Title: Shapley Fields Reveal Chemotopic Organization in the Mouse Olfactory Bulb Across Diverse Chemical Feature Sets.

Nikola Milicevic[1], Shawn D. Burton[2,3], Matt Wachowiak[2], Vladimir Itskov[1]

[1]Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania, United States
[2]Department of Neurobiology, University of Utah School of Medicine, Salt Lake City, Utah, United States
[3]Present address: Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania, United States

Abstract

Representations of chemical features in the neural activity of the olfactory bulb (OB) are not well-understood, unlike the neural code for stimuli of the other sensory modalities. This is because the space of olfactory stimuli lacks a natural coordinate system, and this significantly complicates characterizing neural receptive fields (tuning curves), analogous to those in the other sensory modalities. The degree to which olfactory tuning is spatially organized across the OB, often referred to as *chemotopy*, is also not well-understood. To advance our understanding of these aspects of olfactory coding, we introduce an interpretable method of *Shapley fields*, as an olfactory analog of retinotopic receptive fields. Shapley fields are spatial distributions of chemical feature importance for the tuning of OB glomeruli. We used this tool to investigate chemotopy in the OB with diverse sets of chemical features using widefield epifluorescence recordings of the mouse dorsal OB in response to stimuli across a wide range of the chemical space.

We found that Shapley fields reveal a weak chemotopic organization of the chemical feature sensitivity of dorsal OB glomeruli. This organization was consistent across animals and mostly agreed across very different chemical feature sets: (i) the expert-curated PubChem database features and (ii) features derived from a Graph Neural Network trained on human olfactory perceptual tasks. Moreover, we found that the principal components of the Shapley fields often corresponded to single commonly accepted chemical classification groups, that therefore could be “recovered” from the neural activity in the mouse OB. Our findings suggest that Shapley fields may serve as a chemical feature-agnostic method for investigating olfactory perception.

Introduction

Primary sensory areas of the mammalian brain, such as the visual and the auditory systems, are known to encode characteristic features of the stimuli, such as the wavelength of light and

frequency of sound. Furthermore, encoding of such stimulus features is spatially organized within topographic maps across primary and downstream cortical areas of visual, auditory, gustatory, and other sensory systems (Hubel & Wiesel, 1977; Kandler et al., 2009; Chen et al., 2011). In contrast to these sensory systems, however, how physicochemical features of olfactory stimuli are encoded within the olfactory system, including whether they are topographically organized within the olfactory bulb (OB) – the primary olfactory sensory area in the mammalian brain – remains poorly understood.

Axons of the olfactory sensory neurons (OSNs) expressing the same olfactory receptor (OR) gene converge to deliver sensory input from the periphery to stereotypically positioned glomeruli on the OB surface. Due to the large number of OR genes, the dimensionality of olfactory stimulus space is potentially quite large. However, dimensionality reduction has been found to account for odorant similarity in chemical, neuronal and perceptual spaces in humans (Koulakov et al., 2011; Keller & Vosshall, 2016), mice (Ma et al., 2012) and across species (Haddad et al., 2008, 2010; Saito et al., 2009). These results indicate that olfaction follows a combinatorial logic rather than a simple labeled-line code, similar to receptive field patterns in lower-dimensional sensory systems.

Several studies indicate that glomeruli in mammalian OBs tuned to similar combinations of odorant chemical features are also located in close proximity on the OB and form chemical-feature clusters (Mori et al., 2006; Burton et al., 2022). For example, in the rat OB, a homologous series of odorant molecules can evoke distinct but spatially clustered responses that change systematically with carbon chain length (Rubin & Katz, 1999; Johnson et al., 1999; Johnson & Leon, 2000; Meister & Bonhoeffer, 2001). Moreover, glomerular responses to certain odorant functional groups and hydrocarbon substructures also follow a systematic organization in the rodent OB (Uchida et al., 2000; Takahashi et al., 2004; Johnson et al., 2005; Johnson & Leon, 2007; Matsumoto et al., 2010; Soelter et al., 2020). These findings provide evidence for a topographic mapping of chemical features – or “chemotopy” – in glomeruli of the OB.

Several studies have cast doubt on the validity of the chemotopic organization, however. It was found in (Soucy et al., 2009; Chae et al., 2019; Ma et al., 2012) that commonly used physicochemical features were not systematically represented by glomerular activity in the mouse OB. Instead, (Ma et al., 2012) argued that there is no chemotopic representation of chemical features in either fine or broad scales, while there is still a significant correlation between pairwise tuning similarity and distances between glomeruli. In (Soucy et al., 2009; Chae et al., 2019; Burton et al., 2022) the authors argued for a functional ‘’mosaic” organization of glomerular activity.

It is natural to define ***weak chemotopy*** as the property by which glomeruli, tuned to chemical features in "close proximity" in chemical feature space, are also in a spatial proximity on the OB. We call this chemotopy “weak” in reference to the imprecise and overlapping nature of the topographic organization with respect to the chemical features. This is in contrast to retinotopy in the primary visual cortex, where the visual receptive fields are sharply localized. However, the nature of chemotopy in the OB seems to greatly depend on the choice of chemical features,

because different choices of chemical features significantly change the distances in the poorly understood chemical space. This dependence on the considered features may explain the inconsistencies of the previous studies of chemotopy.

Here, we set out to characterize chemotopy and its relationship to the choice of the chemical features that are being represented in the neural responses. To this end we developed a novel method of *Shapley fields*, as an olfactory analog of the retinotopic receptive fields. Shapley fields are distributions of chemical feature importance for the tuning of OB glomeruli. They are derived as an aggregate from Shapley values, a machine learning tool that allocates credit to features that explain model predictions (Lundberg & Lee, 2017), originally introduced in the context of game theory (Shapley, 1951). Another reason to use Shapley fields is that this tool automates feature selection and thus allows analysis of olfactory receptive properties using any significantly diverse chemical feature set. We used this tool to characterize chemotopy relative to diverse sets of chemical features using epifluorescence recordings of the mouse dorsal OB in response to stimuli across a wide range of odorants.

Our findings are consistent with a weak chemotopic organization of odor representations across OB glomeruli. We also established that our analyses do not depend on different choices of chemical feature sets: (i) the expert-curated PubChem chemical feature set and (ii) an artificial chemical feature set produced by a neural network, trained on human perception and molecular graphs. Both feature sets captured similar receptive properties in the mouse OB.

Results

Odorant-evoked responses were imaged with widefield epifluorescence across the dorsal surface of the OB in anesthetized mice, using artificial inhalation to ensure consistent odorant sampling (Figure 1A), see Methods and (Burton et al., 2022).

To evaluate the sensitivity of glomeruli to odorant chemical features, we trained a *Chemotuning model* using the PubChem binary molecular fingerprint features (Figure 1B), (Kim et al., 2023). The Chemotuning model assumes that each glomerulus activates based on a set of chemical features of a given odorant stimulus, irrespective of the glomerulus position in the OB. For each glomerulus, a random forest binary classifier was fit to predict activation of the glomerulus in response to the odorant, from the chemical features. More precisely, we trained a classifier to only detect activation or no activation. Our analysis did not consider whether a particular odorant activated a glomerulus more than another odorant (see Methods).

To understand the chemical receptive properties of glomeruli, we used Shapley values. Shapley values are a machine learning tool that enable allocation of credit to features explaining model predictions (Lundberg & Lee, 2017). This tool is inspired by classical work in game theory (Shapley, 1951); our usage is similar to that in (Zoabi et al., 2021), and (Rodríguez-Pérez & Bajorath, 2020). Shapley values are a general framework that unifies several other popular additive feature attribution methods such as LIME (Ribeiro et al., 2016), DeepLIFT (Shrikumar et al., 2017) and Layer-wise relevance propagation (Bach Sebastian AND Binder, 2015). We used mean absolute Shapley values, similar to (Lundberg & Lee, 2017) (see Methods), to

determine which chemical features are *most responsible* for predicting glomerular activation by any given odorant (Figure 1C).

Shapley fields reveal a stereotypical mosaic organization of chemical receptive fields within the olfactory bulbs. To understand the organization of OB chemical tuning, we constructed *Shapley fields.* Shapley fields represent signed distributions of feature credit allocations for prediction of glomerular activation, mapped on the surface of the OB. In other words, they are a signed projection of odorant chemical feature importance according to the Chemotuning model on the OB. Alternatively, a Shapley field of a chemical feature is the vector of Shapley values for each glomerulus. Shapley fields associated with four of the PubChem features are shown in Figure 2A.

To investigate the spatial organization of the feature importance on the surface of the OB, we computed the 20-dimensional PCA projections of Shapley fields (see Methods) and then performed a hierarchical clustering of glomeruli, based on the values of Shapley fields after the projection assigned to the glomeruli (Figure 2B). The resulting clusters revealed a mosaic organization of glomeruli across the left and the right OBs of an animal. This organization was inferred from the Shapley values of our Chemotuning model. This organization was also observed across animals (see SuppFig 1).

Notably, while the Chemotuning model does not consider glomerular positional data, spatial clustering by chemical feature responsivity nevertheless emerged. This is similar to the clustering previously reported by some studies (Mori et al., 2006; Johnson & Leon, 2007; Burton et al., 2022). This organization proved highly statistically significant (Figure 2C), as revealed by shuffling the identity of glomeruli (*N=1000 trials*) and computing spatial-autocorrelation statistics, Moran's I and Geary's C (see Methods) of the original clustering and the clustering after each shuffling. For the sake of clarity, we displayed 8 clusters, however any choice of number of clusters between 6 and 12 was statistically significant ($p<0.05$) with respect to autocorrelation scores of Moran's I and Geary's C (Figure 2C, SuppFig 1D).

The Chemotopy model is more predictive than the Chemotuning model. The Chemotuning model prediction does not take advantage of the position of glomeruli on the OB. If there is a spatial organization based on chemistry, then adding positional information as a feature for learning should improve the model's performance. To this end, we also developed the *Chemotopy model*. The Chemotopy model assumes that the response of a glomerulus is a function of *both* position and the chemical features of the odorant and predicts glomerular activation from both the chemical features of the odorant as well as the positional information of the glomerulus. The Chemotopy model is thus a *single* random forest model for predicting activation of all the observed glomeruli on the OB.

We compared the predictions of the Chemotuning and the Chemotopy models, as well as two different baseline models. Here we considered the baseline of predicting that the glomerulus would not be activated (because glomerular activation is sparse in the dataset) as well as the *peer prediction* model (see below). In order to have a relatively small number of model

parameters, we used the principal components of the PubChem features (“PCA-PubChem”), while ordering them according to their Shapley values (see Shapley rankings in Methods). The Chemotopy model outperformed the Chemotuning model when we used the Shapley ranking of the PCA-PubChem features. To provide more context to this comparison, we additionally considered a *peer prediction model* as the other baseline.

Our peer prediction model (see Methods) is inspired by (Harris et al., 2003; McKenzie et al., 2021, Luczak et al., 2022),where it has been observed that one could predict spiking of individual neurons based on the activity of other neurons. In our data, the peer prediction model provided high prediction accuracy (Figure 3) because the olfactory representations are likely redundant for the stimuli that we considered, as should be expected from combinatorial coding of olfactory stimuli in the OB (Malnic et al., 1999; A. Koulakov et al., 2006; Grabe & Sachse, 2018; de March et al., 2020). We observed that approximately 70-80% of the difference between the peer prediction model’s quality of prediction and the baseline is accounted for by the chemical feature sets that we used. Of note, our odorant sample size is rather small (n=59) as compared to the number of recorded glomeruli in a single OB (m=113,108,128,132). Thus, we expect the quality of prediction of Chemotopy and Chemotuning models to catch up to the peer prediction model’s performance with a significantly larger sample of odorants.

