The anatomical logic of smell

Mayra L. Ruiz Tejada Segura[1,2,3,#]; Eman Abou Moussa[4,#]; Elisa Garabello[5,6]; Thiago S. Nakahara[7]; Melanie Makhlouf[4]; Lisa S. Mathew[4]; Filippo Valle[5]; Susie S.Y. Huang[4]; Joel D. Mainland[8,9]; Michele Caselle[5]; Matteo Osella[5]; Stephan Lorenz[4,10]; Johannes Reisert[8]; Darren W. Logan[10]; Bettina Malnic[7]; Antonio Scialdone[1,2,3,†,*]; Luis R. Saraiva[4,8,11, †,*]

[1] Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München; Feodor-Lynen-Strasse 21, 81377 München, Germany
[2] Institute of Functional Epigenetics, Helmholtz Zentrum München; Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[3] Institute of Computational Biology, Helmholtz Zentrum München; Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[4] Sidra Medicine; PO Box 26999; Doha, Qatar.
[5] Physics Department, University of Turin and INFN; via P. Giuria 1, 10125 Turin, Italy
[6] Department of Civil and Environmental Engineering, Cornell University; Ithaca, NY 14853, USA
[7] Department of Biochemistry, University of São Paulo; São Paulo, Brazil.
[8] Monell Chemical Senses Center; 3500 Market Street, Philadelphia, PA 19104, USA.
[9] Department of Neuroscience, University of Pennsylvania; Philadelphia, PA 19104, USA.
[10] Wellcome Sanger Institute; Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK.
[11] College of Health and Life Sciences, Hamad Bin Khalifa University; Doha, Qatar.

[#] Equal contribution (co-first)
† These authors jointly supervised this work
*Correspondence and requests for materials should be addressed to A.S. (email: antonio.scialdone@helmholtz-muenchen.de) or L.R.S. (email: saraivalmr@gmail.com; twitter: @saraivalab)

Abstract

The sense of smell helps us navigate the environment, but its anatomical logic remains unknown. The spatial location of odorant receptor genes (*Olfrs*) in the nose is widely thought to be independent of the structural diversity of the odorants they detect. Using spatial transcriptomics, we created a genome-wide 3D atlas of the mouse olfactory mucosa (OM), and identified key genes differentially expressed in space. Expression maps reveal that *Olfrs* are distributed in a continuous and overlapping fashion over five broad zones in the OM. The spatial locations of *Olfrs* correlate with the mucus solubility of the odorants they recognize. Thus, we provide direct evidence for the chromatographic theory of olfaction, and elucidate the basic logic for the peripheral representation of smell.

One-Sentence Summary

Chemical solubility is organized topographically in the mouse nose

Keywords

Spatial transcriptomics; 3D; olfaction; olfactory epithelium; RNA-seq; odorant

Main text

The functional logic underlying the topographic organization of primary receptor neurons and their receptive fields is well-known for all sensory systems, except olfaction (*1*). The mammalian nose is constantly flooded with odorant cocktails. Powered by a sniff, air enters the nasal cavity, until it reaches the olfactory mucosa (OM). There, myriad odorants activate odorant receptors (*Olfrs*) present in the cilia of olfactory sensory neurons (OSNs), triggering a complex cascade of events that culminate in the brain and result in odor perception (*1, 2*). In mice, most mature OSNs express a single allele of one out of ~1100 Olfr genes (*Olfrs*) (*3-6*). Olfrs employ a combinatorial strategy to detect odorants, which maximizes their detection capacity (*4, 7*). OSNs expressing the same *Olfr* share similar odorant response profiles (*4, 7*), and drive their axons to the same glomeruli in the olfactory bulb (*8-10*). Thus, *Olfrs* define functional units in the olfactory system, and function as genetic markers to discriminate between different mature OSN subtypes (*6, 11*).
Another remarkable feature of the OSN subtypes is their spatial distribution in the OM. Early studies postulated that OSNs expressing different *Olfrs* are spatially segregated into four broad areas within the OM called 'zones' (*12, 13*), but recent work has suggested up to nine overlapping zones (*14-16*). The functional relevance of these zones is not fully understood. These studies, combined, have sampled only ~10% of the intact mouse *Olfr* repertoire. We do not currently understand the full complexity of the OM and, most importantly, lack an unbiased and quantitative definition of zones. In effect, the exact number of zones, their anatomical boundaries, their molecular identity and their potential functional relevance are yet to be determined.

One hypothesis is that the topographic distribution of *Olfrs*/OSN subtypes evolved because it plays a key role in the process of *Olfr* choice in mature OSNs and/or in OSN axon guidance (*17, 18*). An alternative hypothesis is that the spatial organization of *Olfrs*/OSN subtypes is tuned to maximize the detection and discrimination of odorants in the peripheral olfactory system (*12*). Interestingly, the receptive fields of mouse OSNs vary with their spatial location (*19*), which in some cases correlate with the patterns of odorant sorption in the mouse OM – this association was proposed as the '*chromatographic hypothesis*' decades before the discovery of the *Olfrs* (*20*), and later rebranded as the '*sorption hypothesis*' in olfaction (*21, 22*). While some studies lend support to these hypotheses (reviewed in (*23*)), others question their validity (*24, 25*). Thus, the logic underlying the peripheral representation of smell in the peripheral olfactory system still remains unknown, and it is subject of great controversy (*23, 26*).

Spatial transcriptomics, which combines spatial information with high-throughput gene expression profiling, has started to expand our knowledge of complex tissues, organs, or even entire organisms (*27-31*). In this study, we employed a spatial transcriptomics approach to create an 3D map of gene expression of the mouse nose,

and combined it with machine learning and chemoinformatics to reveal the anatomical logic of smell in mouse.

A high-resolution spatial transcriptomic map of the mouse olfactory mucosa

We adapted the RNA-seq tomography (Tomo-seq) method (*30*) to create a spatially resolved genome-wide transcriptional atlas of the mouse nose. We obtained cryosections (35 µm) collected along the dorsal-ventral (DV), anterior-posterior (AP), and lateral-medial-lateral (LML) axes (n=3 per axis) of the OM (Fig. 1A), and performed RNA-seq on individual cryosections (gMethods). After quality control (see Methods, fig. S1A-D and table S1), we computationally refined the alignment of the cryosection along each axis, and we observed a high correlation between biological replicates (see Methods and fig. 1B). Hence, we combined the three replicates into a single series of spatial data including 54, 60 and 56 positions along the DV, AP and LML axis, respectively (see Methods and Fig. 1C). On average, we detected >18,000 genes per axis, representing a total of 19,249 unique genes for all axes combined (Fig. 1D). Molecular markers for all canonical cell types known to populate the mouse OM were detected among all axes (Fig. 1E), and were expressed at the expected levels (*6*). Next, we verified the presence of a spatial signal with the Moran's I (*32*) (fig. S1E). Given the left/right symmetry along the LML axis (see Methods; Fig. 1C), the data was centered and averaged on the two sides (see Methods) – henceforth the LML axis will be presented and referred to as the lateral-medial (LM) axis. We could reproduce the expression patterns in the OM for known spatial markers, including the dorsomedial markers *Acsm4* and *Nqo1* (*33, 34*), and the complementary ventrolateral markers *Ncam2* and *Reg3g* (*35, 36*) (Fig. 1F, fig. S1F).

Spatial differential gene expression analysis identifies cell type-specific expression patterns and functional hotspots in the OM

Having established the accuracy of our dataset, we investigated cell-type specific patterns. Previous studies showed that the vast majority of genes displaying spatially segregated expression patterns across the OM are expressed in mature OSNs (*6, 13, 14, 16, 33, 34, 36-48*), but a small fraction are expressed in sustentacular, globose basal, olfactory ensheathing, Bowman's gland, and respiratory epithelial cells (*36, 42, 47-53*). Despite this progress, our knowledge on what genes display zonal expression patterns, and what cell types they are primarily expressed in is still very limited.

To identify axis-specific differentially expressed genes (hereafter referred to as spatial DEGs), we applied the Ljung-Box test to the autocorrelation function calculated on the binarized expression values (fig. S2A, and Methods), and obtained a total of 12,303 DEGs for the 3 axes combined (FDR<0.01; Fig. 2A).

130 To add cell-type resolution to the spatial axes, we combined our data with a single-cell
131 RNA-seq (scRNA-seq) dataset from 13 cell types present in the mouse OM (*54*). We
132 catalogued spatial DEGs based on their expression in mature OSNs (mOSNs) versus
133 the 12 other cell types (non-mOSNs) (Fig. 2B,C; table S2). This led to the
134 identification of 456 spatial DEGs expressed exclusively in non-mOSNs, which are
135 associated with gene ontology (GO) terms such transcription factors, norepinephrine
136 metabolism, toxin metabolism, bone development, regulation of cell migration, T-cell
137 activation, and others (table S2). Some of these genes are expressed across many cell
138 types, but others are specific to a single cell type (Fig. 2C, table S2). For instance,
139 there are cell-specific markers with known spatial expression patterns, such as the
140 sustentacular cells and Bowman's glands markers *Cyp2g1* and *Gstm2* (*36*), the neural
141 progenitor cell markers *Eya2* and *Hes6* (*45*), and the basal lamina and olfactory
142 ensheathing cell markers *Aldh1a7* and *Aldh3a1* (*47*) (table S2). Additionally, we found
143 previously unknown genes with zonal expression patterns along a single axis or
144 multiple axes, such as the ribosomal protein *Rps21* that is primarily expressed in
145 horizontal basal cells (HBCs), the extracellular proteinase inhibitor *Wfdc18* in
146 microvillous cells type 1 (MVC1s), the fibroblast growth factor *Fgf20* in immature
147 sustentacular cells (iSCs), and the adapter protein *Dab2* in mature sustentacular cells
148 (mSCs) (fig. S2 B,C). These results suggest differences either in their expression levels
149 and/or in the relative abundance of cell types across the OM.
150
151 The successful identification of thousands of spatial DEGs, prompted us to examine
152 their distribution patterns along each axis, and the putative functions associated with
153 such spatial clusters of gene expression. We started by using uniform manifold
154 approximation and projection (UMAP) (*55*) and hierarchical clustering to visualize
155 and cluster all spatial DEGs along the three cartesian axes. This analysis uncovered
156 nine patterns of expression in the DV and AP axes each, and five patterns in the LM
157 axis (Fig. 2D-E). These patterns include variations of four major shapes:
158 monotonically increasing (/), monotonically decreasing (\), U-shape (∪), and inverted
159 U-shape (∩) (Fig. 2E). The latter two patterns present clear maximum/minimum at
160 different positions along the axis – for example, the brown, green, pink, magenta, and
161 black AP clusters show a similar inverted ∪ pattern, but their maximum moves along
162 the axis (Fig. 2E). As expected, several known genes with zonal expression patterns
163 are expressed in the cluster mimicking their respective expression pattern in the tissue.
164 For example, the dorsomedial markers *Acsm4* and *Nqo1* belong to the turquoise
165 clusters in both the DV and LM axes, while the ventrolateral marker *Reg3g* belongs to
166 the blue cluster from the DV axis (Fig. 1F, S1F; table S3).
167 GO enrichment analysis on the axis-specific DEGs yielded enriched terms for 14 of
168 the 23 spatial clusters (table S3). For example, the turquoise AP cluster displaying a
169 monotonically increasing pattern (Fig. 2E) yielded GO terms broadly associated with

