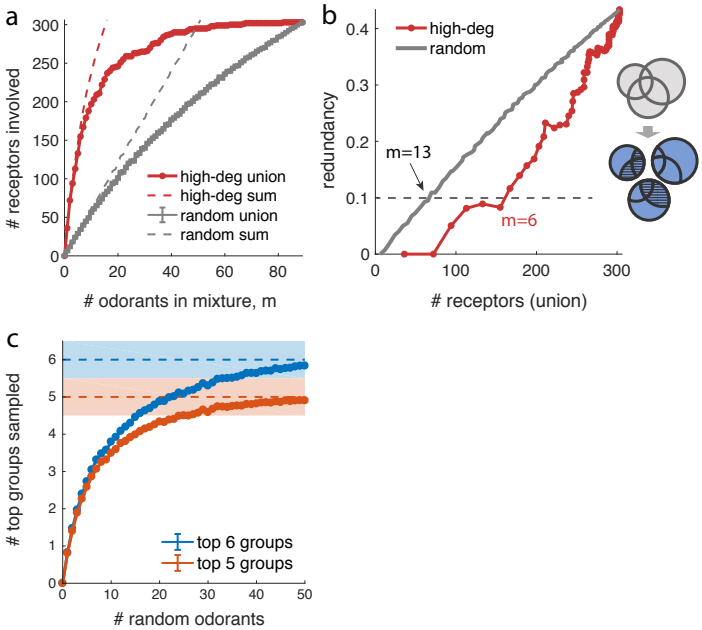


**Figure S3**: **Odorant mixture tests. a.** Receptor coverage and **b.** redundancy in odorant mixtures. The mixture contains either $m$ highest-degree odorants, or a random sampling of $m$ odorants drawn from the dataset without repetition. The receptor code redundancy of the mixture ($\chi_{\text{mix}}$) is defined by the relative size of overlap (striped area) with respect to the size of total signal, $n_{\text{sum}}$ (total blue area). It is plotted against the size of net signal, $n_{\text{union}}$ (unstriped blue area). **c.** Number of distinct secondary odorant groups sampled by a random odorant mixture of varying size.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