Given the previously discussed discrepancies in reported results, which may arise from the differing chemical feature sets used, we sought to explore whether the superior performance of the Chemotopy versus Chemotuning models was influenced by the specific choice of chemical features. To investigate this, we also did our analysis using a second set of chemical features. We generated these features from a graph neural network (GNN) trained on human derived odor labels of approximately 6000 odorant molecules from the GoodScents company (The Good Scents Company, 2023) (see Methods). The Chemotopy model significantly outperformed the Chemotuning model, with the accuracy gap widening as more Shapley-ranked features were included (Figure 3). Notably, the Chemotopy model's prediction accuracy peaked at just five features, whereas the Chemotuning model showed no clear peak. Here, the Shapley values served as a dimensionality reduction technique that improved the quality of prediction by only considering the most relevant features.

We also found that GNN features do *not* outperform the PCA-PubChem feature set with respect to predicting glomerular activation (Figure 3). This finding contrasts with recent works (Achebouche et al., 2022; Qian et al., 2023) where GNN features have been reported to outperform other feature sets such as RDKit (Landrum, 2021) across different olfaction datasets and olfaction tasks (across species). Comparing models trained on GNN features with models trained on PCA-PubChem, we see that PCA-PubChem has more predictive power than the GNN features early on. However, as more features get added, GNN features match PCA-PubChem predictive power both in the Chemotopy and Chemotuning models (Figure 3). These same observations hold for the OBs of animals 2 and 3 (SuppFig 2).

Chemical features derived from human perception yield predictions and mosaic chemotopic organization that is similar to those derived from the PubChem chemical features.

How do the chemical receptive properties of the glomeruli depend on the choice of the chemical feature sets? To answer this question, we first computed the Shapley fields of GNN features (Figure 4A). The first principal component of the GNN Shapley fields was highly correlated to the glomerular carboxylic acid sensitivity vector, which counts how many acids each glomerulus responded to (between 0 and 5 in the stimulus set). Tuning to acids was also correlated to several principal components of the PubChem Shapley fields, albeit to a lesser extent (Figure 4B). Note that these are the Shapley fields of PubChem features, and not PCA-PubChem features.

Pearson correlation analysis also revealed significant correlations among the first ten principal components of both PubChem and GNN Shapley fields (Figure 4C). To validate the apparent correlations between multiple principal components of GNN and PubChem Shapley fields in Figure 4C, we computed the chordal distances between subspaces generated by up to *k=20* principal components of each chemical feature set (see Methods). The chordal distances between the subspaces of different chemical features are significantly lower than expected from random controls (SuppFig 4). Moreover, we found that 20 principal components of the Shapley fields captured 70% of variance for the PubChem feature set and 50% of variance for the GNN feature set respectively (Figure 4D). This suggests that the Shapley fields represent approximately the same chemical sensitivity irrespective of very different choices of the chemical features that we used: (i) a collection of the chemical properties, and (ii) features derived from human perception and molecular graphs. It is thus likely that Shapley fields capture roughly the same information about receptive properties of the OB, from *any diverse enough* set of chemical features.

Principal components of the Shapley values recover major chemical classes.

The n=59 odorants that we used can be organized into 13 chemical classes, as defined by major substructural motifs (see SuppFig 6 for exact class assignment). To understand how the Shapley fields relate to this common chemical classification of odorants, we extended our earlier analysis (Figure 4A,B) to examine how the principal components of Shapley fields correlated with glomerular responses to all other odorant classes, by considering the odorant class sensitivity vectors. An odorant class sensitivity vector represents the sensitivity of all the recorded glomeruli to a given odorant class (see Methods). Surprisingly, we found that the individual principal components of GNN Shapley fields are significantly correlated (see Methods) with only one (8 out of 10 first PCs) or two (one out of 10 first PCs) odorant class sensitivity vectors ($p<0.05$) (Figure 4E). This is not an expected property of the Shapley fields, as GNN features were derived from a graph neural network trained on human perception, combined with molecular graph, yet the first ten principal components of the Shapley fields on the mouse OB often corresponded to a *single* “common chemical class” used in the set of stimuli. The same analyses applied to the PubChem Shapley fields proved less related to individual odorant chemical classes (Figure 4F), with 5 out of 10 first PCs correlated with just one chemical class. See Supplementary Table 1 for the actual correlation values. Note that the

PubChem features show an approximate clustering of the odorant classes used in our experiment, whereas there is barely any clustering by odorant classes in the GNN feature space (SuppFig 7). Surprisingly, the GNN Shapley fields showed more pronounced one-to-one correspondence between principal components and chemical classes as compared to the PubChem Shapley fields, despite the lack of significant clustering of the GNN chemical features.

While there was no significant overperformance of the GNN-based prediction over the PubChem feature-based predictions (Figure 3), it seems that the principal components of the GNN Shapley fields are more consistent with the commonly used chemical classification. This raises the possibility that the human descriptors like “spicy” or “fruity” that were used in training the GNN model, *together* with the mouse OR responses and the molecular graphs used in the GNN, effectively yield the common classification of the odorant molecules (SuppFig 6). We speculate that this is because mouse olfactory perception is still sufficiently close to human olfactory perception despite the extensive evolutionary divergence in the olfactory systems of these two species.

Discussion

Motivated by investigating odorant chemical feature tuning in the OB, we introduced an application of Shapley fields to explore the sensitivity of glomeruli to chemical features as a function of their position on the OB surface. The Shapley fields serve as an olfactory analogue of receptive fields in vision and other sensory areas. While chemical response spectra define receptive fields in the odorant space, they do not correspond to receptive fields in the chemical feature space, which Shapley fields aim to characterize as a more nuanced representation of glomerular sensitivity. Using two different computational approaches, we confirmed that the chemical receptive properties of glomeruli in the OB are consistent with a weak chemotopy in the organization of odorant representations across the OB surface. This agrees with earlier studies in (Rubin & Katz, 1999; Uchida et al., 2000; Johnson et al., 1999, 2005; Johnson & Leon, 2000, 2007; Meister & Bonhoeffer, 2001; Takahashi et al., 2004; Matsumoto et al., 2010; Soelter et al., 2020; Burton et al., 2022) and at least partially addresses the concerns raised in later works (Soucy et al., 2009; Chae et al., 2019; Ma et al., 2012).

To investigate how our conclusions are influenced by inherently different choices of the chemical feature sets, we repeated our analyses using an artificially constructed chemical feature set that combined molecular graph information with human perceptual data, using methods developed in (Achebouche et al., 2022; Qian et al., 2023). We found that the outperformance of the Chemotopy vs. Chemotuning model does not depend on the choice of chemical feature sets, and that both chemical feature sets describe the receptive properties of the OB glomeruli with comparable accuracy. Note that unlike the findings in (Qian et al., 2023), the predictions based on the GNN features were *not* significantly more accurate than those based on the PubChem features.

Furthermore, the Shapley fields derived from the two feature sets considered were strongly correlated, suggesting that the wide array of chemical features in PubChem and chemical

features derived from human perception cover similar properties of mouse olfactory perception, as represented by the neural activity in the dorsal OB. This suggests that if one chooses a sufficiently diverse set of chemical features, the corresponding Shapley fields would capture the same receptive properties of the OB glomeruli and thus may provide a feature set-agnostic tool for investigating olfactory perception.

We also found that each principal component of the Shapley fields often significantly correlates with a single commonly accepted chemical class used in the presented stimuli. This suggests that these chemical classes could have been “discovered” based on the mouse olfactory responses.

Materials and Methods

Data collection: Data were collected from 2 male and 1 female compound heterozygous crosses of OMP-IRES-Cre (Jackson Laboratory stock #006668) and Rosa26-CAG-lox-STOP-lox-GCaMP6f (RCL-GCaMP6f; Ai95D; Jackson Laboratory stock #024105) mouse lines, aged 16 - 18 weeks. Prior to data collection mice were housed up to 5 per cage on a 12-hour light/dark cycle with food and water available ad libitum.

Odorant responses were collected with epifluorescence imaging in isoflurane-anesthetized mice outfitted with a double tracheotomy to decouple respiration and anesthetic delivery from odorant presentation and sampling, using procedures identical to those described in (Burton et al., 2019). All procedures were performed following the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the University of Utah Institutional Animal Care and Use Committee.

Olfactometry: Odorants were obtained from Sigma-Aldrich, TCI America, Bedoukian Research, or ICN Biomedicals and delivered using a custom olfactometer that allowed mixing of odorant vapor in free space prior and computer-controlled timing and choice of odorant identity (Burton et al., 2019). Odorant delivery parameters were identical to those used in (Burton et al., 2022) except that the end-stage eductor was omitted. Odorants were pre-diluted in caprylic/capric medium chain triglyceride oil (C3465, Spectrum Chemical Mfg. Corp.) as described in (Burton et al., 2022) using dilutions ranging from 1:10 - 1:50 in order to achieve estimated delivered concentrations ranging from 1 - 1000 nM (median: 100 nM, interquartile range: 20 - 600 nM).

Data processing: Odorant response matrices (odorant x glomerulus) were generated using the analysis pipeline described in (Burton et al., 2022). Briefly, averaged ΔF response maps were generated from 4 - 6 repeated presentations of each odorant, after which regions of interest

(ROIs) were selected manually using custom visualization software written in MATLAB (https://github.com/WachowiakLab). ROIs representing presumed individual glomeruli from the dorsal OB were segmented manually, with the segmenter blind to odorant identity (Burton et al., 2022). Note that ROIs were drawn either following the rough contours of activation or with stereotyped-sized squares roughly centered within the contours of activation. After segmentation, response matrices were generated using the mean ΔF value across all pixels in an ROI, for all 59 tested odorants.

Random Forests: The random forest binary classifier models were trained using the scikit-learn python library (Pedregosa et al., 2011). The hyperparameters we were optimizing for were n_estimators (the number of trees in the forest), max_features (the number of features to consider when looking for the best split) and max_samples (the number of samples to draw from data to train each base estimator). Every other hyperparameter was chosen to be the default one given by the random forest classifier constructor, except the class_weight hyperparameter. Since we are dealing with heavily imbalanced datasets, the class_weight was set to “balanced_subsample”. Model performances were measured with the AUROC (Area Under Receiver Operating Characteristic curve). The cross-validation details are described in the relevant model descriptions.

Chemotuning model: The Chemotuning model predicts glomerular activation from the chemical features of the odorants alone. More specifically, given a glomerulus-odorant pair (g, o), a random forest binary classifier predicts activation of g to o based on the chemical features of o. Thus, the Chemotuning model is a collection of random forest models predicting activation, one for each glomerulus on the OB. The hyperparameters for the RF that were optimized with respect to running time and performance were: n_estimators=128, max_features=”sqrt”, max_samples=0.5, class_weight=“balanced_subsample”

Chemotopy model: The Chemotopy model predicts glomerular activation from the chemical features of the odorants and position information about the glomeruli. More specifically, given a glomerulus-odorant pair (g, o), a random forest binary classifier predicts activation of g to o based on the chemical features of o, and the position of g. Thus, the Chemotopy model is a single random forest model for predicting activation of all the observed glomeruli. To compare with the Chemotuning model (see below) we added an additional feature, which is a random integer identity for each glomerulus. That is, we wanted to make sure the models are trained on the feature space of the same dimension and then compared. The prediction is cross validated across 10 folds and the AUROC metric is used to evaluate performance. The hyperparameters for the random forest that were optimized with respect to running time and performance were as follows: n_estimators=256, max_features=1.0, max_samples=0.7, class_weight=”balanced_subsample”

Chemotuning control model: The Chemotopy control model takes the same data as the Chemotopy model however the positional information about glomeruli is randomly shuffled and a random number is assigned to each glomerulus to keep track of its identity as an additional feature. The model is cross validated in a leave-one-out fashion and the AUROC metric is used to evaluate performance. The hyperparameters for this model as the same as the hyperparameters for the Chemotopy model.