170 the molecular machinery of mOSNS – such as axonal transportation, RNA
171 processing, protein modification and quality control, ribosomal regulation and
172 regulation of histone deacetylation (fig. S2D and table S3). Interestingly, the brown
173 DV cluster, which displays a monotonically decreasing expression pattern (Fig. 2E)
174 had similar GO term enrichment (fig. S2E and table S3). These results raise the
175 hypothesis that mOSN activity is enriched in the dorsoposterior region of the OM,
176 consistent with previous results (*56, 57*).
177 We then extended our analysis to the remaining clusters, and found additional GO
178 terms shared between several clusters among the three different axes. For example,
179 the GO terms enriched in the dorsomedial region (turquoise DV, pink AP and LM
180 green) suggest that this region is involved in the OM detoxification (Fig. S2F-H).
181 Moreover, the enrichment in terms related to immune system in the anteromedial
182 section along the AP axis (yellow, black and magenta AP clusters, and turquoise LM
183 cluster; table S3) hints at a role of this area in defending OM from pathogenic
184 invaders. Finally, the ventral portion of the DV (red DV cluster) was associated with
185 terms related to cilia movement and function, consistent with both the location and
186 functions of the respiratory epithelium (*36*) (table S3).
187
188 Next, we further explored the relationships between the genes populating each
189 cluster. We found that ventral genes (blue DV cluster) tend to reach a peak in
190 expression in the anterior area of the OM (yellow AP cluster) more often than
191 expected by chance (Fig. 2F). We also observed that medial genes (turquoise LM
192 cluster) are more highly expressed in the dorsal (magenta DV cluster) and anterior
193 regions (black, yellow and magenta AP cluster), while genes peaking in the lateral
194 region (brown LM cluster) tend to be ventral (red DV cluster; Figure 2F). These
195 conclusions hold even when we exclude olfactory receptor genes from the analysis
196 (fig. S2I-K).
197
198 These associations between the clusters of DEGs along different axes strongly
199 suggest that the presence of complex 3D expression patterns in OM is not restricted
200 to either *Olfrs* or mOSNs. Moreover, our data allowed us to uncover spatially
201 restricted functional hotspots within the OM.

202 A 3D transcriptomic atlas of the mouse OM

203 Our Tomo-seq data enables the quantification of gene expression estimates along the
204 three cartesian axes of the OM. Here, we take this analysis one step further and
205 generate a fully browsable tridimensional (3D) gene expression atlas of the mouse
206 OM. To this aim, first we reconstructed the 3D shape of OM based on publicly
207 available images of OM sections (made available by (*58*), see Methods). We then fed
208 the resulting shape information combined with the gene expression data along the

three cartesian axes into the iterative proportional fitting (IPF) algorithm (*30, 59*) (Fig. 3A). Our resulting 3D atlas of the OM faithfully reproduced the known 3D pattern of the dorsomedial marker *Acsm4* (*34*) (Fig. 3B). In addition, we confirmed the predicted 3D expression patterns for two novel spatial DEGs we identified: *Cytl1* and *Moxd2* (table S3) using in situ hybridization (Fig. 3C-F). *Cytl1* is expressed mainly along the septum (Fig. 3C-D), consistent with the role *Cytl1* plays in osteogenesis, chondrogenesis, and bone/cartilage homeostasis (*60, 61*). *Moxd2*, a monooxygenase dopamine hydroxylase-like protein putatively involved in neurotransmitter metabolism (*62*), is expressed mostly in a small ventrolateral patch (Fig. 3E-F). We have made this new 3D gene expression atlas of the mouse OM available to the scientific community via a web-portal (at the address: http://atlas3dnose.helmholtz-muenchen.de:3838/atlas3Dnose), providing access to all the spatial transcriptomic data in a browsable, user-friendly format.

Topographical expression patterns of *Olfrs*

In our combined dataset we detected a total of 959 unique *Olfrs* (Fig. 4A), of which we confidently reconstructed the spatial expression patterns for 689 differentially expressed in space (FDR<0.01; Fig. 4B) – a number six times larger than the combined 112 *Olfrs* characterized by previous ISH studies (*10, 12, 14, 16*). To define *Olfr* expression in 3D space in a rigorous, unbiased and quantitative way, we ran a Latent Dirichlet Allocation algorithm (see Methods)(*63*) on the 689 spatially differentially expressed *Olfr*, which suggested the presence of five zones (see Methods and fig. S3A). Next, we visualized the spatial distribution of these five zones in our 3D model of the OM, with colors representing the probability that a given spatial position belongs to each zone (Fig. 4C). These zones extend from the dorsomedial-posterior to the lateroventral-anterior region, consistent with the previously described zones (*12-14*).

Using the same mathematical framework, we can compute the probabilities that the expression pattern of each *Olfr* gene belongs to a given zone, i.e., the "degree of belonging" (DOB, table S4). The DOBs represent a decomposition of the expression patterns in terms of the five zones shown in Fig. 4C and quantitatively describe the changes in patterns of genes with overlapping areas of expression (e.g., see fig. S3B). The width of the distribution of DOBs across the five zones, which can be measured with entropy, can distinguish genes whose patterns mostly fit in a single zone from those spanning multiple zones (see Methods and fig. S3C). These analyses can be extended also to non-*Olfr* genes, allowing the identification of marker genes for each zone (e.g., *Moxd2* for zone 2; see fig. 3E-F, S3D and Methods). To visualize the global distribution of the 689 *Olfrs*, we applied the diffusion map algorithm (*64*) to their DOBs. This showed that the genes are approximately distributed along a continuous

line spanning the five zones and without clear borders between zones (fig. 4D) (*14-16*). With the diffusion pseudo-time algorithm (*65*), we calculated an index (hereafter referred to as "3D index") that tracks the position of each *Olfr* gene along the 1D curve in the diffusion map and represents its expression pattern (fig. 4E).
While this approach yielded an index for the 689 *Olfr* genes used to build the diffusion map, there were ~700 *Olfrs* that could not be analyzed, either because they were too lowly expressed or not detected at all in our dataset (Fig. 4A). However, since the spatial expression patterns for some *Olfrs* are partly associated with their chromosomal/genomic coordinates (*66-68*), we hypothesized that we could train a machine learning algorithm to predict the 3D indexes for the ~700 *Olfrs* missing from our dataset. Thus, we trained a Random Forest algorithm on the 3D indexes of 681 spatially differentially expressed *Olfrs* in our dataset using 9 genomic features as predictors (e.g., chromosomal position, number of *Olfrs* in cluster, distance to nearest known enhancer, etc; see Methods and Fig. 4F). The five most important predictors were features associated with chromosomal position, distance to the closest *Olfr* enhancer (*69*), length of the *Olfr* cluster, position in the *Olfr* cluster, and phylogenetic class of *Olfrs* (Fig. 4F).

We assessed the performance of the algorithm with a cross-validation test (see Methods and Fig. S4A), and then predicted the 3D indexes for the 697 *Olfrs* missing reliable expression estimates in our dataset. Importantly, we found that our 3D indexes strongly correlate with the zonal indexes inferred using ISH in (*14*)(rho=0.88, $p=3x10^{-27}$; Fig. 4G) and in two other studies (*16, 66*)(fig. S4B,C). Furthermore, we performed ISH for two *Olfrs* that have not been characterized before, *Olfr309* and *Olfr727*, which were correctly predicted to be expressed primarily in zone 2 and 4 respectively (Fig. 4H-M and table S4)

Taken together, through multiple unsupervised and supervised computational methods, we have quantitatively defined five spatial expression domains in the OM (called zones), and have provided accurate 3D spatial indexes for 1386 *Olfrs*, which represents ~98% of the annotated *Olfrs*.

The anatomical logic of smell

Next, we used our 3D, quantitative reconstruction of the OM to address whether there is an underlying logic linked to *Olfrs* zonal distribution. We first compiled a list of known 738 Olfr-ligand pairs, representing 153 *Olfrs* and 221 odorants (Fig. 5A, table S5). Interestingly, we found that *Olfr* pairs sharing at least one common ligand have more similar expression patterns (i.e., more similar 3D indexes) than *Olfrs* detecting different sets of odorants (Fig. 5B). This observation is consistent with the

286 hypothesis that the *Olfr* zonal expression depends, at least partially, on the properties
287 of the odorants they bind to.
288 Previous studies showed that the sorption patterns of the odorants in the mouse OM
289 depend on their mucus solubility, which can be quantified by the air/mucus partition
290 coefficient (*21, 70*). In particular, odorants with higher mucus solubility reach a higher
291 concentration in the antero-dorsomedial region of the OM (*12, 21*). Hence, we
292 hypothesized that the Olfr zonal expression patterns correlate with the mucus
293 solubility of the cognate odorants, in such a way to mimic the sorption pattern of the
294 odorants.
295 To test this hypothesis, we compared the air/mucus partition coefficients (K_{am}) of
296 221 odorants to the average 3D indexes of the *Olfrs* detecting them (see Methods,
297 table S5). Strikingly, we found a positive correlation (rho=0.55, p-value=$7x10^{-10}$; Fig.
298 5C), indicating that the most soluble odorants (lower K_{am}) preferentially activate *Olfrs*
299 in the most antero-dorsomedial OM region (zone 1) of the OM, while the least
300 soluble odorants (higher K_{am}) activate *Olfrs* in the postero-ventrolateral OM region
301 (zones 4-5). In other words, gradients of odorants sorption (as defined by their K_{am})
302 correlate with the gradients of *Olfr* expression from zone 1 to zone 5, consistent with
303 the *chromatographic/sorption* hypothesis in olfaction (*20, 21*). This is exemplified by the
304 plots in Fig. 5D, illustrating the predicted average expression levels across OM
305 sections of the *Olfrs* binding to five odorants with different values of K_{am}. These
306 results show, for the first time, a direct association between *Olfr* spatial patterns and
307 the sorption patterns of their cognate ligands in the OM, providing a potential
308 explanation for the physiological function of the zones in the OM.