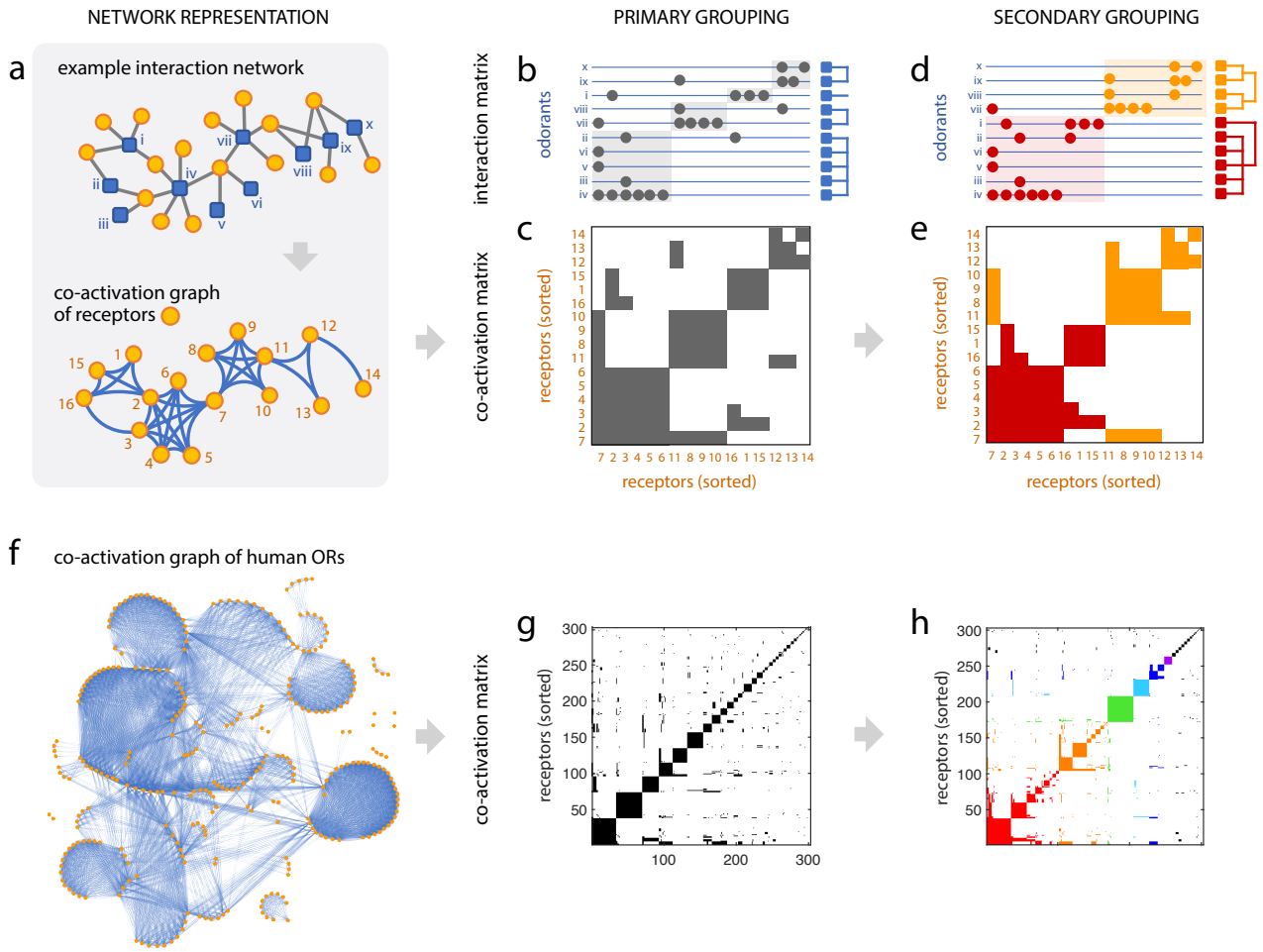


**Figure S4**: **The grouping procedure. a-e.** Illustration using an example interaction network. **a.** The co-activation graph is a one-mode projection from the bipartite interaction network. Co-activation graph of either node type can be considered, but disjoint hubs (high-degree nodes) in one node type give rise to isolated cliques in the co-activation graph on the other node type. In this example case, almost disjoint odorant hubs (blue squares) are reflected in the almost isolated cliques in the co-activation graph of receptors (orange circles). **b.** Odorants are grouped consequently. **c.** A co-activation matrix, where each row or column is a receptor; receptor cliques appear as square blocks. Receptors are sorted such that cliques are ordered by the size; this is called the primary grouping. **d.** Odorants are again grouped consequently. **e.** Receptors are grouped once again by the redundancy, and re-sorted according to this secondary grouping. See text for details. **f-h.** Analysis on the odorant-receptor interaction dataset [14], binarized by counting all interactions. **f.** The co-activation graph from the real dataset. **g.** Sorted co-activation matrix for the 303 receptors. Diagonal blocks indicate the primary grouping of receptors. each around a shared hub odorant. **h.** Secondary grouping of receptors, based on the receptor code redundancy between the odorant hubs. We colored the 6 largest groups, which already span more than $250$ receptors cumulatively. The hub odorants for several large primary groups are labeled.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