Peer prediction model: The peer prediction model predicts glomerular activation from the activation of its peers (the other glomeruli on the OB). More specifically, given a glomerulus-odorant pair (g, o), a random forest binary classifier predicts activation of g to o using as features the activations of all glomeruli except g, to o. Thus, the peer prediction model is a collection of random forest models predicting activation, one for each glomerulus on the OB. The model is cross validated in a leave-one-out fashion and the AUROC metric is used to measure performance. The hyperparameters for the random forest were optimized with respect to running time and performance and were the same as the ones for the Chemotuning model.

Position of glomeruli: Each glomerulus was approximated by a minimal rectangle that encloses the region of interest (ROI) associated to the glomerulus. The center of the rectangle was taken to be the position of the glomerulus. For each OB, the x and y coordinates of the glomeruli were normalized to both have mean 0 and variance 1.

The PubChem Substructure Fingerprint: The PubChem System generates a binary substructure fingerprint for chemical structures (Kim et al., 2023). When considering unique and nonzero PubChem fingerprints for our datasets we ended up with 140 features for the 59 odorants dataset and 132 features for the 52 odorants dataset.

PCA PubChem features: PCA dimensionality reduction was performed on the PubChem Substructure Fingerprint binary features with 59 and 52 odorants. Both were projected to 40 dimensions. The first 40 principal components accounted for around 90% of the variance in both datasets.

GNN features: We used a dataset of 5955 small molecules that were made publicly available, after pre-processing, by the authors in (Achebouche et al., 2022). Each molecule comes with a set of human perceptual odor labels. Originally these molecules, with their odor labels, were obtained from The Good Scents Company (TGSC) Database (The Good Scents Company, 2023), and Leffingwell Database (Leffingwell & Associates, n.d.). The 160 odor labels were grouped into 21 categories and used for training a multitask binary classification Graph Neural Network (GNN) (Scarselli et al., 2009; Zhou et al., 2020). We trained a multitask binary classification GNN using the DeepChem library (Ramsundar Bharath et al., 2019) and the same network architecture as the one specified in (Achebouche et al., 2022). Furthermore, we cross validated our GNN model across 5 folds and evaluated its performance with the AUPRC (Area Under Precision-Recall Curve) score. Our mean AUPRC score of approximately 0.40 agreed with the reported AUPRC score in (Achebouche et al., 2022). We then trained a network by

using the whole human perceptual dataset of 5955 odorants. Inspired by the idea of transfer learning (Tan et al., 2018), we produced learned GNN features for our smaller dataset of odorants. That is, we used the learned weights in the network to human-derived learned features for our odorants, by considering their representations in the *penultimate layer* of the network, before the multitask softmax classification layer. We call these the GNN features. There were 256 GNN features in total. All the odorants in our datasets were present in the human perceptual dataset, except for 7; 6 of the 7 amines and methyl salicylate.

Shapley values: Shapley values are a game theoretic approach to explain the outputs of machine learning models and we used the SHAP (SHapley Additive exPlanations) python library to compute them (Lundberg & Lee, 2017). Shapley values calculate the optimal credit allocation to features which locally explains the model outputs (Figure 1C). This is done by treating features as players in a cooperative game, and the model output as the game prize and then calculating the classic Shapley values from game theory (Shapley, 1951). For an arbitrary model, computing Shapley values has exponential time complexity and in practice Shapley values are only approximated. Instead, we used shap.TreeExplainer which uses dynamic programming to exactly compute Shapley values for tree-based models, in polynomial time (Lundberg et al., 2020).

Shapley rankings: Because Shapley values are associated with a given model, and the Chemotuning model is in fact a collection of individual models, one for each glomerulus, we ranked the chemical features using the Chemotopy model. Let $F = \{f_i\}_{i=1}^{N}$ be any of the sets of chemical features that were described above. For a given an OB, we trained a Chemotopy random forest model, M, on every glomerulus-odorant pair (g, o) (no test/train split, no cross-validation) that predicted glomerular activation based on the position of g and chemical features of o. Let f_i denote a chemical feature in F, and $\varphi_i(g, o)$ denote the Shapley value of f_i, which gives the contribution of $f_i(o)$ towards the prediction $M(g, o)$. The global Shapley value of f_i is the mean absolute value of Shapley values of f_i, over all observations (g, o), that is

$$\Phi_i = \frac{\sum_{g,o} |\varphi_i(g, o)|}{|G||O|},$$

where $|G|$ is the number of glomeruli on the OB, and $|O|$ is the number of odorants. We thus have a collection $\{\Phi_i\}_{i=1}^{N}$ of nonnegative numbers that we rank in decreasing order. We call these ranks the *Shapley rankings* of the set of chemical features F. Note that we did not use a cross-validated training here, because we observed that the ranks didn't change significantly, and the cross-validation procedure used for all the analyses in Figure 3 already insured that no overfitting had been possible.

Due to the randomness inherent to random forest models, a natural concern is the stability of these ranks. However, we found the Shapley ranks are for the most part stable with respect to the random choices of random forest models (SuppFig 5). The Shapley ranks of PubChem features were more stable than the ranks of GNN features. This is most likely due to GNN being

a larger feature set (256 dimensions) as opposed to the PubChem features (140 or 132 dimensions) and hence there was more possibility of ranks changing during random forest training. The stability analysis was done by considering the Kendall rank correlation coefficients (Kendall, 1938) of Shapley rankings from different random seeds.

Shapley fields: Given an OB, and a set of chemical features $F = \{f_i\}_{i=1}^{N}$, for each glomerulus g we train a Chemotuning model, T_g, using the chemical features F (no train/test split, no cross-validation). Let $\varphi_i(g, o)$ be the Shapley value of f_i, which gives the contribution of $f_i(o)$ towards the prediction $T_g(o)$. Let O_g^+, O_g^- be the collections of odorants for which $\varphi_i(g, o) \geq 0, \varphi_i(g, o) < 0$, respectively. Let

$$\Phi_i^+(g) = \frac{1}{|O_g^+|} \sum_{o \in O_g^+} \varphi_i(g, 0), \Phi_i^-(g) = \frac{1}{|O_g^-|} \sum_{o \in O_g^-} \varphi_i(g, 0),$$

where $|O_g^+|, |O_g^-|$ are the numbers of odorants in O_g^+ and O_g^-, respectively. This construction yields, for each feature f_i, two distributions on the OB which we call the *positive and negative Shapley fields* associated to f_i. Intuitively, the Shapley fields represent the distribution of feature importance on the OB, very much analogous to receptive fields; one can also think of them as signed measures on the OB. We could have done a cross-validated training and made sure we add positive and negative Shapley values appropriately from different testing folds, and then define positive and negative Shapley fields. However, we observed that the values of the Shapley fields don't change significantly and thus the additional computation time, which is not insignificant, would yield diminishing returns.

Hierarchical clustering of Shapley fields: We concatenate the positive and negative Shapley fields, calculated from a fixed set of chemical features, from different OBs. Given the concatenated Shapley fields $\Phi_i^+(g), \Phi_i^-(g)$, we perform a principal component dimensionality reduction and project to the first 20 principal components, which accounted for 70% (PubChem) or 50% (GNN) of the variance, see Figure 4D and SuppFig 3D. After the projection, we perform hierarchical ward linkage clustering with the cosine distance. The resulting dendrogram (Figure 2B and SuppFig 1B) and a height parameter are used to assign glomeruli to clusters.

Odorant class sensitivity vectors: For each odorant class in our dataset (13 odorant classes in total) we define its *odorant class sensitivity vector*. It is a vector on the OB, meaning its coordinates are indexed by the glomeruli on the OB, and it counts to how many odorants from the given odorant class a given glomerulus was sensitive to. For example, if the odorant class in question is the acids, then the acid sensitivity vector counts for each glomerulus on the OB, to how many acids the glomerulus responded to. Since we used 5 acids in our experiments, this vector assigns values between 0 and 5 to each glomerulus.

Correlations of Shapley fields principal components to odorant class sensitivity vectors:

A principal component of Shapley fields from a chemical feature set is represented by a vector whose indices correspond to the glomeruli on the OB. This vector therefore lives in an *n*-dimensional space, where *n* is the number of glomeruli. The 13 odorant classes sensitivity vectors are also vectors in this *n*-dimensional space, and they span a 13-dimensional linear subspace. We measured the correlation between a principal component of Shapley fields and its projection to this linear subspace. The projection is a linear combination of odorant class tuning vectors with some weights. We calculated the statistical significance of these weights (*p<0.05*) by randomly permutating the odorant identities that define the odorant class sensitivity vectors (N=10000 trials). The statistically significant coefficients gave us the odorant classes' tuning vectors the principal component of Shapley fields is correlated to.

Distances between linear subspaces of Shapley fields: To compare Shapley fields induced by different chemical feature sets, we compared the linear subspaces spanned by k principal components they generate, where $1 \leq k \leq 20$. Here, each principal component is a high-dimensional vector whose coordinates are its values on each of the glomeruli (see Figure 4A). These k-dimensional subspaces can be represented as a matrix whose columns are the principal components. Given Shapley fields from different chemical feature sets we can then compare distances between these resulting linear subspaces, as we vary the PCA dimension *k* we are projecting to. The distances between these subspaces are calculated via principal angles, described below.

Principal angles between subspaces: Angles between vectors in an n-dimensional space can be determined from the cosine-dot-product formula, i.e., if θ is the vector between angles x and y we have $\cos\theta = \frac{x \cdot y}{\|x\|\|y\|}$. Treating vectors as the one-dimensional subspaces they span; this angle represents the angle between these one-dimensional subspaces. This idea extends to higher dimensions, and it is possible to define so-called principal angles between linear subspaces of any dimension. Specifically, given real-valued (n x k) matrices *A* and *B*, each one with pairwise orthogonal column vectors of length 1, one can find k principal angles $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$ between the linear subspaces that are spanned by the column vectors of A and B. In practice the principal angles can be computed from the singular values of the matrix $A^T B$ (Björck & Golub, 1973). That is given σ_i, the i-th singular value of $A^T B$, we define $\cos\theta_i = \sigma_i$.

Chordal distance between subspaces: Given linear subspaces V and W of dimension p, let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$ be the principal angles between V and W. The chordal distance between V and W is the quantity $d_c(V, W) = (\sum_{i=1}^{p} \sin^2\theta_i)^{\frac{1}{2}}$, (Conway et al., 1996). If we write A and B for matrices of column vectors of orthonormal basis vectors that generate V and W, respectively, by definition $\cos^2\theta_i = 1 - \sin^2\theta_i = \sigma_i^2$, where σ_i is the i-th singular value of $C = A^T B$. In particular, $\sum_i \sigma_i^2 = trace(C^T C)$. From this it follows that $d_c(V, W) = (p - trace(C^T C))^{\frac{1}{2}}$.

Spatial autocorrelation metrics:
Let N be the number of spatial units indexed by i and j ; x is the variable of interest; $\bar{x}$ is the mean of x; $\bar{x}_i$ is the difference $x_i - \bar{x}$; w_{ij} is a matrix of spatial weights with zeroes on the

diagonal; and W the sum of all w_{ij}. Then the Moran's I and Geary's C spatial autocorrelations are defined as:

Moran's I: Moran's I is a measure of spatial autocorrelation that is given by the formula:

$$I = \frac{N}{W \sum_{i=1}^{N} \bar{x}_i^2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \bar{x}_i \, \bar{x}_j.$$

Geary's C: Geary's C is a measure of spatial autocorrelation that is given by the formula:

$$C = \frac{N-1}{2W \sum_{i=1}^{N} \bar{x}_i^2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (x_i - x_j)^2.$$

ACKNOWLEDGMENTS: The authors would like to thank Michael Schmuker for helpful discussions and suggestions. This work was supported by the NSF Next Generation Networks for Neuroscience Program, award 2014217, NIH grant F32MH115448, NSF grant 1555919, and NIH NINDS grant NS109979.