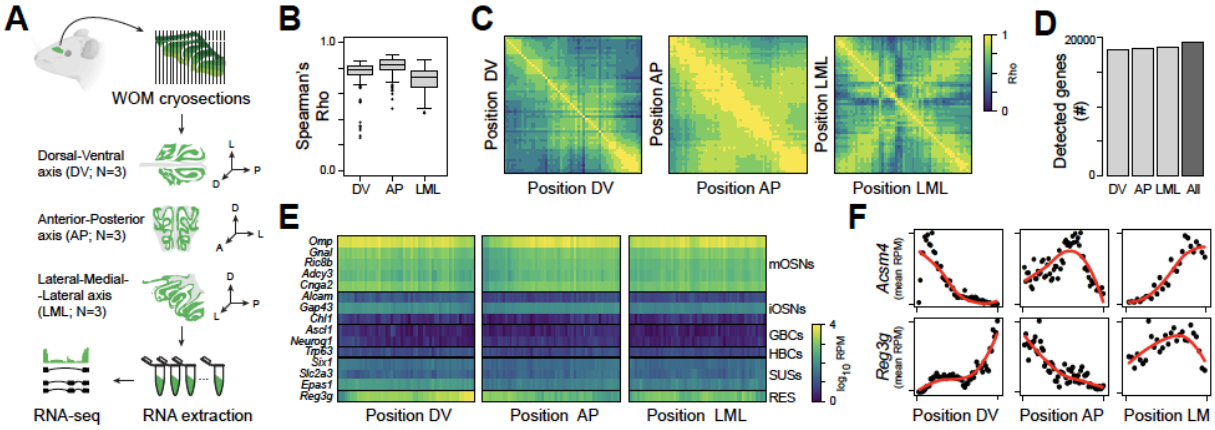
309 Discussion

310 Past studies yielded inconclusive and sometimes contradictory views on the basic logic
311 underlying the peripheral representation of smell, partly because the topographic
312 distribution of OSN subtypes and their receptive fields still remained vastly uncharted,
313 data on Olfr-ligand pairs was scarce, and the many pitfalls associated with electro-
314 olfactogram recordings used to study spatial patterns of odor recognition in the nose
315 (*26, 71, 72*).
316 Here, we combined RNA-seq and computational approaches that utilize unsupervised
317 and supervised machine learning methods to discover and quantitatively characterize
318 spatial expression patterns in the OM. We created the first 3D transcriptional map of
319 the mouse OM, which allowed us to spatially characterize thousands of genes,
320 including ~98% of the annotated *Olfrs*. We identified and validated by ISH several
321 new spatial marker genes, and a clustering analysis pinpointed the OM locations
322 where specific functions related to, e.g., the immune response, might be carried out.
323 We also identified five broad Olfr expression zones in the OM, which were
324 mathematically defined and used to decompose the expression patterns of all genes.

Our analysis enabled us to answer fundamental and longstanding questions about the rationale behind the spatial organization of the peripheral olfactory system. Specifically, we provide evidence to the hypothesis that the spatial zones increase the discriminatory power of the olfactory system by distributing receptors in the areas of the OM that are more likely to be reached by their cognate ligands, based on their solubility in mucus.

The quantitative framework we built for this dataset will facilitate interrogation of gene expression patterns via an online tool we provide, and help answer important questions on the spatial patterns in the nose. Moreover, our approach could be easily applied to spatial transcriptomic data collected from other tissues, or the same tissue across multiple developmental stages. Results from this study will also serve as a template to start answering other important questions about olfaction, such as whether Olfr spatial expression maps can also encode maps of odor perception. Because the general molecular mechanisms of olfaction, zonal organization of Olfrs, and components of olfactory perception are conserved in mammals (*73-77*), findings from our and other subsequent studies can likely be extrapolated to other mammals, including humans. Finally, the functional logic underlying the topographic organization of primary receptor neurons and their receptive fields in smell is now starting to be exposed.

345 # Figures



346 **FIGURE 1**

347 **Figure 1. Application of TOMO-seq to mouse OM.** (A) Experimental design.
348 TOMO-seq was performed on 9 tissue samples, from which 3 were sliced along the
349 Dorsal-Ventral axis, 3 along the Posterior-Anterior axis and 3 along the Lateral-
350 Medial-Lateral axis. (B) Boxplots showing the distributions of Spearman's correlation
351 coefficients (Rho) between replicates in each axis. (C) Heatmaps showing Spearman's
352 correlation between gene expression patterns at different positions along the three
353 axes. (D) Number of detected genes along each axis separately or across the whole
354 dataset. Genes were considered as detected when they had at least one normalized
355 count in at least 10% of the samples from one axis. (E) Heatmaps of $\log_{10}$ normalized
356 expression (after combining the three replicates per axis) of OM canonical markers
357 along the 3 axes (RPM = reads per million; mOSNs = mature Olfactory Sensory
358 Neurons; iOSNs = immature Olfactory Sensory Neurons; GBCs = Globose Basal
359 Cells; HBCs = Horizontal Basal Cells; SUSs = Sustentacular cells; RES = Respiratory
360 epithelium cells). (F) Normalized expression of canonical OM spatial marker genes
361 along the 3 axes. Red line showing fits with local polynomial models.

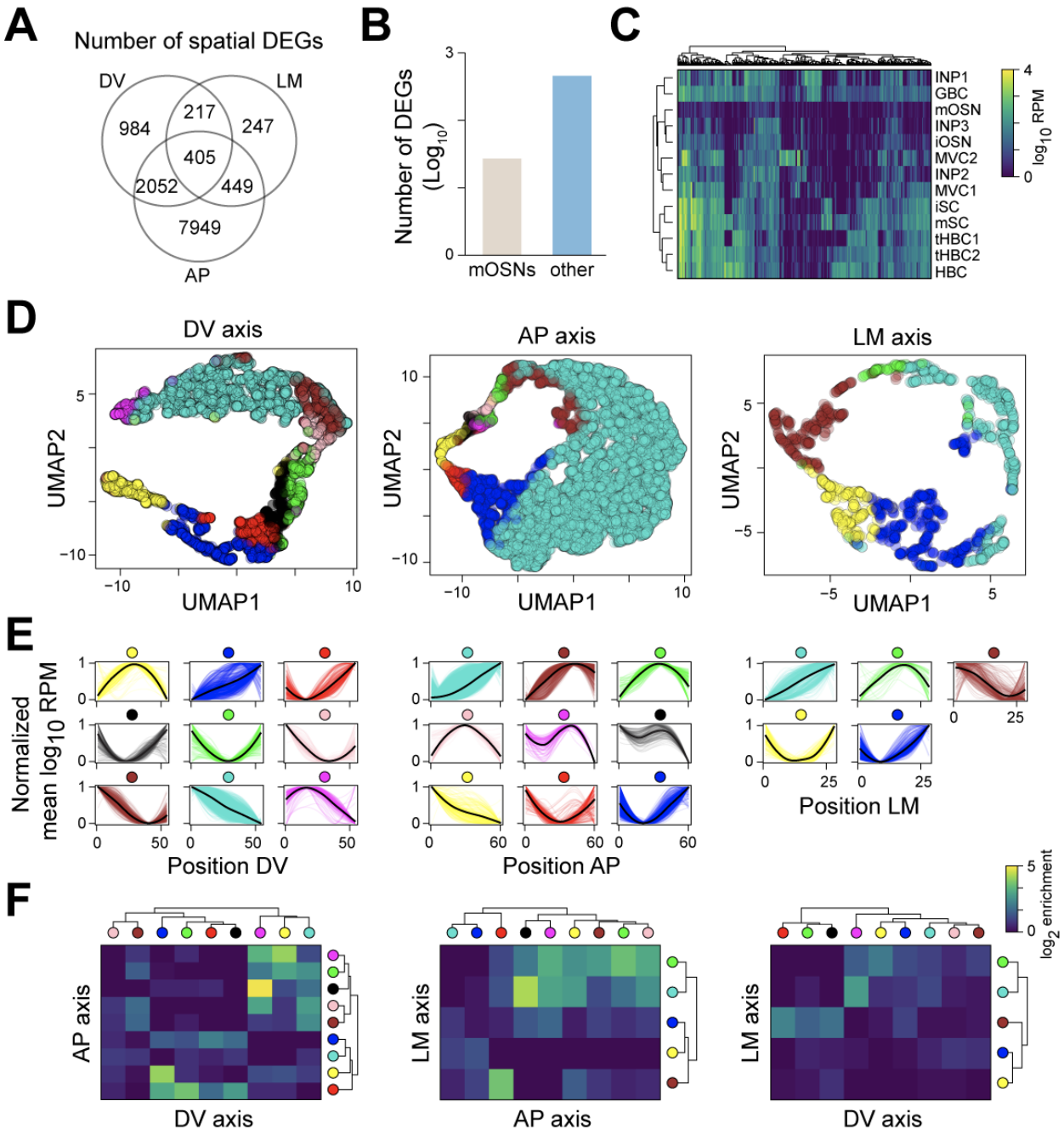

A
Number of spatial DEGs
DV
LM
AP
984
217
247
405
2052
449
7949
B
Number of DEGs (Log10)
mOSNs
other
C
INP1
GBC
mOSN
INP3
iOSN
MVC2
INP2
MVC1
iSC
mSC
tHBC1
tHBC2
HBC
log10 RPM
D
DV axis
AP axis
LM axis
UMAP1
UMAP2
E
Normalized mean log10 RPM
Position DV
Position AP
Position LM
F
log2 enrichment
AP axis
DV axis
LM axis

FIGURE 2

363 **Figure 2. Genes with non-random spatial patterns across different cell types in**
364 **the OM.** (A) Venn diagram showing the numbers of spatial differentially expressed
365 genes (DEGs) along each axis. (B) Bar plot showing the $\log_{10}$ number of spatial
366 DEGs that are mOSN-specific ("mOSNs"), or that are detected only in cell types
367 other than mOSNs ("other"). (C) Heatmap of log10 mean expression per cell type of
368 genes that are not expressed in mOSNs, but only in other OM cell types (INP =
369 Immediate Neuronal Precursors; GBC = Globose Basal Cells; mOSNs = mature
370 Olfactory sensory neurons; iOSNs = immature Olfactory Sensory Neurons; MVC =
371 Microvillous Cells; iSC = Immature Sustentacular Cells; mSC = Mature Sustentacular
372 Cells; HBCs = Horizontal Basal Cells). (D) UMAP plots of spatial DEGs along the
373 three axes. Each gene is colored according to the cluster it belongs to. (E) Normalized
374 average expression patterns of spatial DEGs clusters along the three axes. (F)
375 Heatmap showing the log2 enrichment over the random case for the intersection
376 between lists of genes belonging to different clusters (indicated by colored circles)
377 across pairs of axes.