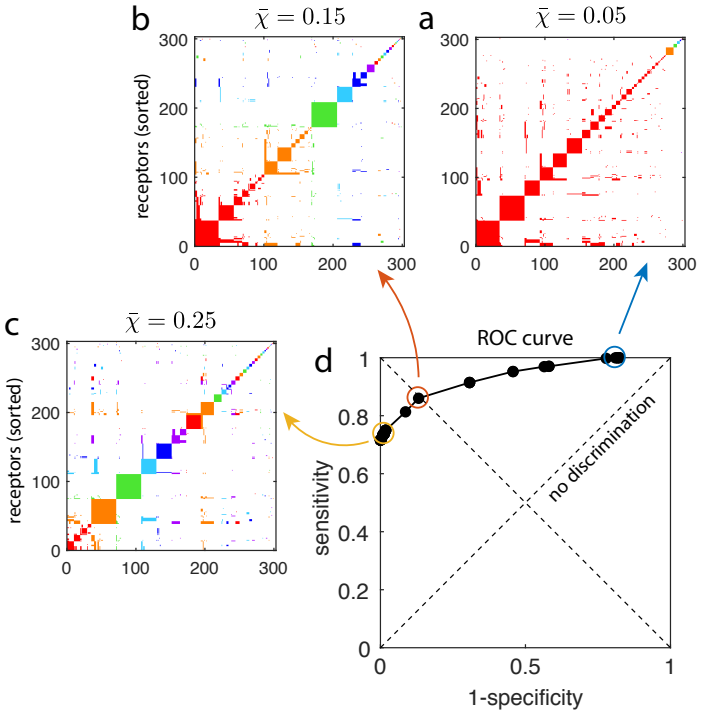


**Figure S5**: **Finding the optimal threshold $\bar{\chi}$ for the secondary grouping of receptors. a-c.** Secondary grouping results at different thresholds, $\bar{\chi} = 0.05$, $\bar{\chi} = 0.15$ and $\bar{\chi} = 0.25$ respectively. Each panel shows the co-activation matrix of receptors, sorted and colored by the corresponding secondary groups. **d.** The ROC curve of grouping at varying threshold $\bar{\chi}$. The two diagonals, the line of no discrimination which corresponds to random guesses (positive diagonal) and the line of balance points where sensitivity equals specificity (negative diagonal), are shown in dashed lines.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

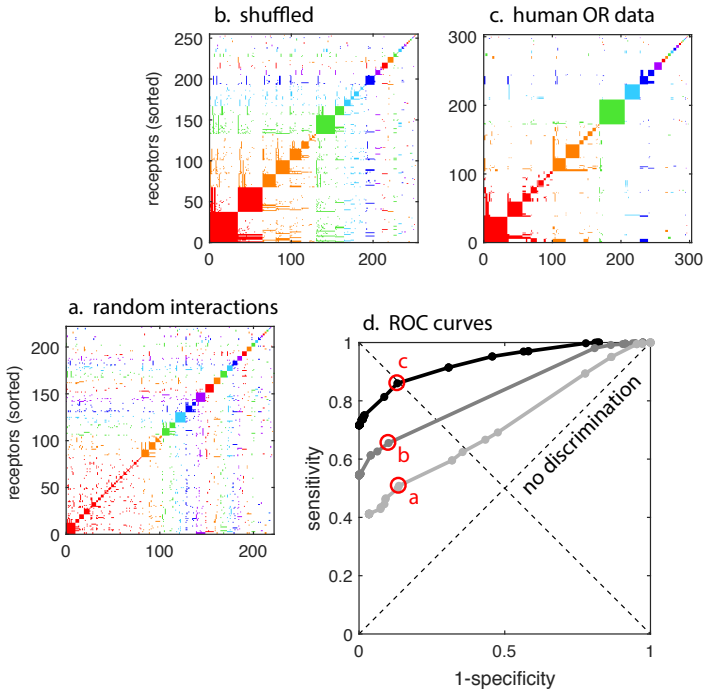


**Figure S6**: **No clear grouping by chance.** Best global partitioning results are shown for: **a.** Random network with fixed average degree for odorants. **b.** Network with shuffled receptor indices and preserved degree distribution. For a-b, the displayed numbers of receptors are smaller than the total $N = 303$, because we excluded receptors that do not respond to any odorant in the network. Non-responding receptors arise in the randomization process due to the overall sparsity of interactions. **c.** Best grouping result from the real data from human ORs, reprinted for comparison. **d.** Comparison of the prediction rates, plotted in the ROC space. Circles indicate the best partitioning results in the respective cases (a-c).