References

Achebouche, R., Tromelin, A., Audouze, K., & Taboureau, O. (2022). Application of artificial intelligence to decode the relationships between smell, olfactory receptors and small molecules. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-23176-y

Bach Sebastian AND Binder, A. A. N. D. M. G. A. N. D. K. F. A. N. D. M. K.-R. A. N. D. S. W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, *10*(7), 1–46. https://doi.org/10.1371/journal.pone.0130140

Björck, ke, & Golub, G. H. (1973). Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, *27*(123). https://doi.org/10.1090/s0025-5718-1973-0348991-3

Burton, S. D., Brown, A., Eiting, T. P., Youngstrom, I. A., Rust, T. C., Schmuker, M., & Wachowiak, M. (2022). Mapping odorant sensitivities reveals a sparse but structured representation of olfactory chemical space by sensory input to the mouse olfactory bulb. *ELife*, *11*. https://doi.org/10.7554/ELIFE.80470

Burton, S. D., Wipfel, M., Guo, M., Eiting, T. P., & Wachowiak, M. (2019). A Novel Olfactometer for Efficient and Flexible Odorant Delivery. *Chemical Senses*, *44*(3), 173–188. https://doi.org/10.1093/chemse/bjz005

Chae, H., Kepple, D. R., Bast, W. G., Murthy, V. N., Koulakov, A. A., & Albeanu, D. F. (2019). Mosaic representations of odors in the input and output layers of the mouse olfactory bulb. *Nature Neuroscience*, *22*(8). https://doi.org/10.1038/s41593-019-0442-z

Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. P., & Zuker, C. S. (2011). A gustotopic map of taste qualities in the mammalian brain. *Science*, *333*(6047). https://doi.org/10.1126/science.1204076

Conway, J. H., Hardin, R. H., & Sloane, N. J. A. (1996). Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental Mathematics*, *5*(2). https://doi.org/10.1080/10586458.1996.10504585

de March, C. A., Titlow, W. B., Sengoku, T., Breheny, P., Matsunami, H., & McClintock, T. S. (2020). Modulation of the combinatorial code of odorant receptor response patterns in odorant mixtures. *Molecular and Cellular Neuroscience*, *104*, 103469. https://doi.org/https://doi.org/10.1016/j.mcn.2020.103469

Grabe, V., & Sachse, S. (2018). Fundamental principles of the olfactory code. *Biosystems*, *164*, 94–101. https://doi.org/https://doi.org/10.1016/j.biosystems.2017.10.010

Haddad, R., Khan, R., Takahashi, Y. K., Mori, K., Harel, D., & Sobel, N. (2008). A metric for odorant comparison. *Nature Methods*, *5*(5). https://doi.org/10.1038/nmeth.1197

Haddad, R., Weiss, T., Khan, R., Nadler, B., Mandairon, N., Bensafi, M., Schneidman, E., & Sobel, N. (2010). Global features of neural activity in the olfactory system form a parallel code that predicts olfactory behavior and perception. *Journal of Neuroscience*, *30*(27). https://doi.org/10.1523/JNEUROSCI.0398-10.2010

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, *424*(6948), 552–556. https://doi.org/10.1038/nature01834

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture - Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *198*(1130), 1–59.

Johnson, B. A., Farahbod, H., & Leon, M. (2005). Interactions between odorant functional group and hydrocarbon structure influence activity in glomerular response modules in the rat olfactory bulb. *Journal of Comparative Neurology*, *483*(2). https://doi.org/10.1002/cne.20409

Johnson, B. A., & Leon, M. (2000). Odorant molecular length: One aspect of the olfactory code. *Journal of Comparative Neurology*, *426*(2). https://doi.org/10.1002/1096-9861(20001016)426:2<330::AID-CNE12>3.0.CO;2-5

Johnson, B. A., & Leon, M. (2007). Chemotopic odorant coding in a mammalian olfactory system. In *Journal of Comparative Neurology* (Vol. 503, Issue 1). https://doi.org/10.1002/cne.21396

Johnson, B. A., Woo, C. C., Hingco, E. E., Pham, K. L., & Leon, M. (1999). Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. *Journal of Comparative Neurology*, *409*(4). https://doi.org/10.1002/(SICI)1096-9861(19990712)409:4<529::AID-CNE2>3.0.CO;2-N

Kandler, K., Clause, A., & Noh, J. (2009). Tonotopic reorganization of developing auditory brainstem circuits. In *Nature Neuroscience* (Vol. 12, Issue 6). https://doi.org/10.1038/nn.2332

Keller, A., & Vosshall, L. B. (2016). Olfactory perception of chemically diverse molecules. *BMC Neuroscience*, *17*(1). https://doi.org/10.1186/s12868-016-0287-2

Kendall, M. G. (1938). *A New Measure of Rank Correlation* (Vol. 30, Issue 1). https://about.jstor.org/terms

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, *51*(D1). https://doi.org/10.1093/nar/gkac956

Koulakov, A. A., Kolterman, B. E., Enikolopov, A. G., & Rinberg, D. (2011). In search of the structure of human olfactory space. *Frontiers in Systems Neuroscience*, *SEPTEMBER 2011*. https://doi.org/10.3389/fnsys.2011.00065

Koulakov, A., Gelperin, A., & Rinberg, D. (2006). *Combinatorial on/off model for olfactory coding*.

Landrum, G. (2021). RDKit: Open-Source Cheminformatics Software. In *Http://Www.Rdkit.Org/*.

Leffingwell & Associates. (n.d.). *Flavor-Base* (9th ed.).

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1). https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-December*.

Ma, L., Qiu, Q., Gradwohl, S., Scott, A., Yu, E. Q., Alexander, R., Wiegraebe, W., & Yu, C. R. (2012). Distributed representation of chemical features and tunotopic organization of glomeruli in the mouse olfactory bulb. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14). https://doi.org/10.1073/pnas.1117491109

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial Receptor Codes for Odors. *Cell*, *96*(5), 713–723. https://doi.org/10.1016/S0092-8674(00)80581-4

Matsumoto, H., Kobayakawa, K., Kobayakawa, R., Tashiro, T., Mori, K., Sakano, H., & Mori, K. (2010). Spatial arrangement of glomerular molecular-feature clusters in the odorant-receptor class domains of the mouse olfactory bulb. *Journal of Neurophysiology*, *103*(6). https://doi.org/10.1152/jn.00035.2010

McKenzie, S., Huszár, R., English, D. F., Kim, K., Christensen, F., Yoon, E., & Buzsáki, G. (2021). Preexisting hippocampal network dynamics constrain optogenetically induced place fields. *Neuron*, *109*(6), 1040-1054.e7. https://doi.org/https://doi.org/10.1016/j.neuron.2021.01.011

Meister, M., & Bonhoeffer, T. (2001). Tuning and topography in an odor map on the rat olfactory bulb. *Journal of Neuroscience*, *21*(4). https://doi.org/10.1523/jneurosci.21-04-01351.2001

Mori, K., Takahashi, Y. K., Igarashi, K. M., & Yamaguchi, M. (2006). *Maps of Odorant Molecular Features in the Mammalian Olfactory Bulb*. https://doi.org/10.1152/physrev.00021.2005.-The

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Qian, W. W., Wei, J. N., Sanchez-Lengeling, B., Lee, B. K., Luo, Y., Vlot, M., Dechering, K., Peng, J., Gerkin, R. C., & Wiltschko, A. B. (2023). Metabolic activity organizes olfactory representations. *ELife*, *12*. https://doi.org/10.7554/eLife.82502

Ramsundar Bharath, Peter Eastman, Walters Patrick, Pande Vijay, Leswing Karl, & Wu Zhenqin. (2019). *Deep Learning for the Life Sciences* (1st ed.). O'Reilly Media.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rubin, B. D., & Katz, L. C. (1999). Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron*, *23*(3). https://doi.org/10.1016/S0896-6273(00)80803-X

Saito, H., Chi, Q., Zhuang, H., Matsunami, H., & Mainland, J. D. (2009). Odor coding by a mammalian receptor repertoire. *Science Signaling*, *2*(60). https://doi.org/10.1126/scisignal.2000016

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Shapley, L. S. (1951). Notes on the n-Person Game II: The Value of an n-Person Game. In *Rand Corporation Research Memoranda: Vol. RM-670*.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 3145–3153). PMLR. https://proceedings.mlr.press/v70/shrikumar17a.html

Soelter, J., Schumacher, J., Spors, H., & Schmuker, M. (2020). Computational exploration of molecular receptive fields in the olfactory bulb reveals a glomerulus-centric chemical map. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-56863-4

Soucy, E. R., Albeanu, D. F., Fantana, A. L., Murthy, V. N., & Meister, M. (2009). Precision and diversity in an odor map on the olfactory bulb. *Nature Neuroscience*, *12*(2). https://doi.org/10.1038/nn.2262

Takahashi, Y. K., Kurosaki, M., Hirono, S., & Mori, K. (2004). Topographic representation of odorant molecular features in the rat olfactory bulb. *Journal of Neurophysiology*, *92*(4). https://doi.org/10.1152/jn.00236.2004

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11141 LNCS*. https://doi.org/10.1007/978-3-030-01424-7_27

The Good Scents Company. (2023). *www.TheGoodScentsCompany.com*. TSGC Information System.

Uchida, N., Takahashi, Y. K., Tanifuji, M., & Mori, K. (2000). Odor maps in the mammalian olfactory bulb: Domain organization and odorant structural features. *Nature Neuroscience*, *3*(10). https://doi.org/10.1038/79857

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. In *AI Open* (Vol. 1, pp. 57–81). Elsevier B.V. https://doi.org/10.1016/j.aiopen.2021.01.001