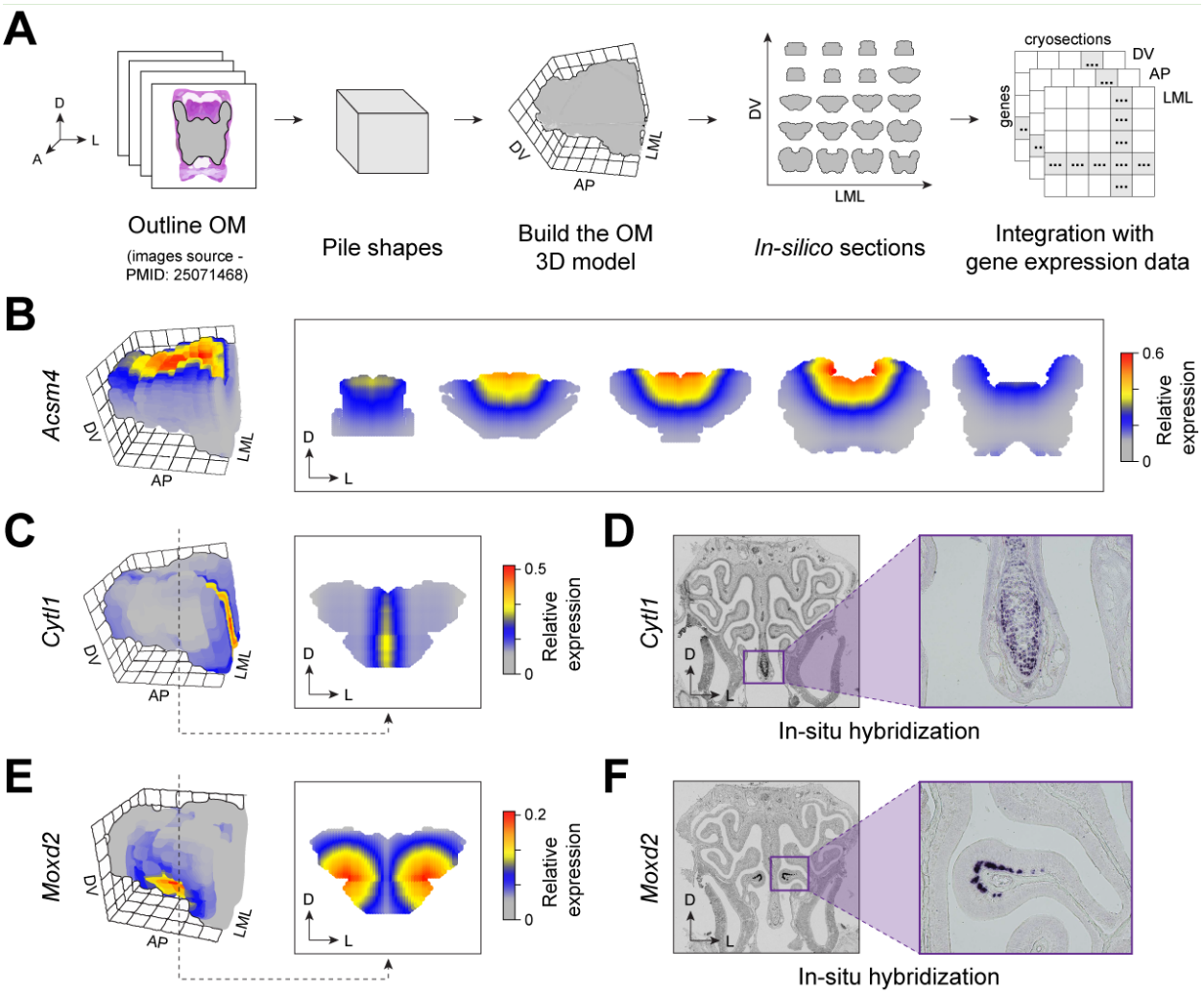


FIGURE 3

Figure 3. The 3D reconstruction of the OM. (A) Schematic of 3D shape reconstruction strategy. Images of 2D slices along the AP axis of the OM were piled together to build an in silico 3D model of OM, which can also be used to visualize in silico sections. This 3D model, together with the gene expression data along each axis, was the input of the iterative proportional fitting algorithm, which allowed us to estimate a 3D expression pattern for any gene. (B) Reconstruction of the 3D expression pattern of the gene *Acsm4* in the OM, visualized in 3D and in OM slices taken along the AP axis. (C) Reconstruction of the 3D expression pattern of the gene *Cytl1* in the OM. (D) In-situ hybridization experiment validating *Cytl1* spatial expression pattern reconstructed in panel D. (E) Reconstruction of the 3D expression pattern of the gene *Moxd2* in the OM. (F) In-situ hybridization experiment validating *Moxd2* spatial expression pattern reconstructed in panel E.

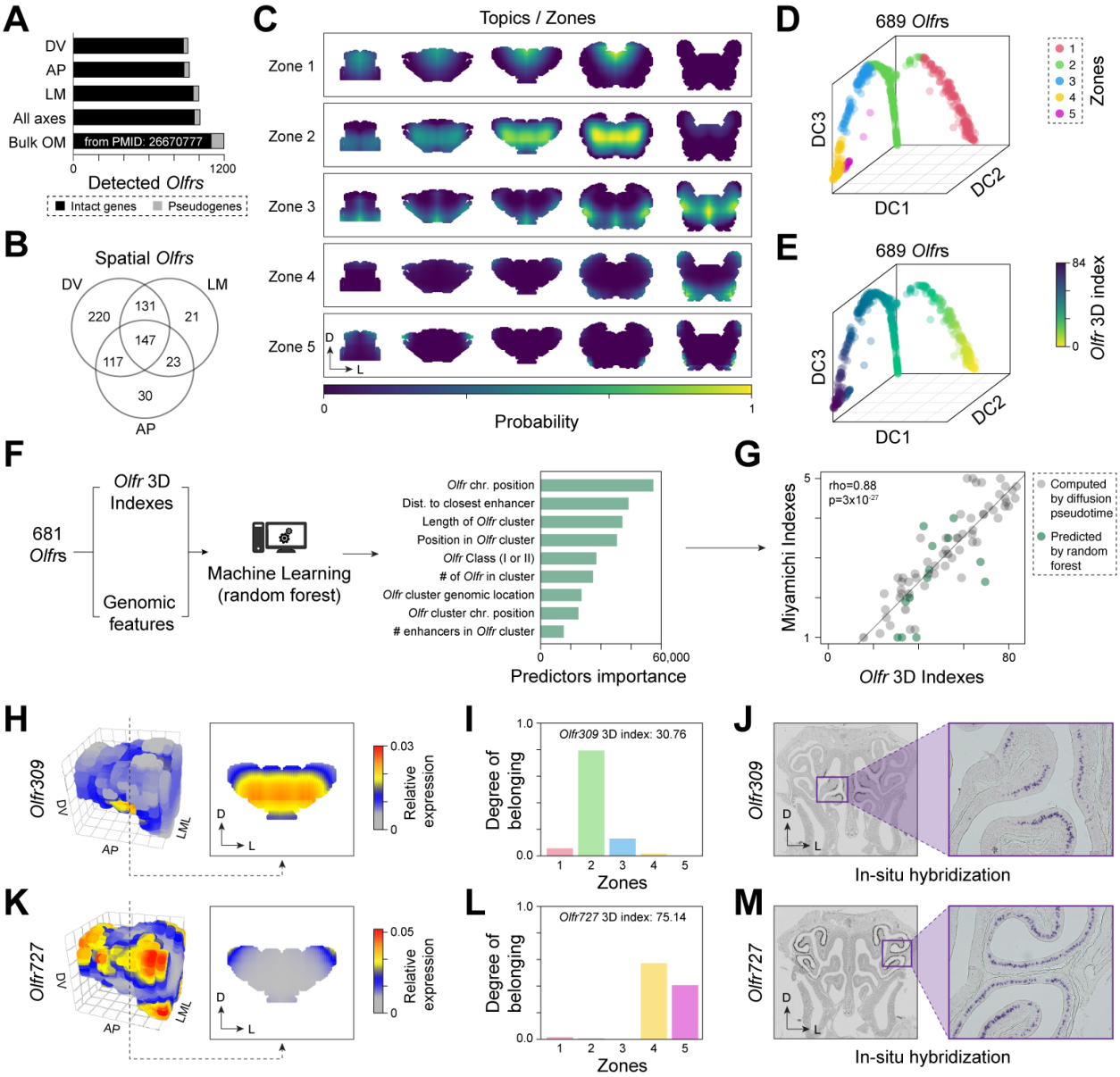

A
DV
AP
LM
All axes
Bulk OM
from PMID: 26670777
0
1200
Detected Olfrs
Intact genes
Pseudogenes
B
Spatial Olfrs
DV
LM
AP
220
131
21
147
117
23
30
C
Topics / Zones
Zone 1
Zone 2
Zone 3
Zone 4
Zone 5
D
L
0
1
Probability
D
689 Olfrs
DC3
DC1
DC2
1
2
3
4
5
Zones
E
689 Olfrs
DC3
DC1
DC2
84
0
Olfr 3D index
F
681 Olfrs
Olfr 3D Indexes
Genomic features
Machine Learning (random forest)
Olfr chr. position
Dist. to closest enhancer
Length of Olfr cluster
Position in Olfr cluster
Olfr Class (I or II)
# of Olfr in cluster
Olfr cluster genomic location
Olfr cluster chr. position
enhancers in Olfr cluster
0
60,000
Predictors importance
G
rho=0.88
p=3x10-27
Miyamichi Indexes
5
1
0
80
Olfr 3D Indexes
Computed by diffusion pseudotime
Predicted by random forest
H
Olfr309
DV
AP
LM/L
D
L
0.03
0
Relative expression
I
Olfr309 3D index: 30.76
Degree of belonging
1.0
0.0
1
2
3
4
5
Zones
J
Olfr309
D
L
In-situ hybridization
K
Olfr727
DV
AP
LM/L
D
L
0.05
0
Relative expression
L
Olfr727 3D index: 75.14
Degree of belonging
1.0
0.0
1
2
3
4
5
Zones
M
Olfr727
D
L
In-situ hybridization