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

| Idx | Odorant | Degree | Group1 | Group2 | Descriptors |
|---|---|---|---|---|---|
| 1 | cis-3-hexen-1-ol | 36 | 1 | 1 | fresh **green** grassy foliage vegetable herbal oily |
| 2 | phenyl acetaldehyde | 10 | 1 | 1 | **green** sweet floral hyacinth clover honey cocoa |
| 3 | n-amyl acetate | 8 | 1 | 1 | ethereal **fruity** banana pear apple |
| 4 | amyl laurate | 7 | 1 | 1 | fatty **oily** waxy |
| 5 | 2-ethylfenchol | 4 | 1 | 1 | **earthy** humus rooty musty patchouli humus camphoreous |
| 6 | 3-methyl-2-hexenoic acid | 3 | 1 | 1 | **sweaty** |
| 7 | methyl salicylate | 2 | 1 | 1 | wintergreen **minty** |
| 8 | octyl aldehyde | 1 | 1 | 1 | **aldehydic** waxy citrus orange peel green fatty |
| 9 | pyrazine | 1 | 1 | 1 | **nutty** pungent sweet corn roasted hazelnut barley |
| 10 | geranyl acetate | 27 | 3 | 1 | **floral** rose lavender green waxy |
| 11 | sandalwood | 5 | 3 | 1 | sweet **woody** balsamic cashew terpenic herbal spicy |
| 12 | lilial | 5 | 3 | 1 | **floral** muguet watery green powdery cumin |
| 13 | Cyclohexanone | 1 | 3 | 1 | **minty** acetone |
| 14 | cinnamaldehyde | 20 | 7 | 1 | sweet **spicy** cinnamon warm |
| 15 | coumarin | 11 | 7 | 1 | sweet **tonka** new mown hay coumarinic |
| 16 | spearmint | 9 | 7 | 1 | **spearmint** |
| 17 | quinoline | 7 | 7 | 1 | **medicinal** musty tobacco rubbery earthy |
| 18 | 1-octanethiol | 6 | 7 | 1 | sulfurous |
| 19 | coffee difuran | 5 | 7 | 1 | sulfurous **coffee** roasted chicken meaty onion cabbage |
| 20 | thioglycolic acid | 3 | 7 | 1 | strong unpleasant odor characteristic of mercaptan |
| 21 | dicyclohexyl disulfide | 1 | 7 | 1 | **alliaceous** onion eggy nut skin coffee clam seafood |
| 22 | eugenol methyl ether | 15 | 10 | 1 | sweet fresh warm **spicy** clove carnation cinnamon |
| 23 | terpineol | 1 | 10 | 1 | pine **terpenic** lilac citrus woody floral |
| 24 | linalool | 11 | 12 | 1 | citrus **floral** sweet bois de rose woody green blueberry |
| 25 | (-)-carvone | 6 | 21 | 1 | sweet spearmint herbal **minty** |
| 26 | (+)-carvone | 5 | 21 | 1 | **minty** spicy bready caraway |
| 27 | r-carvone | 1 | 21 | 1 | sweet spearmint herbal **minty** |
| 28 | lyral | 5 | 23 | 1 | **floral** muguet cyclamen rhubarb woody |
| 29 | anisaldehyde | 1 | 25 | 1 | **anisic** sweet powdery mimosa floral hawthorn balsamic |
| 30 | methyl octanoate | 1 | 32 | 1 | **waxy** green sweet orange aldehydic vegetable herbal |
| 31 | butyric acid | 23 | 5 | 2 | sharp acetic **cheesy** buttery fruity |
| 32 | acetophenone | 2 | 5 | 2 | sweet pungent hawthorn mimosa almond acacia chemical |
| 33 | butyl acetate | 2 | 5 | 2 | **ethereal** solvent fruity banana |
| 34 | androstenone | 24 | 6 | 2 | **sweaty**, urinous, woody, or floral to some people |
| 35 | anise | 6 | 6 | 2 | fresh spicy sweet clean crushed-fruit |
| 36 | 3-phenyl propyl propionate | 4 | 6 | 2 | **floral** balsamic hyacinth mimosa fruity |
| 37 | 1-butanol | 1 | 6 | 2 | **fermented** fusel oily sweet balsamic whiskey |
| 38 | jasmine | 11 | 13 | 2 | **jasmin** floral |
| 39 | androstadienone | 14 | 14 | 2 | **sweaty** |
| 40 | heptanoic acid | 1 | 14 | 2 | rancid sour **cheesy** sweaty |
| 41 | p-cymene | 1 | 14 | 2 | fresh citrus **terpenic** woody spicy |
| 42 | 5,5-Dimethyl-1,3-cyclohexanedione | 1 | 14 | 2 | odorless or weak |
| 43 | 2-methyl-1-propanethiol | 1 | 14 | 2 | **sulfurous** cooked vegetable mustard |
| 44 | caramel furanone | 1 | 14 | 2 | sweet **caramellic** maple burnt sugar coffee |
| 45 | amyl butyrate | 1 | 14 | 2 | sweet **fruity** banana pineapple cherry tropical |