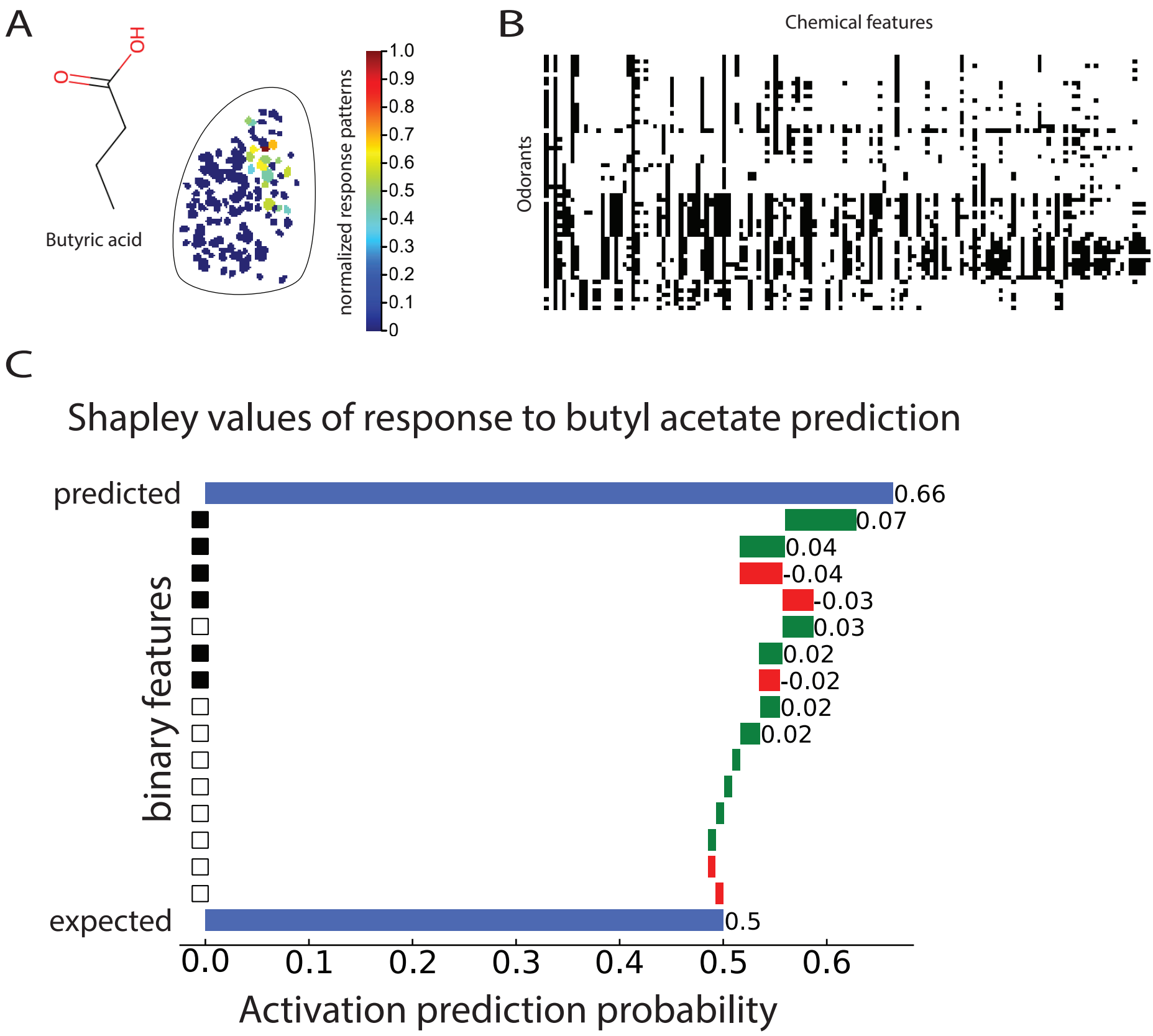


Fig. 1: Predicting glomerular activation in response to odorants. A. Glomerular activation. An example of glomerular activation pattern in response to butyric acid. **B. PubChem chemical features.** Binary chemical features of the presented odorants, from the PubChem database (Methods). Black indicates presence of a chemical feature, while white indicates absence. **C. Explanation of chemical feature contribution towards predicting activation of a glomerulus to a particular odorant (butyl acetate).** The contribution of each feature towards model prediction is calculated with Shapley values (Methods). Positive Shapley values (green) show contribution towards predicting activation; negative Shapley values (red) show contribution towards predicting no activation. The black squares on the y-axis indicate presence of a particular chemical feature in the odorant, the white squares indicate absence. The bottom blue line is the expected prediction; probability of activation the model would predict if no chemical features were given to the model. In this example, all Shapley values of features together with the model expected prediction sum to give the 66% probability of activation (top blue bar). This is represented in the waterfall plot; each feature's Shapley value contributes to the model prediction starting from 0.5 and moving the prediction to the left (if Shapley value is negative) or the right (if Shapley value is positive). Only the 15 features with the largest absolute Shapley values are shown in decreasing order from top to bottom.

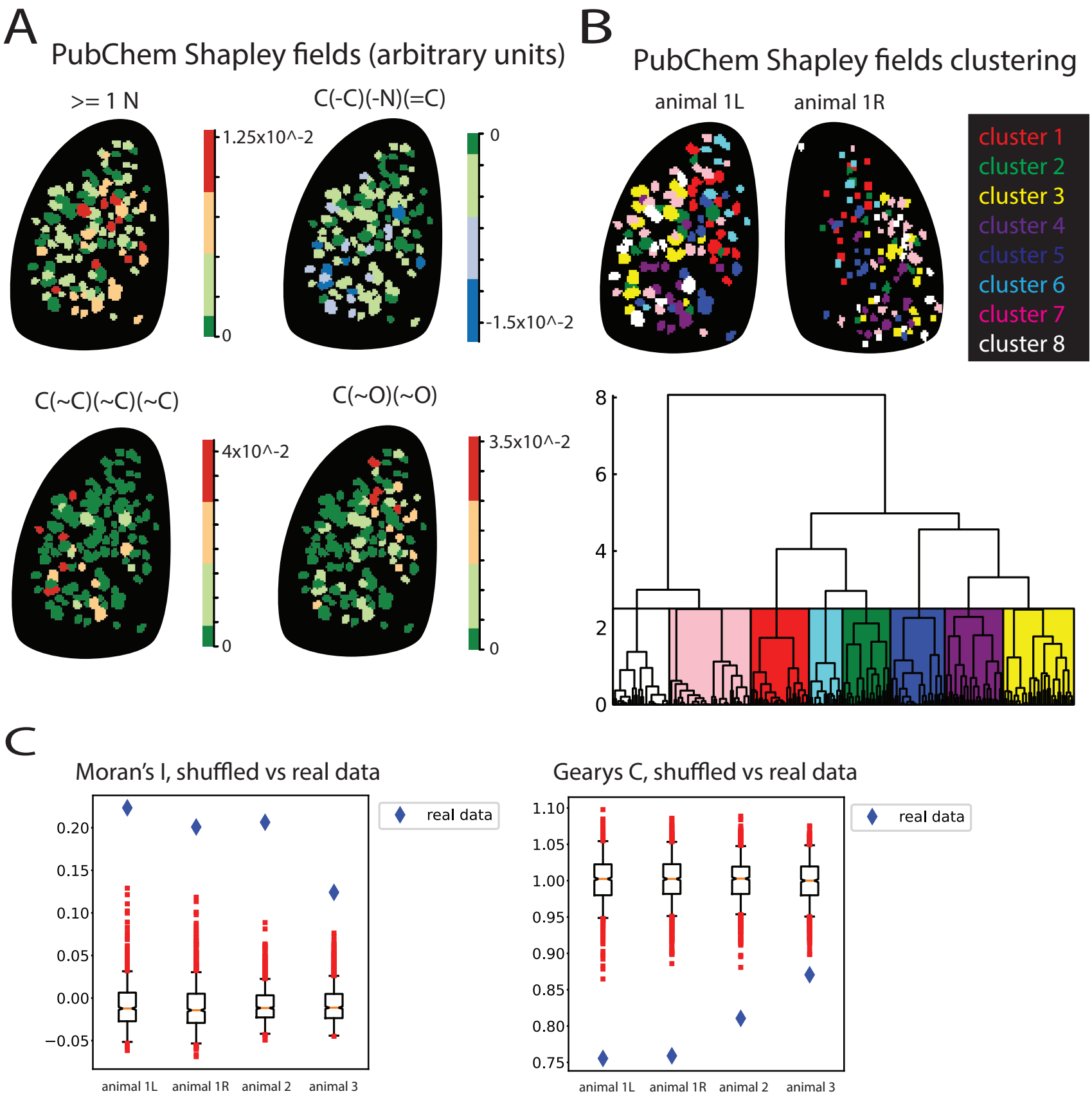


Fig. 2: Mosaic chemotopic organization of feature importance on the OB. A. Shapley fields. Importance of four chemical features from the PubChem feature set is displayed on the surface of the left OB, of animal 1, via Shapley fields (Methods). The top right figure is the negative Shapley field of the feature C(-C)(-N)(=C). The rest of the figures are the positive Shapley fields of the respective features. **B. Clustering of Shapley fields results in spatially correlated clusters on the surface of OB.** Hierarchical clustering of glomeruli on left and right OBs of animal 1, based on the first 20 principal components of Shapley fields, for the Pubchem chemical features. Cosine distance ward linkage clustering was used on the principal components of Shapley fields. Most clusters have a "mosaic" distribution that is roughly consistent between the left and the right OBs. **C. The spatial clustering inferred from Shapley fields is statistically significant.** Glomeruli positions on the OB were randomly shuffled n times (n=1000). After each shuffling, a glomerulus was assigned to its original cluster from Panel B. The Moran's I and Gearys C clustering scores (Methods), across all 4 datasets show a highly significant difference (p=0) between the clusters from the data (blue) and the randomly shuffled ones (red).

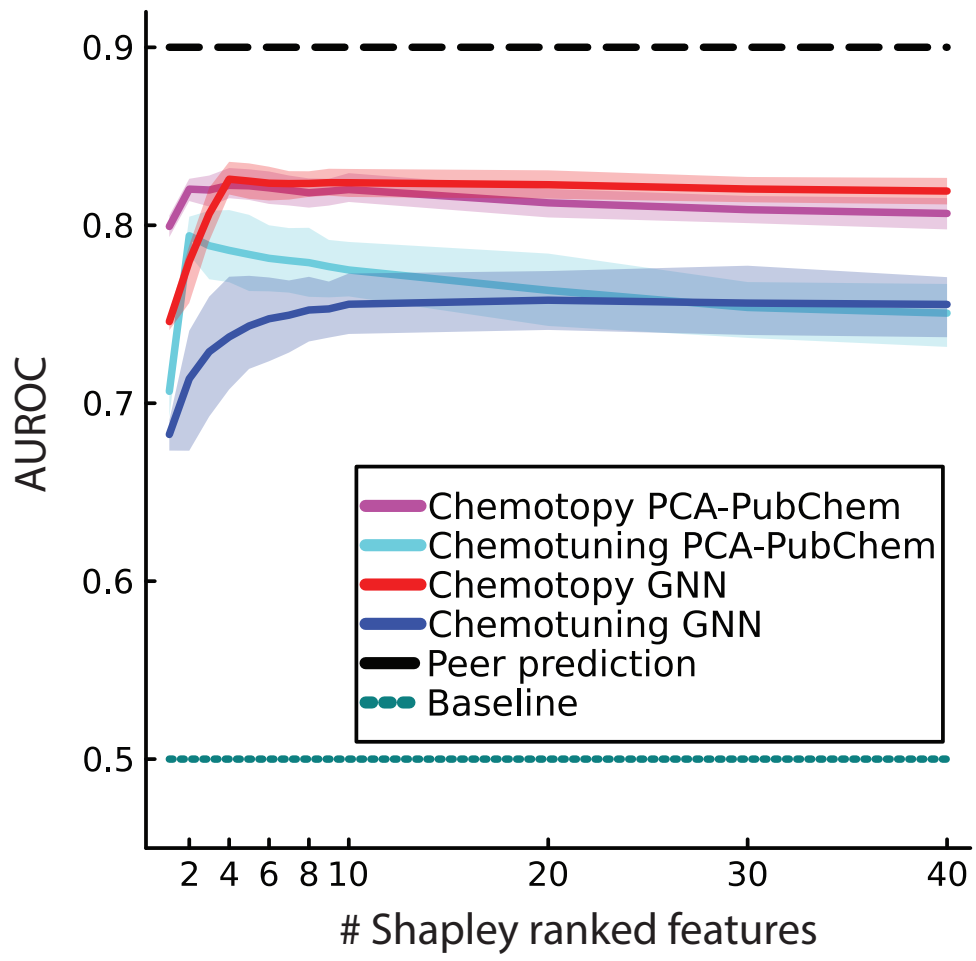


Fig. 3. Chemotopy model outperforms chemotuning model. Three different models: chemotuning, chemotopy and the peer prediction are compared vs baseline, across chemical feature sets. The features in a chemical feature set were ranked according to their global Shapley values, largest global Shapley value first (Methods). The features were added, one at a time according to their Shapley ranks, to the two models of glomerular activation prediction from chemical features; chemotopy and chemotuning (Methods). The mean cross-validated AUROC score is plotted along with its 95% confidence interval (shaded regions), N=200, against the number of Shapley ranked chemical features used. The mean cross-validated peer prediction model (black dashed line) AUROC score is also shown as well as the baseline (teal) having value of 0.5 which is the AUROC score of a random binary classifier. The 95% confidence interval for the peer prediction was too narrow to be perceptible on the plot (0.8970-0.9026).