FIGURE 4

392 **Figure 4. Zonal organization of the OM.** (A) Number of *Olfr* genes detected in our
393 data and in an OM bulk RNA-seq data (*6*) (B) Venn diagram of spatially differentially
394 expressed *Olfrs* per axis. (C) Visualization of the five zones estimated with a Latent
395 Dirichlet Allocation algorithm. The colors indicate the probability (scaled by its
396 maximum value) that a position belongs to a given zone. (D) Diffusion map of *Olfrs*.
397 Genes are colored based on the zone they fit in the most. DC, diffusion component.
398 (E) Same as panel D, with *Olfrs* colored by their 3D index. (F) We fitted a Random
399 Forest algorithm to the 3D indexes of 681 spatially differentially expressed *Olfr* using
400 nine genomic features as predictors. After training, the Random Forest was used to
401 predict the 3D indexes of 697 *Olfrs* that have too low levels in our data. (G) 3D
402 indexes versus the indexes of 80 Olfrs estimated in (*14*) from ISH data. Black circles
403 indicate *Olfrs* detected in our dataset; green circles are *Olfrs* whose indexes were
404 predicted with Random Forest. (H, K) Predicted expression patterns of *Olfr309* and
405 *Olfr727*. (I,L) Degrees of belonging for *Olfr309* and *Olfr727*. (J,M) In-situ hybridization
406 for *Olfr309* and *Olfr727*.

407
408
409
410
411
412

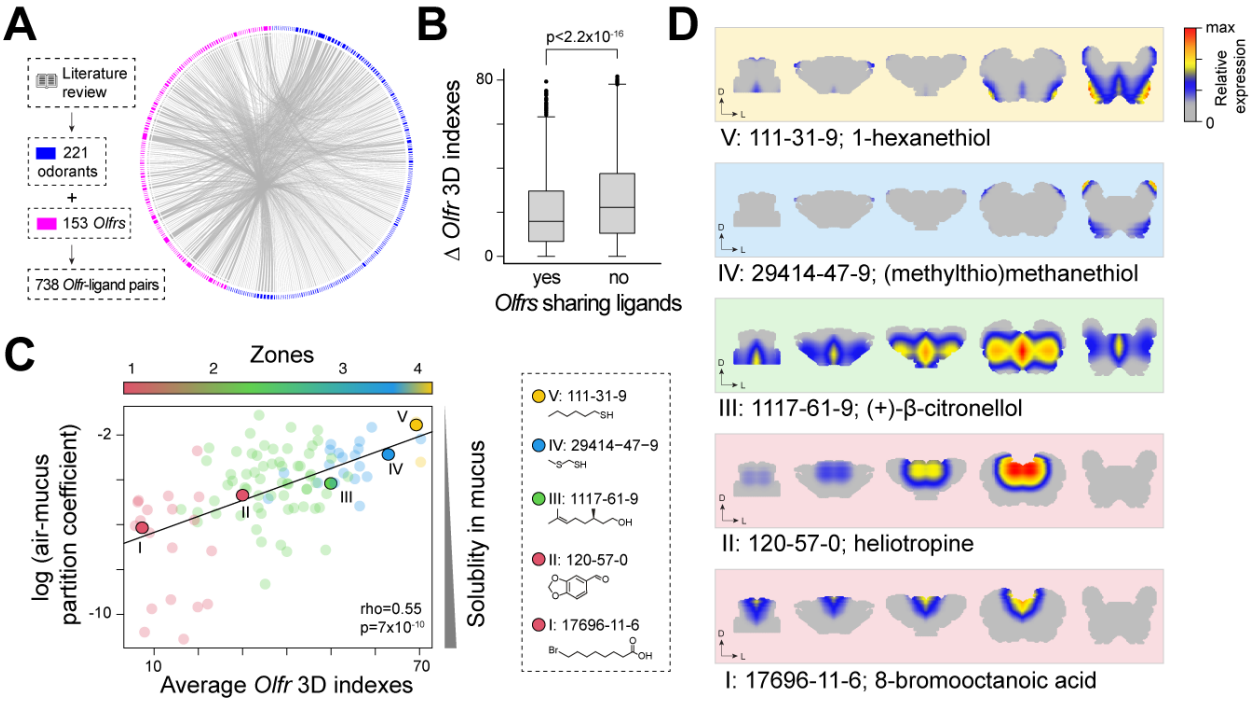


FIGURE 5

Figure 5. Physiological role of the zones. (A) Circular network illustrating the pairs of Olfrs and ligands that we found in literature. (B) Box plots showing the distributions of the absolute value of 3D index differences between pairs of Olfrs sharing at least one ligand versus pairs of Olfrs without cognate ligands in common. The difference between the two distributions is statistically significant ($p < 2.2x10^{-16}$, Wilcoxon Rank-Sum test). (C) Scatter plot illustrating the correlation between odorants' air / mucus partition coefficients and the average 3D indexes of their cognate Olfrs. Only odorants for which we know at least two cognate Olfrs (110) are used here. Odorants are colored according to the zone they belong to (defined as the zone with the highest average degree of belonging computed over all cognate receptors). Five odorants, indicated on the right-hand side, are highlighted in the plot by slightly larger circles. (D) Average expression patterns of the cognate Olfr for the five odorants highlighted in panel C.

429 # METHODS

430 ## Ethics statement

431 The use and care of animals used in this study was approved by the Internal Animal
432 Care and Use Committee (IACUC) of Monell Chemical Senses Center, by the IACUC
433 of the University of São Paulo, and by the Wellcome Trust Sanger Institute Animal
434 Welfare and Ethics Review Board in accordance with UK Home Office regulations,
435 the UK Animals (Scientific Procedures) Act of 1986.
436

437 ## Dissection of the olfactory mucosa, cryosections, and RNA-sequencing

438 All mice used were C57BL/6J (The Jackson Laboratory, Stock # 00664), 8-9 weeks
439 old and group-housed. The olfactory mucosa (OM) of 9 mice was carefully dissected,
440 and all the surrounding tissue (including glands and bone) removed. The OMs were
441 then embedded in OCT (Tissue Tek), immediately frozen in dry-ice and kept at -80ºC.
442 Each OM was then cryosectioned along each of the 3 cartesian axes: dorsal-ventral
443 (DV, N=3), anterior-posterior (AP, N=3), or lateral-medial-lateral (LML, N=3).
444 Every second cryosections (35 μm thick) was collected into 1.5 mL eppendorf tubes
445 containing 350 μl RLT Plus Buffer (Qiagen) supplemented with 1% 2-
446 mercaptoethanol, immediately frozen in dry-ice and kept at -80C until extraction.
447 RNA was extracted using the RNeasy Plus Micro Kit (Qiagen), together with a
448 genomic DNA eliminator column and a 30-minute incubation with DNAse I
449 (Qiagen). Reverse transcription and cDNA pre-amplification were performed using
450 the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Clontech/Takara).
451 cDNA was harvested and quantified with the Bioanalyzer DNA High-Sensitivity kit
452 (Agilent Technologies). Libraries were prepared using the Nextera XT DNA Sample
453 Preparation Kit and the Nextera Index Kit (Illumina). Multiplexed libraries were
454 pooled and paired-end 150-bp sequencing was performed on the Illumina HiSeq 4000
455 platform at Sidra Medicine, except for one library (DV-I) for which 125-bp paired-
456 end sequencing was performed on the Illumina HiSeq 2500 platform at the Wellcome
457 Sanger Institute. For the remaining eight libraries, 150-bp sequencing was performed
458 on the Illumina HiSeq 4000 platform. The raw data are available through
459 ArrayExpress under accession number E-MTAB-10211.
460

461 ## RNA-seq data mapping and gene counting

462 Reads were aligned to the mm10 mouse genome (release 99). The sequences of the
463 genes "*Xntrpc*" and "*Capn5*" were removed from the genome files as in (*6*). The
464 alignment was performed with the software STAR version 2.7.3a (*78*). Genome
465 indexes were generated using STAR --runMode genomeGenerate with default
466 parameters. Then, alignment of reads was performed with the following options: --
467 runThreadN 48 --outSAMunmapped Within --outFilterMultimapNmax 1000 --
468 outFilterMismatchNmax 4 --outFilterMatchNmin 100 --alignIntronMax 50000 --

469 alignMatesGapMax 50500 --outSAMstrandField intronMotif --outFilterType BySJout.
470 The resulting SAM files were converted to bam format and sorted using samtools
471 (version 0.1.19-44428cd)(*79*). The multi mapping reads were eliminated using the
472 same software (samtools view -q 255). Finally, the reads for each gene were counted
473 using htseq-count (version 0.11.2) with the options -m intersection-nonempty -s no -i
474 gene_name -r pos (*80*).
475

476 **Quality Control**

477 We excluded all the samples that fulfilled any of these criteria: they had less than 50%
478 mapped reads, less than 4,000 detected genes, more than 20% mitochondrial reads,
479 less than 10,000 total number of reads, or did not express any of the 3 canonical OSN
480 markers *Omp*, *Cnga* and *Gnal*. This resulted in ~51 good-quality sections along the DV
481 axis (~84% out of the collected sections), ~76 (~91% of total) along the AP axis and
482 ~59 (~93% of total) along the LML axis, as averaged across the three replicates per
483 axis.
484