**Table S1**: **List of odorants with perceptual descriptors (1/2).** The 89 odorants are listed according to the groups identified in our work. The order of odorants is the same as in the row order in Fig. 4 in the main text. The columns "Group1" and "Group2" show the primary odorant group rank and the secondary odorant group rank, respectively, ranked by the size of the corresponding receptor groups. Descriptor sources are separately provided in Table S3.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

| Idx | Odorant | Degree | Group1 | Group2 | Descriptors |
|---|---|---|---|---|---|
| 46 | isoeugenol | 13 | 15 | 2 | sweet **spicy** clove woody carnation floral |
| 47 | ambrette | 8 | 16 | 2 | **musk** sweet ambrette seedy |
| 48 | (+)-dihydrocarvone | 3 | 30 | 2 | herbal **minty** mentholic |
| 49 | eugenol | 36 | 2 | 3 | sweet fresh warm **spicy** clove carnation cinnamon |
| 50 | TMT | 5 | 2 | 3 | fox odor |
| 51 | 2-methoxy-4-methylphenol | 4 | 2 | 3 | **spicy** clove vanilla phenolic medicinal leathery |
| 52 | butyl anthranilate | 4 | 2 | 3 | petitgrain plum sweet grape **fruity** winey floral berry powdery |
| 53 | r-limonene | 3 | 2 | 3 | **citrus** orange fresh sweet |
| 54 | 2-heptanone | 2 | 2 | 3 | **cheesy** fruity spicy sweet herbal - typical blue cheese character |
| 55 | (+)-menthol | 2 | 2 | 3 | **mentholic** cooling minty |
| 56 | nutmeg | 2 | 2 | 3 | **spicy** woody nutmeg powdery terpenic aromatic pine resinous |
| 57 | banana | 1 | 2 | 3 | banana **fruity** creamy tropical |
| 58 | guaiacol | 1 | 2 | 3 | **phenolic** smoky spicy vanilla woody |
| 59 | eugenol acetate | 23 | 4 | 4 | fresh sweet woody clove floral carnation malty **spicy** |
| 60 | caproic acid | 2 | 4 | 4 | sour **fatty** sweaty cheesy |
| 61 | pentadecalactone | 1 | 4 | 4 | **musk** animal powdery natural fruity |
| 62 | ethyl vanillin | 19 | 8 | 5 | sweet creamy **vanilla** caramellic |
| 63 | helional | 4 | 8 | 5 | **floral** watery fresh green ozone cyclamen hay |
| 64 | dihydrojasmone | 1 | 8 | 5 | **floral** fresh outdoors jasmin myrrh woody spicy herbal |
| 65 | hexyl octanoate | 1 | 8 | 5 | fruity **green** waxy berry apple estery |
| 66 | cinnamon | 14 | 11 | 5 | sweet cinnamon **spicy** warm woody |
| 67 | benzene | 1 | 11 | 5 | **aromatic** sweet gasoline-like |
| 68 | nonanedioic acid | 1 | 11 | 5 | **fatty** |
| 69 | pyridine | 1 | 11 | 5 | sour **fishy** ammoniacal, pleasant coffee or chocolate when dilute |
| 70 | isovaleric acid | 11 | 9 | 6 | sour stinky feet sweaty **cheesy** tropical |
| 71 | propionic acid | 1 | 9 | 6 | pungent **acidic** cheesy vinegar |
| 72 | beta-ionone | 1 | 9 | 6 | **floral** woody sweet fruity berry tropical beeswax |
| 73 | decyl aldehyde | 5 | 17 | 7 | sweet **aldehydic** waxy orange peel citrus floral |
| 74 | undecanal | 1 | 17 | 7 | waxy soapy floral **aldehydic** citrus green fatty cloth laundered cloth |
| 75 | geraniol | 9 | 18 | 8 | sweet **floral** fruity rose waxy citrus |
| 76 | ethylene brassylate | 8 | 19 | 9 | powdery sweet floral ambrette **musk** woody |
| 77 | vanillin | 8 | 20 | 10 | sweet creamy **vanilla** caramellic |
| 78 | (-)-menthol | 5 | 22 | 11 | **mentholic** cooling minty |
| 79 | allyl phenylacetate | 4 | 24 | 12 | **honey** fruity rummy |
| 80 | nonyl aldehyde | 3 | 29 | 13 | waxy **aldehydic** rose fresh orris orange peel fatty peely |
| 81 | 2-decenal | 4 | 27 | 14 | **fatty** orange rose aldehydic floral green |
| 82 | diacetyl | 4 | 28 | 15 | **buttery** sweet creamy pungent caramellic |
| 83 | 1-formylpiperidine | 1 | 28 | 15 | odorless |
| 84 | isobutyric acid | 2 | 31 | 16 | **acidic** sour cheesy dairy buttery rancid |
| 85 | galaxolide | 1 | 33 | 17 | sweet floral **musk** |
| 86 | (-)-citronellal | 1 | 43 | 22 | sweet dry **floral** herbal waxy aldehydic citrus |
| 87 | 1-octanol | 1 | 44 | 23 | **waxy** green orange aldehydic rose mushroom |
| 88 | nonanal | 1 | 45 | 24 | waxy **aldehydic** rose fresh orris orange peel fatty peely |
| 89 | r-limonene | 1 | 46 | 25 | **citrus** orange fresh sweet |