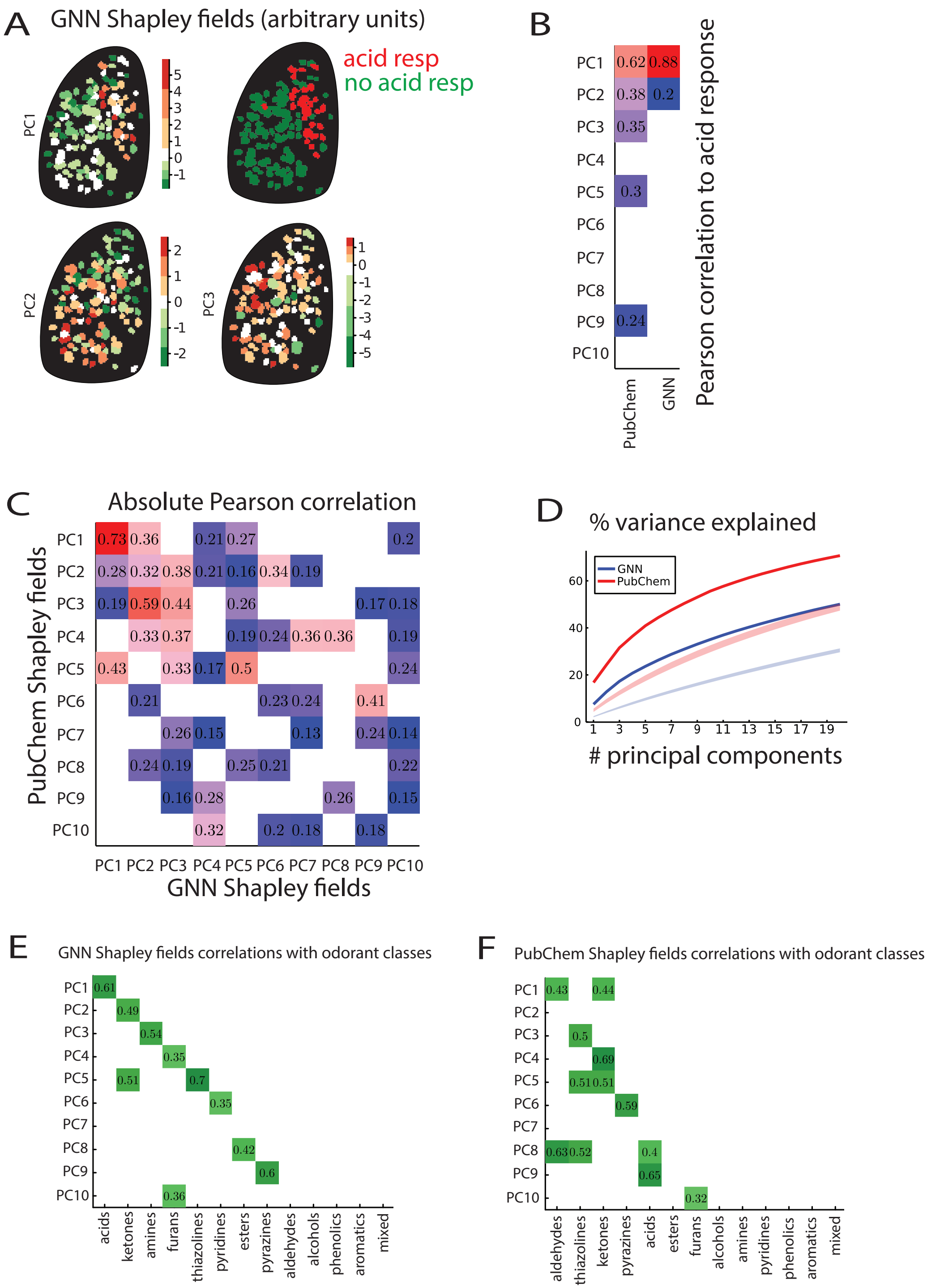
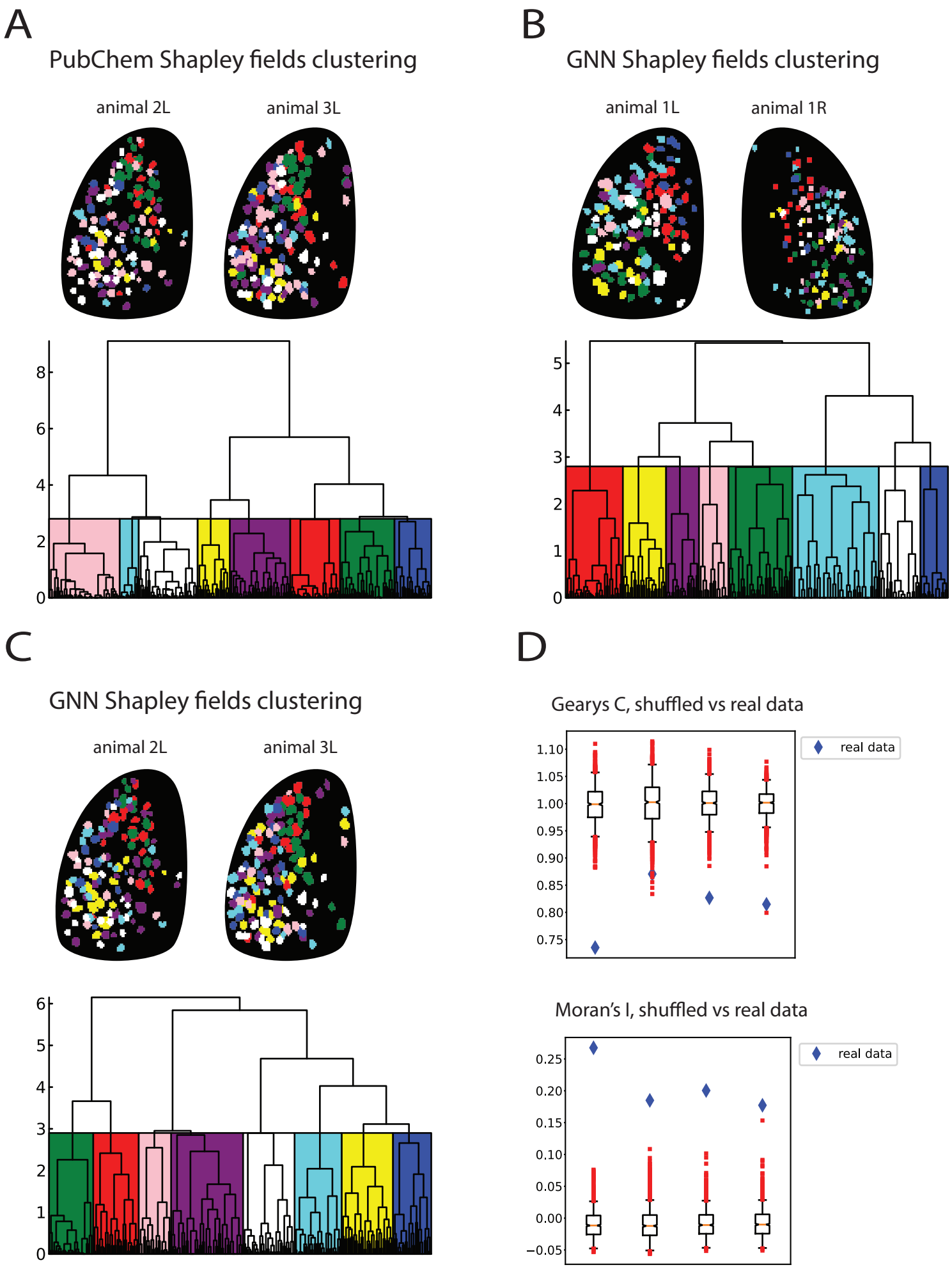
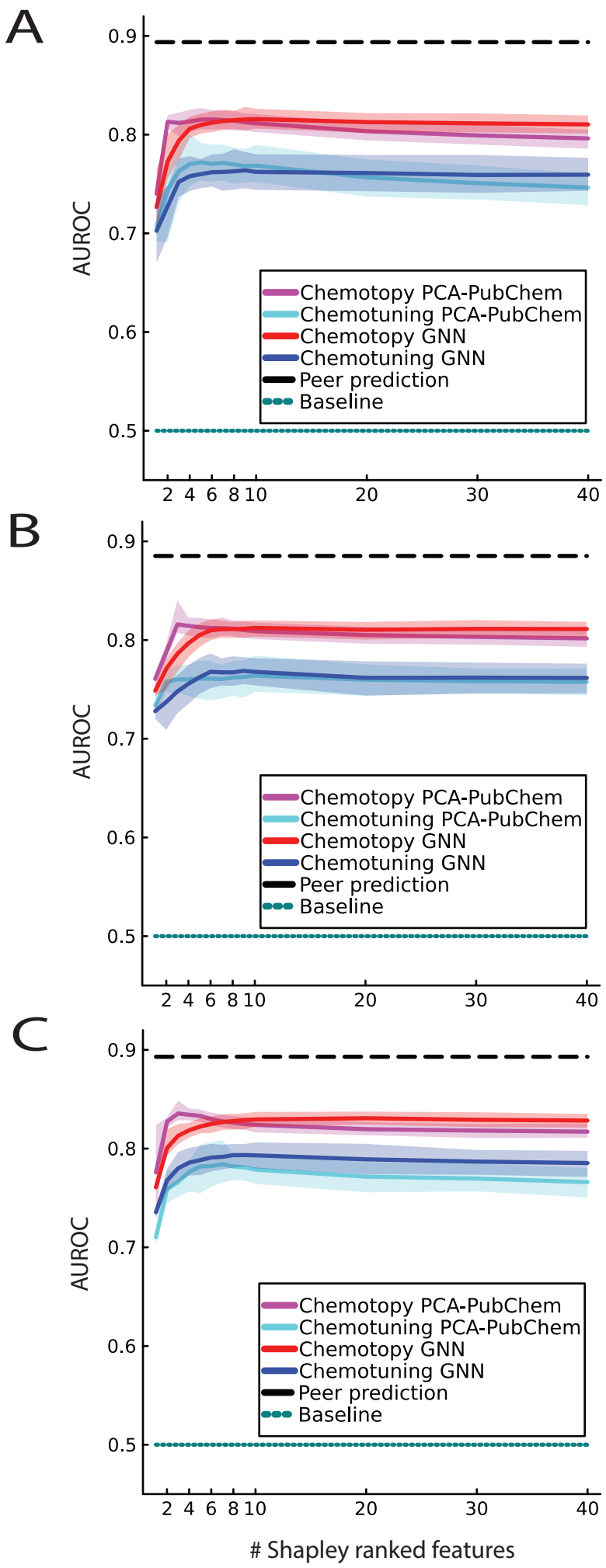

A
GNN Shapley fields (arbitrary units)
acid resp
no acid resp
PC1
PC2
PC3
B
Pearson correlation to acid response
PubChem
GNN
C
Absolute Pearson correlation
PubChem Shapley fields
GNN Shapley fields
D
% variance explained
GNN
PubChem
principal components
E
GNN Shapley fields correlations with odorant classes
acids
ketones
amines
furans
thiazolines
pyridines
esters
pyrazines
aldehydes
alcohols
phenolics
aromatics
mixed
F
PubChem Shapley fields correlations with odorant classes