485 **Data normalization**

486 Gene expression counts were normalized by reads-per-million (RPM), then genes
487 detected in only one replicate and genes that were detected in less than 10% of all
488 samples along one axis were eliminated. To check the similarity between replicates, we
489 calculated Spearman correlations between the transcriptional profiles of sections along
490 each axis (using the top 1000 Highly Variable Genes per axis). Close positions had the
491 most similar transcriptional profiles (Fig. 1C). Afterwards, the 3 replicates for each
492 axis were aligned as follows: the top 3,000 highly variable genes (HVGs) from each
493 replicate were identified using the method implemented in the scran library in R (*81*)
494 and the intersection of these 3 groups was used in the next steps. For the replicates'
495 alignment, we took as reference the replicate with the smallest number of slices. We
496 used a sliding window approach that identified the range of consecutive positions on
497 each replicate along which the average value of the Spearman's correlation coefficient
498 computed with the reference replicate over the HVG was maximum (mean
499 Spearman's Rho= 0.80, $p<0.05$). To mitigate batch effects, the level of every gene was
500 scaled in such a way that their average value in each replicate was equal to the average
501 calculated across all replicates. After this scaling transformation, the data was then
502 averaged between replicates. Once the 3 biological replicates were combined, we had
503 54 sections along the DV axis, 60 along the AP and 56 along the LML. Along the
504 LML axis a symmetric pattern of expression is expected around the central position,
505 where the septal bone is located. To confirm this in our data, first we identified the
506 central position by analyzing the expression pattern of neuronal markers like *Cnga2*,
507 *Omp* and *Gnal*, whose expression is lowest in the area around the septal bone. Indeed,
508 all three marker genes reach a minimum at the same position along the LML axis
509 (slice 28), which we considered to be the center. The expression patterns of ~90% of

genes on either side of the central position show a positive correlation, and ~70% reach statistical significance (Spearman's correlation computed on the highly variable genes having more than 50 normalized counts in at least 3 slices), further supporting the hypothesis of the bilateral symmetry. Hence, after replicates were averaged, LML axis was made symmetric averaging positions 1:28 and 56:29. Moreover, *Olfrs* were normalized by the geometric mean of neuronal markers *Omp*, *Gnal* and *Cnga2*, as done previously (*11*).

To verify the presence of a spatial signal, we calculated the Moran's I and the associated p-values for the top 100 Highly Variable genes along each axis using the "Moran.I" function from the "ape" library in R with default parameters (*82*). The p-values of the genes along each axis were combined with the Simes' method (*83*) using the function combinePValues from the scran R library (Fig. S1E).

Identification of differentially expressed genes and gene clustering

Before testing for differential expression along a given axis, we filtered out genes whose expression levels had low variability. To this aim, for each gene we estimated their highest and lowest expression by taking the average of its three highest and three lowest values respectively. Then, we considered for downstream analyses only the genes that meet either of these two criteria: the highest expression value is greater than or equal to 5 normalized counts and the fold-change between the highest and lowest value is greater than 2; or the difference between the highest and the lowest value is greater than or equal to 4 normalized counts.

The expression levels of the genes were binarized according to whether their value was higher or lower than their median expression along the axis. Finally, we used the "ts" function in R to transform the binarized expression values into time series objects, and we applied on them the Ljung-Box test (Box.test function in R with lag=(axis length)-10) which identifies genes with statistically significant autocorrelations, i.e., with non-random expression patterns along an axis. The resulting p-values were adjusted using the FDR method and genes with an FDR < 0.01 were considered as differentially expressed. For the next steps, the $\log_{10}$ normalized expression of differentially expressed genes along each axis was fitted with a local regression using the locfit function in the R library locfit (*84*). Smoothing was defined in the local polynomial model term of the locfit model using the function "lp" from the same library with the following parameters: nn = 1 (Nearest neighbor component of the smoothing parameter) and deg = 2 (degree of polynomial). The fitted expression values of these genes along each axis were normalized between 0 and 1. Clustering was performed separately for each axis on the fitted and normalized patterns of the differentially expressed genes. We used the R function "hclust" to perform hierarchical clustering on the gene expression patterns, with a Spearman's

550 correlation-based distance (defined as $\sqrt{0.5\ (1-\rho)}$) and the "average" aggregation
551 method. The number of clusters were defined with the cutreeDynamic function from
552 the dynamicTreeCut R library, with the parameters minClusterSize = 50, method =
553 "hybrid" and deepSplit = 0. To visualize the data in two dimensions, we applied the
554 UMAP dimensionality reduction algorithm (umap function in the R library umap with
555 default options; see Fig. 2D) (*55, 85*). To analyze the relationship between the
556 expression patterns of genes along different axes, we computed the intersections of
557 the gene clusters between any pair of axes. The expected number of elements in each
558 intersection under the assumption of independent sets is given by:

559 $$|A \cap B|_{exp} = \frac{|A|\ |B|}{|A \cup B|}$$

560 where A and B indicate the sets of genes in two clusters identified along two different
561 axes and $|\cdot|$ indicates the cardinality of a set (i.e., the number of its elements). The
562 ratio between the observed and the expected number of elements in the intersection
563 $|A \cap B|_{obs}\ /\ |A \cap B|_{exp}$ quantifies the enrichment/depletion of genes having a given
564 pair of patterns across two axes with respect to the random case. The $\log_2$ values of
565 $(1 + |A \cap B|_{obs}\ /\ |A \cap B|_{exp})$ are shown in Fig. 2F
566

567 **Combining Tomo-seq with single-cell RNA-seq data**

568 The TPM (transcripts per million)-normalized single cell RNA-seq (scRNA-seq) data
569 collected from mouse olfactory epithelium available from (*54*) was used to identify
570 cell-type specific genes. To this aim, we computed the average expression level for
571 each cell type in the scRNA-seq dataset for all the differentially expressed genes that
572 we identified in our TOMO-seq data. The genes with an average expression above
573 100 TPM in mOSNs and below 10 TPM in all other cell types were considered
574 mOSN-specific. Conversely, genes with an average expression above 100 TPM in any
575 of the non-mOSN cell types and below 10 in mOSNs were considered to be specific
576 for non-mOSN cells.
577

578 **Gene Ontology (GO) enrichment analysis**

579 GO Enrichment analyses were performed using the GOrilla online tool (http://cbl-
580 gorilla.cs.technion.ac.il) with the option "Two unranked lists of genes (target and
581 background lists)". For each axis, we used as background list the list of the genes we
582 tested.
583

584 **3D spatial reconstruction**

585 The olfactory mucosa shape was obtained from publicly available images of the
586 mouse nasal cavity along the posterior to the anterior axis published in (*58*). The area
587 of the slices corresponding to the OM was manually selected and images of their

silhouettes were made. Those images were then transformed into binary matrices having 1's in the area occupied by the OM and 0's in the remaining regions. The binary matrices were rescaled in order to match the spatial resolution in our dataset, which is composed of 54 voxels along the DV axis, 56 along the LML axis and 60 along the AP axis. Finally, matrices were piled in a 3D array in R to obtain an in-silico representation of the 3D shape of the OM, which, in total, was composed of 77,410 voxels. To perform the 3D reconstruction of the expression pattern for a given gene, first we normalized its expression levels by the volume of the slice at each corresponding position along the three axes, which was estimated using our 3D in silico representation of the OM. Then, we rescaled the data in such a way that the sum of the expression levels along each axis was equal to the average expression computed across the whole dataset. This rescaled dataset together with the binary matrix representing the 3D OM shape was used as input of the Iterative Proportional Fitting algorithm, which produced an estimation of the expression level of a gene in each voxel (*30*). Iterations stopped when the differences between the true and the reconstructed 1D values summed across the three axes was smaller than 1.

Definition of zones by topic modelling

In order to identify zones, we fitted a Latent Dirichlet Allocation (LDA) (*86*) algorithm to the 3D gene expression patterns (in log10 scale) of the differentially expressed *Olfrs* (689 *Olfrs* x 77,410 voxels).

The LDA algorithm was originally employed for document classification: based on the words included in each document, LDA can identify "topics", in which the documents can then be classified. Using this linguistic analogy, in our application of LDA, we considered the genes as "documents", and the spatial locations as "words", with the matrix of gene expression levels being the analogous of the "bag-of-word" matrix (*63*). In this representation, the zones are the equivalent of "topics", and they are automatically identified by LDA. We used the LDA implementation included in the R package "Countclust (*87*) ", developed by based on the "maptpx" library (*88*), which performs a maximum a posteriori estimation to for model fitting. LDA was run for all possible numbers of topics $K \in [2,9]$. The following parameters were chosen: convergence tolerance = 0.1; max time optimization step = 180 seconds; n_init = 3. For each number of topics k, three independent runs were performed with different starting points, in order to avoid biases due to the choice of the initial condition. We estimated the number of topics by computing the log-likelihood for each value of $K \in [2,9]$. As seen in fig. S3A, while the log-likelihood is a monotonically increasing function of the number of topic (as expected), for a number of topics around ~5 it shows a "knee" and starts to increase more slowly. This suggests that ~5 is the optimal number of topics needed to describe the complexity of the data without overfitting. Hence, we fix a number of topics equal to 5; however, we also verified

628 that all our conclusions remain substantially unaffected if a different number of topics
629 (e.g., 4 or 6) is chosen.
630 After running LDA with K=5, we retrieved the model output, which consists of two
631 probability distributions: the first is P(position| k) with k ∈ [1,5], which is the
632 conditional probability distribution defining the topic k; the second probability
633 distribution is P(k | gene), namely the probability distribution that quantifies the
634 degrees of belonging of a given gene to the topics k∈[1,5]. With these probability
635 distributions, we can identify the spatial positions that form each topic and how the
636 different topics can be combined to generate the spatial expression pattern of each
637 gene.
638 Being a generative model, once trained, LDA can also decompose into topics the
639 spatial expression patterns of genes that were not used during the training procedure.
640 We exploited this feature of LDA to estimate the degrees of belonging of non-
641 olfactory receptor genes. To this aim, we utilized an algorithm based on the python
642 gensim library Lda.Model.inference function (*89*), using as input the estimated
643 probability distribution P(position | k) with k ∈ [1,5]. The model fitting was
644 performed using the Open Computing Cluster for Advanced data Manipulation
645 (OCCAM), the High-Performance Computer designed and managed in collaboration
646 between the University of Torino and the Torino division of the Istituto Nazionale di
647 Fisica Nucleare (*90*).
648

649 **Definition of *Olfr* 3D indexes via diffusion pseudo-time**

650 As explained in the section above, we can describe the spatial expression pattern of
651 each gene through a set of five numbers, which represent the degrees of belonging to
652 the five topics identified by LDA. We applied a diffusion map (*64*) to the degrees of
653 belonging of the *Olfrs* to visualize them in two dimensions by using the
654 "DiffusionMap" function from the "destiny" R package (*91*) (with
655 distance="rankcor" and default parameters). In this two-dimensional map, the *Olfrs*
656 are approximately distributed along a curve that joins the most dorsal/medial genes
657 (those in zones 1-2) with those that are more ventral/lateral (zones 3-5). To track the
658 position of the genes along this curve, we computed a diffusion pseudo-time (DPT)
659 coordinate (*65*) with the "DPT" function from the "destiny" R package (taking as
660 starting point the gene with the smallest first diffusion component among the genes
661 suggested by the function find_tips from the same package). In order to make the
662 indexes go from Dorsal to Ventral, as in previous studies (*14*), we reversed the order
663 of the DPT coordinates by substracting the maximum coordinate from all coordinates
664 and multiplying them by (-1). By doing this, we obtained for each *Olfr* an index, which
665 we called 3D index, representing its spatial expression pattern in the 3D space: more
666 dorsal/medial genes (zones 1-2) have smaller 3D indexes than *Olfrs* expressed in the
667 ventral/lateral regions (zones 3-5).