**Table S2**: **List of odorants with perceptual descriptors (2/2).** Continued from Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/525287; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

| RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Deposited data** | | |
| Pairwise odorant-receptor interaction | [14] | Data Record 3 (dose-response) |
| Information of olfactory receptors | [14] | Data Record 4 (receptors) |
| Information of odorants | [14] | Data Record 5 (odorants) |
| Pairwise interaction properties | [16] | S1 Table |
| Modular layout of interaction network | This paper | https://github.com/jihyunbak/ORnetwork |
| **Odor descriptors** | | |
| 3-methy-2-hexenoic acid | [54] | N/A |
| thioglycolic acid | NIOSH Pocket Guide to Chemical Hazards | https://www.cdc.gov/niosh/npg/npgd0610.html |
| androstenone | [43] | N/A |
| androstadienone | [55] | N/A |
| 5,5-Dimethyl-1,3-cyclohexanedione | Fisher Scientific | SDS: 5,5-Dimethyl-1,3-cyclohexanedione |
| | EMD Millipore Corporation | MSDS: Dimedone GR for analysis (reagent for aldehydes) |
| nonanedioic acid | Columbus Chemical Industries | SDS: Azaleic Acid |
| 1-formylpiperidine | EMD Millipore Corporation | MSDS: N-Formylpiperidine for synthesis |
| anise, pyridine | [56] | N/A |
| All other odorants in dataset | W. Luebke, The Good Scents Company Information System | http://www.thegoodscentscompany.com |

**Table S3**: **List of data sources.** SDS: Safety Data Sheet; MSDS: Material Safety Data Sheet.