Fig. 4: Shapley fields reveal an underlying "universal" chemical organization on the OB. A. Principal components of Shapley fields. Importance of GNN features is projected on the OB via Shapley fields (Methods). First 3 principal components of Shapley fields on the left OB of animal 1, along with a classification of glomeruli into acid responsive (red) and acid non-responsive (green) are displayed. Visually the first principal component of GNN Shapley fields corresponds to acid tuning separation on the OB. **B. Interpretation of principal components of Shapley fields.** Absolute Pearson correlation values between first 10 principal components of PubChem and GNN Shapley fields with the acid sensitivity vector. The acid sensitivity vector is a count of the number of acids that activate a given glomerulus. Only significant correlations ($p < 0.05$) are shown. GNN Shapley fields PC1 is highly correlated with responsiveness to acids ($r = 0.88$). All the principal components of the GNN Shapley fields beyond PC2 show no significant correlation with acid sensitivity vector. In contrast, several PubChem Shapley fields principal components show correlation to the acid sensitivity vector to some degree, albeit none as high as the PC1 of GNN Shapley fields. **C. The Shapley fields of PubChem features and GNN features are highly correlated.** The absolute values of Pearson correlations of GNN's and PubChem's first 10 principal components of the respective Shapley fields are displayed. Only significant correlations ($p < 0.05$) are shown. There are significant degrees of correlation among the 10 principal components of Pubchem and GNN Shapley fields. **D. Variance explained by principal components.** We find that the first 20 principal components of PubChem Shapley fields account for about 70% of the total variance in Shapley fields (red curve). In contrast, the first 20 principal components of GNN Shapley fields account for about 50% (blue curve). The most probable reason for this difference is the simple difference in the number of Shapley fields; there are 280 PubChem Shapley fields and 512 GNN Shapley fields. The percentage of variance explained is statistically significant when compared to the shuffled Shapley fields principal components, from panel D, for both PubChem (red shaded region) and GNN (blue shaded region). **E. Principal components of Shapley fields are correlated with sensitivities of ≤ 2 odorant classes.** Principal components of GNN Shapley fields were projected onto the linear subspace spanned by odorant classes sensitivity vectors on the OB (Methods). The projection of principal components of Shapley fields can be written as a linear combination of different odorant classes sensitivity vectors. Statistical analysis involving the shuffling of odorant identities (N=10000) determined which of the coefficients in the projection are statistically significant ($p < 0.05$), see Methods. Only the absolute values of statistically significant coefficients, for each projection, are shown in the table. **F. Principal components of PubChem fields are correlated with sensitivities of different odorant classes on the OB.** Same analysis as in panel E, but for the PubChem feature set.



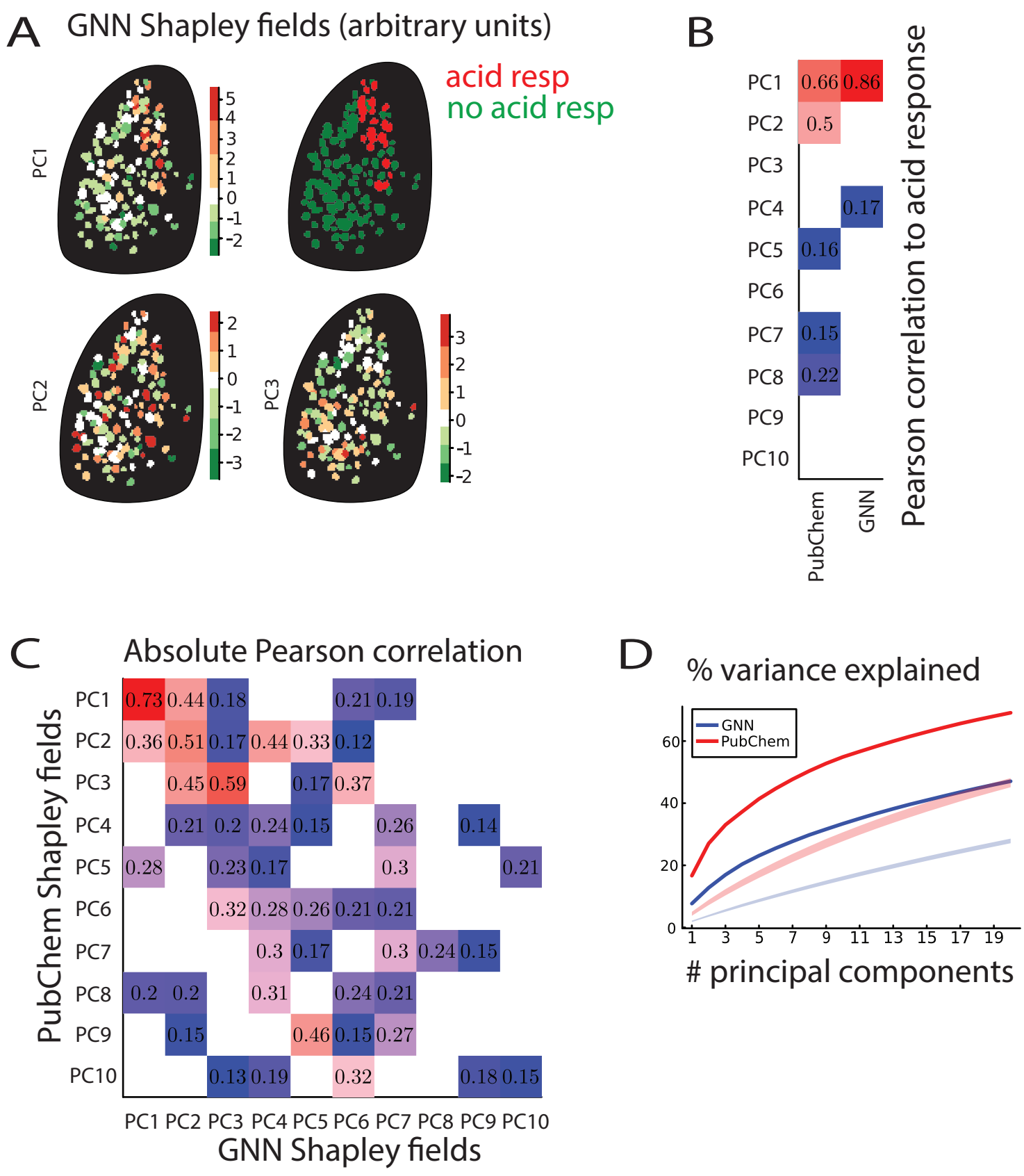
SuppFig. 1: Mosaic chemotopic organization of feature importance on the OB. A. The clustering in the space of PubChem Shapley fields results in spatially correlated clusters on the surface of OB. Hierarchical clustering of glomeruli on left OBs of animals 2 and 3, based on the first 20 principal components of Shapley fields, for the Pubchem chemical features. Cosine distance ward linkage clustering was used on the principal components of Shapley fields. A cutoff for clustering purposes was chosen as shown on the dendrogram. Most clusters have a "mosaic" distribution that is roughly consistent between the OBs. **B and C. The clustering in the space of GNN Shapley fields results in spatially correlated clusters on the surface of OB.** Hierarchical clustering of glomeruli on all OBs in our dataset based on the first 20 principal components of Shapley fields, for the GNN chemical features. Cosine distance ward linkage clustering was used on the principal

components of Shapley fields. A cutoff for clustering purposes was chosen as shown on the dendrogram. Most clusters have a "mosaic" distribution that is roughly consistent between the OBs. **D. The spatial clustering inferred from Shapley fields is statistically significant.** Glomeruli positions on the OB were randomly shuffled n times (n=1000). After each shuffling, a glomerulus was assigned to its original cluster from Panel B. The Moran's I and Gearys C clustering scores (Methods), across all 4 datasets show a highly significant difference (p=0) between the clusters from the data (blue) and the randomly shuffled ones (red).



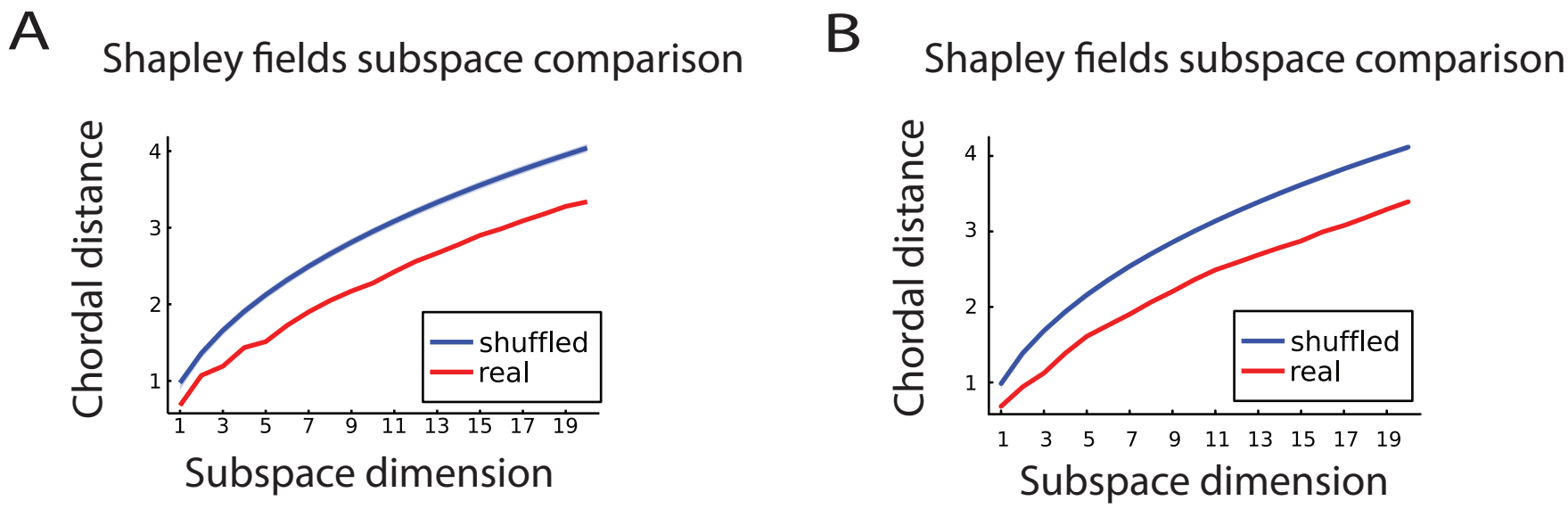
SuppFig. 2: Chemotopy model outperforms chemotuning model. A. Three different models: chemotuning, chemotopy and the peer prediction are compared vs baseline, across chemical feature sets. The features in a chemical feature set were ranked according to their global Shapley values, largest global Shapley value first (Methods). The features were added, one at a time according to their Shapley ranks, to the two models of glomerular activation prediction from chemical features; chemotopy and chemotuning (Methods). The mean cross-validated AUROC score is plotted along with its 95% confidence interval (shaded regions) against the number of Shapley ranked chemical features used. The mean cross-validated peer prediction model (black dashed line) AUROC score is also shown as well as the baseline (teal) having value of 0.5 which is the AUROC score of a random binary classifier. The 95% confidence interval for the peer prediction was too narrow to be perceptible on the plot (0.8913-0.8962). The data models performances shown are for the right OB of animal 1. **B. Three different models: chemotuning, chemotopy and the peer prediction**

are compared vs baseline, across chemical feature sets. The same as in A, but for the left OB of animal 2. **C. Three different models: chemotuning, chemotopy and the peer prediction are compared vs baseline, across chemical feature sets.** The same as in A, but for the left OB of animal 3.