668

669 **Prediction of zone index for undetected *Olfrs* with Random Forest**

670 We fitted a Random Forest model to the 3D indexes of 681 of the 689 *Olfrs* we
671 characterized with our dataset (i.e., those that are located in genomic clusters). The
672 following nine features of each *Olfr* were used as predictors: genomic position (i.e.,
673 gene starting position divided by chromosome length); genomic cluster; genomic
674 cluster length; number of *Olfrs* in the genomic cluster; number of enhancers in the
675 genomic cluster; cluster position (i.e., starting position of the cluster divided by the
676 chromosome length); distance to the closest enhancer; gene position within the cluster
677 (i.e., the distance of the gene starting position from the end of the cluster divided by
678 the cluster length); and phylogenetic class.
679 These features were computed using the mm10 mouse genome in Biomart (*92*), while
680 the list of enhancers and the genomic clusters assigned to each *Olfr* were taken from
681 (*69*). The Random Forest model was fitted with the function "randomForest" (R
682 library "randomForest" (*93*), with option "na.action = na.omit"). Afterwards, we
683 performed a cross-validation test with the function "rf.crossValidation" from the
684 "rfUtilities" package (*94*) with default parameters. Over 100 cross-validation iterations,
685 the root mean square error (RMSE) was $\lesssim$10% of the mean 3D index. The feature
686 importance was computed with the "importance" function from the randomForest
687 library with default parameters. Finally, the Random Forest model trained on the 681
688 *Olfrs* was used to predict the 3D indexes of 697 *Olfrs* that were too lowly expressed or
689 were undetected in our dataset. Overall, we were able to compute or predict with
690 Random Forest a 3D index for all the *Olfrs* annotated in the mouse genome, except
691 for 28 of them that do not have any genomic cluster assigned.

692

693 **Odorant information and Olfr-ligand pairs**

694 All odorant structures and associated CAS numbers were retrieved from either Sigma-
695 Aldrich (www.sigmaaldrich.com) or PubChem (https://pubchem.ncbi.nlm.nih.gov).
696 A comprehensive catalog of the cognate mouse Olfr-ligand pairs was collected (last
697 update: March 2021) by combining data from the ODORactor database (*95*) and
698 additional literature sources (*4, 7, 74, 96-123*).
699 This catalog includes 738 Olfr-ligand interactions for a total of 153 Olfrs and 221
700 odorants. These 153 *Olfrs* include 100 spatial *Olfrs* in our dataset and for which we
701 have 3D indexes, and 49 additional *Olfrs* with predicted 3D indexes (see above). Next,
702 we checked whether *Olfrs* pairs sharing at least one cognate ligand have more similar
703 spatial expression patterns than pairs not sharing ligands. To do this, we computed
704 the absolute values of the differences between the 3D indexes (Δ) of 1706 pairs of
705 ORs sharing at least one odorant and 9922 pairs of ORs that are known to bind to
706 different odorants (Fig. 5B). The two sets of Δ values were significantly different
707 (Mann-Whitney U test, p-value $< 2.2\times10^{-16}$).

708

709 **Estimation of the air/mucus partition coefficients**

710 For each odorant, we estimated the air/mucus partition coefficients (K_{am}) as done
711 previously (*21, 70*). Briefly, we calculated the air/water partition coefficients (K_{aw}) for
712 each odorant from the Henry's Law constants obtained using the HENRYWIN
713 model in the US EPA Estimation Program Interface (EPI) Suite (version 4.11;
714 www.epa.gov/oppt/exposure/pubs/episuite.htm). Then, we computed the air/mucus
715 partition coefficients (K_{am}) according to the formula:

716 $$Log(K_{am}) = 0.524 \cdot Log(K_{aw}) \cdot Log(K_{ow})$$

717 where K_{ow} indicates the octanol/water partition coefficient, which were obtained
718 using the KOWWIN model in the EPI Suite.

719

720 Finally, we plotted in Fig. 5C the Log(K_{am}) of each odorant versus the mean 3D
721 index of the cognate *Olfrs*, using only the 110 odorants for which we could find two
722 or more cognate receptors. These two quantities have a Spearman's correlation
723 coefficient equal to 0.55 (p-value=$7x10^{-10}$). The correlation remained significant even
724 when we removed *Olfrs* whose indexes were predicted with the Random Forest
725 (Spearman's correlation = 0.69, p-value= $2x10^{-3}$), or when only class II *Olfrs* were
726 included (Spearman's correlation = 0.5, p-value= $1.3x10^{-7}$).

727

728 ***In-situ* hybridization**

729 *In-situ* hybridization was basically performed as previously described (Ibarra-Soria et
730 al. 2017). 12-week-old male C57BL/6J mice were perfused with 4% paraformaldehyde
731 and decalcified in RNase-free 0.45M EDTA solution (in 1x PBS) for two weeks.
732 Decalcified heads were cryoprotected in RNase-free 30% sucrose solution (1x PBS),
733 dried, embedded in OCT embedding medium, and frozen at -80°C. Sequential 16 mm
734 sections were prepared with a cryostat and the sections were hybridized to
735 digoxigenin-labeled cRNA probes prepared from the different genes using the
736 following oligonucleotides: Cytl (5'–AAAGACACTACCTCTGTTGCTGCTG – 3'
737 and 5' – GTAAGCAGAGACCAGAAAGAAGAGTG – 3'), Moxd2 (5' –
738 TGTACCTTTCTCCCACTCCCTATTGTC – 3' and 5'–
739 CCCATGCAACTGGAGATGTTAATTCTG – 3'), Olfr309 (5'–
740 TACAATGGCCTATGACCGCTATGTG – 3' and 5' –
741 TCCTGACTGCATCTCTTTGTTCCTG – 3') , Spen (5'–
742 GGTGGGAAACTTACCGGAGAACGTG – 3' and 5' –
743 TGCTGCTGATGGAGTCACTACTG – 3'), Olfr727 (5' –
744 CGCTATGTTGCAATATGCAAGCCTC – 3' and 5'–
745 GCTTTGACATTGCTGCTTTCACCTC – 3'). The PCR products were cloned into
746 pGEM-T Easy vector and the probes were obtained by in vitro transcription of the

747 plasmids, using SP6 or T7 RNA Polymerases (Ambion) and DIG RNA Labeling mix
748 (Roche).
749

750 **Data availability**

751 The processed data can be visualized and downloaded from the website
752 http://atlas3dnose.helmholtz-muenchen.de:3838/atlas3Dnose.

753 **Acknowledgments**

754 We would like to thank Maria-Elena Torres-Padilla, Bob Datta, and the members of
755 the Saraiva and Scialdone Labs for the constructive feedback, and helpful comments.
756 We are thankful to Giorgia Greco for insightful discussions about chemical properties
757 of ligands. This work was supported by the Helmholtz Association (Scialdone Lab)
758 and Sidra Medicine (SDR400040).
759

760 **Author contributions**

761 M.L.R.T.S. analyzed data and wrote the initial version of the manuscript.
762 E.A.M. performed the RNA-seq experiments and analyzed data.
763 T.S.N., L.S.M., and S.L. performed experiments.
764 M.M., S.S.Y.H., J.D.M., F.V., M.C., and M.O. analyzed data.
765 E.G., J.R., D.W.L., and B.M. analyzed data, and helped write the manuscript.
766 A.S. and L.R.S. conceived and supervised the project, analyzed data, and wrote the
767 final version of the manuscript.
768
769

References

1. E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, A. Hudspeth, *Principles of neural science*. (McGraw-Hill, New York, ed. 5th, 2013), pp. l, 1709 p.
2. L. Buck, R. Axel, A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-187 (1991).
3. A. Chess, I. Simon, H. Cedar, R. Axel, Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823-834 (1994).
4. B. Malnic, J. Hirono, T. Sato, L. B. Buck, Combinatorial receptor codes for odors. *Cell* **96**, 713-723 (1999).
5. N. K. Hanchate *et al.*, Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* **350**, 1251-1255 (2015).
6. L. R. Saraiva *et al.*, Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Scientific reports* **5**, 18178 (2015).
7. K. Nara, L. R. Saraiva, X. Ye, L. B. Buck, A large-scale analysis of odor coding in the olfactory epithelium. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 9179-9191 (2011).
8. P. Mombaerts *et al.*, Visualizing an olfactory sensory map. *Cell* **87**, 675-686 (1996).
9. K. J. Ressler, S. L. Sullivan, L. B. Buck, Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245-1255 (1994).
10. R. Vassar *et al.*, Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981-991 (1994).
11. X. Ibarra-Soria *et al.*, Variation in olfactory neuron repertoires is genetically controlled and environmentally modulated. *eLife* **6**, (2017).
12. K. J. Ressler, S. L. Sullivan, L. B. Buck, A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597-609 (1993).
13. R. Vassar, J. Ngai, R. Axel, Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309-318 (1993).
14. K. Miyamichi, S. Serizawa, H. M. Kimura, H. Sakano, Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **25**, 3586-3592 (2005).
15. J. Strotmann, I. Wanner, J. Krieger, K. Raming, H. Breer, Expression of odorant receptors in spatially restricted subsets of chemosensory neurones. *Neuroreport* **3**, 1053-1056 (1992).
16. B. Zapiec, P. Mombaerts, The Zonal Organization of Odorant Receptor Gene Choice in the Main Olfactory Epithelium of the Mouse. *Cell reports* **30**, 4220-4234 e4225 (2020).
17. E. Bashkirova *et al.*, Homeotic Regulation of Olfactory Receptor Choice via NFI-dependent Heterochromatic Silencing and Genomic Compartmentalization. *bioRxiv*, 2020.2008.2030.274035 (2020).
18. D. M. Coppola, C. T. Waggener, S. M. Radwani, D. A. Brooks, An electroolfactogram study of odor response patterns from the mouse olfactory