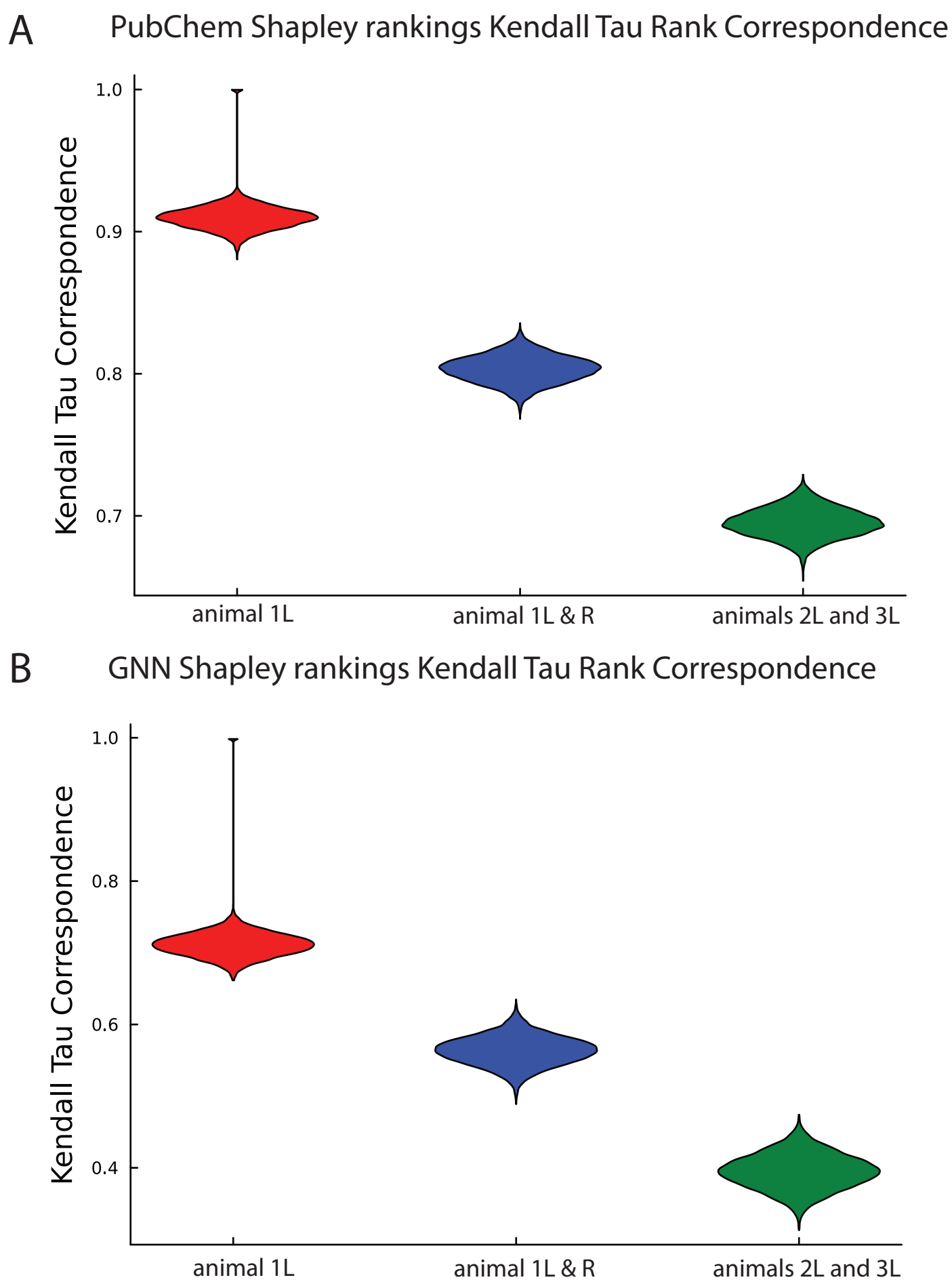


SuppFig. 3: Shapley fields reveal an underlying "universal" chemical organization on the OB. A. Principal components of Shapley fields. Importance of GNN features is projected on the OB via Shapley fields (Methods). First 3 principal components of Shapley fields on the left OB of animal 2, along with a classification of glomeruli into acid responsive (red) and acid non-responsive (green) are displayed. Visually the first principal component of GNN Shapley fields corresponds to acid tuning separation on the OB. **B. Interpretation of principal components of Shapley fields.** Absolute Pearson correlation values between first 10 principal components of PubChem and GNN Shapley fields with the responsiveness to acids vector. The responsiveness to acids vector is a count of the number of acids that activate a given glomerulus. Only significant correlations ($p < 0.05$) are shown. GNN Shapley fields PC1 is highly correlated with responsiveness to acids ($r = 0.86$). All the principal components of the GNN Shapley fields beyond PC2 show no significant correlation with acid response. In contrast, several PubChem Shapley fields principal components show correlation to responsiveness to acids to some degree, albeit none as high as the PC1 of GNN Shapley fields. **C. The Shapley fields of PubChem features and GNN features are highly correlated.** The absolute values of Pearson correlations of GNN's and PubChem's first 10 principal components of the respective Shapley fields are displayed. Only significant correlations ($p < 0.05$) are shown. There are significant degrees of correlation among the 10 principal components of Pubchem and GNN Shapley fields. **D. Variance explained by principal components.**

We find that the first 20 principal components of PubChem Shapley fields account for about 69% of the total variance in Shapley fields (red curve). In contrast, the first 20 principal components of GNN Shapley fields account for about 47% (blue curve). The most probable reason for this difference is the simple difference in the number of Shapley fields; there are 264 PubChem Shapley fields and 512 GNN Shapley fields. The percentage of variance explained is statistically significant when compared to the shuffled Shapley fields principal components, from panel D, for both PubChem (red shaded region) and GNN (blue shaded region).

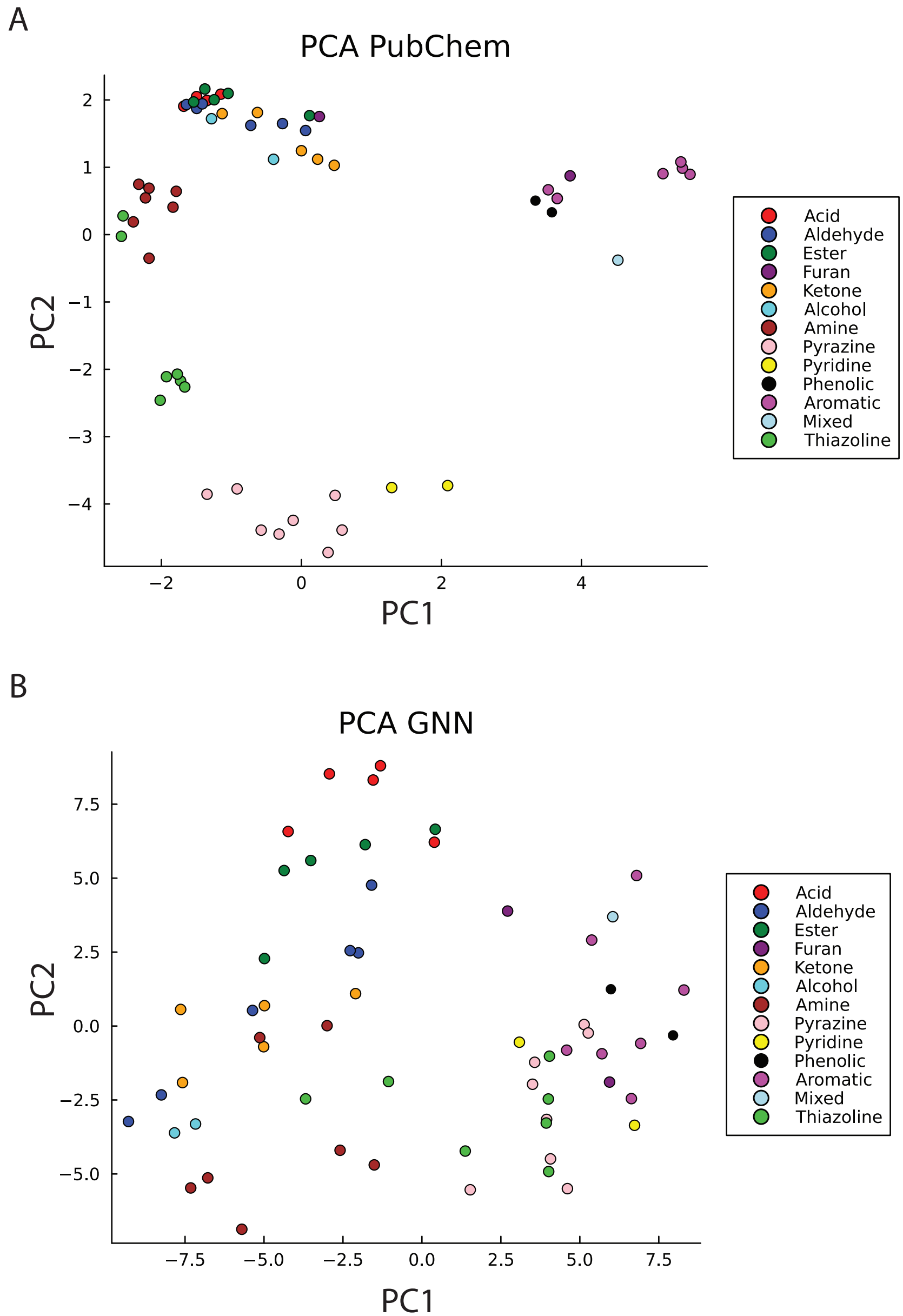


SuppFig. 4: Human-derived chemical features and molecular fingerprints share chemical organization on the OB. A. Distances between linear subspaces of PCs of GNN and PubChem Shapley fields. Principal angles and the resulting chordal distance between GNN and PubChem Shapley field principal component subspaces (Methods) were computed. We only considered the first 20 principal components of both GNN and PubChem Shapley fields and computed the principal angles of the linear subspaces they span, starting with PC1s and adding the other PCs one at a time. The resulting distance vs dimension curve is shown in red. We shuffled the activation patterns on both OBs of animal 1 (N=1000), and recomputed both GNN and PubChem Shapley fields along with their chordal distances of subspaces generated by the first 20 principal components. The resulting curve (blue) with its 95% confidence interval shows that the real chordal distance between GNN and PubChem Shapley fields is statistically significant. **B. Distances between linear subspaces of PCs of GNN and PubChem Shapley fields.** Same analysis as in B but for left OBs of animals 2 and 3.



SuppFig. 5: Stability of Shapley ranks of chemical feature sets. A. Kendall rank correlation coefficients of Shapley rankings of PubChem features. Chemotopy model (see Methods) was trained 100 times (100 different random seeds), and the Shapley rankings of PubChem features used by the model were computed (see Methods). To analyze the stability of rankings (how much they change between random seeds) we computed the Kendall rank correlation coefficients between rankings. This comparison was done for Shapley rankings of PubChem features on the same OB (red), same animal but different OBs (blue), and on different OBs in different animals (green). Each violin plot contains 10000 such rank comparisons. The high values of Kendall rank correlation coefficients (the y-axis) indicate the stability of Shapley ranks of features. **B. Kendall rank correlation coefficients of Shapley rankings of GNN features.** Same analysis as in A, but for the the GNN feature set. Overall there is less stability for GNN Shapley ranks than for the PubChem Shapley ranks.

SuppFig. 6: Odorants used. Odorant names and odorant classes used in experiments.



SuppFig. 7: PCA projections of chemical feature sets. A. PCA projections of the PubChem feature set. We reduced the dimension of the PubChem feature space of the 59 odorants used in our experiments to 2 dimensions, using PCA. Each odorant is represented as a point in the 2 dimensional space, with a color representing its odorant class. **B. PCA projections of the GNN feature set.** Same is in A, but for the GNN feature set.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
PubChem Shapley fields	0.87 p=1.6e-68	0.81 p=8.3e-52	0.76 p=1.6e-42	0.66 p=4.7e-49	0.85 p=2.9e-62	0.67 p=3.1e-30	0.45 p=1.3e-12	0.6 p=4.9e-23	0.53 p=1.6e-17	0.65 p=1.1e-27
GNN Shapley fields	0.93 p=8.2e-97	0.73 p=1.6e-37	0.65 p=1.2e-37	0.49 p=1.7e-14	0.73 p=7.1e-38	0.48 p=6.2e-14	0.47 p=2.7e-13	0.43 p=2.6e-11	0.5 p=1.7e-15	0.51 p=3.3e-16

Supplementary Table 1: Correlations of Shapley fields to odorant classes sensitivity vectors. Principal components of Shapley fields were projected onto the linear subspace spanned by odorant classes sensitivity vectors on the OB (Methods). The Pearson correlations between the principal components and their respective projections are shown in the table, along with the p-value of the correlation.