816 epithelium with reference to receptor zones and odor sorptiveness. *J*
817 *Neurophysiol* **109**, 2179-2191 (2013).
818 19. M. Ma, G. M. Shepherd, Functional mosaic organization of mouse olfactory
819 receptor neurons. *Proceedings of the National Academy of Sciences of the*
820 *United States of America* **97**, 12869-12874 (2000).
821 20. M. M. Mozell, The spatiotemporal analysis of odorants at the level of the olfactory
822 receptor sheet. *J Gen Physiol* **50**, 25-41 (1966).
823 21. J. W. Scott, L. Sherrill, J. Jiang, K. Zhao, Tuning to odor solubility and sorption
824 pattern in olfactory epithelial responses. *The Journal of neuroscience : the official*
825 *journal of the Society for Neuroscience* **34**, 2025-2036 (2014).
826 22. T. A. Schoenfeld, T. A. Cleland, Anatomical contributions to odorant sampling
827 and representation in rodents: zoning in on sniffing behavior. *Chemical senses*
828 **31**, 131-144 (2006).
829 23. L. Secundo, K. Snitz, N. Sobel, The perceptual logic of smell. *Curr Opin*
830 *Neurobiol* **25**, 107-115 (2014).
831 24. D. M. Coppola, E. Fitzwater, A. D. Rygg, B. A. Craven, Tests of the
832 chromatographic theory of olfaction with highly soluble odors: a combined
833 electro-olfactogram and computational fluid dynamics study in the mouse. *Biol*
834 *Open* **8**, (2019).
835 25. T. Abaffy, A. R. Defazio, The location of olfactory receptors within olfactory
836 epithelium is independent of odorant volatility and solubility. *BMC research notes*
837 **4**, 137 (2011).
838 26. S. M. Kurian *et al.*, Odor coding in the mammalian olfactory epithelium. *Cell*
839 *Tissue Res* **383**, 445-456 (2021).
840 27. M. Asp, J. Bergenstrahle, J. Lundeberg, Spatially Resolved Transcriptomes-Next
841 Generation Tools for Tissue Exploration. *Bioessays* **42**, e1900221 (2020).
842 28. M. Asp *et al.*, A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of
843 the Developing Human Heart. *Cell* **179**, 1647-1660 e1619 (2019).
844 29. G. Peng *et al.*, Spatial Transcriptome for the Molecular Annotation of Lineage
845 Fates and Cell Identity in Mid-gastrula Mouse Embryo. *Dev Cell* **36**, 681-697
846 (2016).
847 30. J. P. Junker *et al.*, Genome-wide RNA Tomography in the zebrafish embryo. *Cell*
848 **159**, 662-675 (2014).
849 31. K. Achim *et al.*, High-throughput spatial mapping of single-cell RNA-seq data to
850 tissue of origin. *Nature biotechnology* **33**, 503-509 (2015).
851 32. C. Schmal, J. Myung, H. Herzel, G. Bordyugov, Moran's I quantifies spatio-
852 temporal pattern formation in neural imaging data. *Bioinformatics* **33**, 3072-3079
853 (2017).
854 33. F. Gussing, S. Bohm, NQO1 activity in the main and the accessory olfactory
855 systems correlates with the zonal topography of projection maps. *Eur J Neurosci*
856 **19**, 2511-2518 (2004).
857 34. Y. Oka *et al.*, O-MACS, a novel member of the medium-chain acyl-CoA
858 synthetase family, specifically expressed in the olfactory epithelium in a zone-
859 specific manner. *Eur J Biochem* **270**, 1995-2004 (2003).

35. M. Alenius, S. Bohm, Identification of a novel neural cell adhesion molecule-related gene with a potential role in selective axonal projection. *J Biol Chem* **272**, 26083-26086 (1997).
36. T. T. Yu *et al.*, Differentially expressed transcripts from phenotypically identified olfactory sensory neurons. *J Comp Neurol* **483**, 251-262 (2005).
37. S. D. Liberles, L. B. Buck, A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645-650 (2006).
38. R. Pacifico, A. Dewan, D. Cawley, C. Guo, T. Bozza, An olfactory subsystem that mediates high-sensitivity detection of volatile amines. *Cell reports* **2**, 76-88 (2012).
39. H. J. Fulle *et al.*, A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3571-3575 (1995).
40. P. L. Greer *et al.*, A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction. *Cell* **165**, 1734-1748 (2016).
41. J. H. Cho, M. Lepine, W. Andrews, J. Parnavelas, J. F. Cloutier, Requirement for Slit-1 and Robo-2 in zonal segregation of olfactory sensory neuron axons in the main olfactory bulb. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 9094-9104 (2007).
42. J. F. Cloutier *et al.*, Neuropilin-2 mediates axonal fasciculation, zonal segregation, but not axonal convergence, of primary accessory olfactory neurons. *Neuron* **33**, 877-892 (2002).
43. Y. Yoshihara *et al.*, OCAM: A new member of the neural cell adhesion molecule family related to zone-to-zone projection of olfactory and vomeronasal axons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **17**, 5830-5842 (1997).
44. I. Tietjen, J. Rihel, C. G. Dulac, Single-cell transcriptional profiles and spatial patterning of the mammalian olfactory epithelium. *Int J Dev Biol* **49**, 201-207 (2005).
45. I. Tietjen *et al.*, Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161-175 (2003).
46. S. S. Wang, J. W. Lewcock, P. Feinstein, P. Mombaerts, R. R. Reed, Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. *Development* **131**, 1377-1388 (2004).
47. E. M. Norlin *et al.*, Evidence for gradients of gene expression correlating with zonal topography of the olfactory sensory map. *Mol Cell Neurosci* **18**, 283-295 (2001).
48. D. M. Juilfs *et al.*, A subset of olfactory neurons that selectively express cGMP-stimulated phosphodiesterase (PDE2) and guanylyl cyclase-D define a unique olfactory signal transduction pathway. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3388-3395 (1997).
49. P. M. Heron, A. J. Stromberg, P. Breheny, T. S. McClintock, Molecular events in the cell types of the olfactory epithelium during adult neurogenesis. *Molecular brain* **6**, 49 (2013).

50. C. E. Peluso, W. Jang, U. C. Drager, J. E. Schwob, Differential expression of components of the retinoic acid signaling pathway in the adult mouse olfactory epithelium. *J Comp Neurol* **520**, 3707-3726 (2012).
51. G. K. Whitby-Logan, M. Weech, E. Walters, Zonal expression and activity of glutathione S-transferase enzymes in the mouse olfactory mucosa. *Brain Res* **995**, 151-157 (2004).
52. A. Miyawaki, H. Homma, H. Tamura, M. Matsui, K. Mikoshiba, Zonal distribution of sulfotransferase for phenol in olfactory sustentacular cells. *EMBO J* **15**, 2050-2055 (1996).
53. C. D. Duggan, S. DeMaria, A. Baudhuin, D. Stafford, J. Ngai, Foxg1 is required for development of the vertebrate olfactory system. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 5229-5239 (2008).
54. R. B. Fletcher *et al.*, Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell* **20**, 817-830 e818 (2017).
55. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, (2018).
56. V. Vedin, M. Molander, S. Bohm, A. Berghard, Regional differences in olfactory epithelial homeostasis in the adult mouse. *J Comp Neurol* **513**, 375-384 (2009).
57. M. D. Nickell, P. Breheny, A. J. Stromberg, T. S. McClintock, Genomics of mature and immature olfactory sensory neurons. *J Comp Neurol* **520**, 2608-2629 (2012).
58. A. W. Barrios, G. Nunez, P. Sanchez Quinteiro, I. Salazar, Anatomy, histochemistry, and immunohistochemistry of the olfactory subsystems in mice. *Front Neuroanat* **8**, 63 (2014).
59. S. E. Fienberg, An Iterative Procedure for Estimation in Contingency Tables. *The Annals of Mathematical Statistics* **41**, 907-917, 911 (1970).
60. S. Zhu *et al.*, Protein Cytl1: its role in chondrogenesis, cartilage homeostasis, and disease. *Cell Mol Life Sci* **76**, 3515-3523 (2019).
61. Y. Shin, Y. Won, J. I. Yang, J. S. Chun, CYTL1 regulates bone homeostasis in mice by modulating osteogenesis of mesenchymal stem cells and osteoclastogenesis of bone marrow-derived macrophages. *Cell Death Dis* **10**, 47 (2019).
62. C. J. Goh, D. Choi, D. B. Park, H. Kim, Y. Hahn, MOXD2, a Gene Possibly Associated with Olfaction, Is Frequently Inactivated in Birds. *PloS one* **11**, e0152431 (2016).
63. L. Liu, L. Tang, W. Dong, S. Yao, W. Zhou, An overview of topic modeling and its current applications in bioinformatics. *Springerplus* **5**, 1608 (2016).
64. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989-2998 (2015).
65. L. Haghverdi, M. Buttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845-848 (2016).
66. L. Tan, X. S. Xie, A Near-Complete Spatial Map of Olfactory Receptors in the Mouse Main Olfactory Epithelium. *Chemical senses* **43**, 427-432 (2018).

948 67. X. Zhang *et al.*, High-throughput microarray detection of olfactory receptor gene
949 expression in the mouse. *Proceedings of the National Academy of Sciences of*
950 *the United States of America* **101**, 14168-14173 (2004).
951 68. S. L. Sullivan, M. C. Adamson, K. J. Ressler, C. A. Kozak, L. B. Buck, The
952 chromosomal distribution of mouse odorant receptor genes. *Proceedings of the*
953 *National Academy of Sciences of the United States of America* **93**, 884-888
954 (1996).
955 69. K. Monahan *et al.*, Cooperative interactions enable singular olfactory receptor
956 expression in mouse olfactory neurons. *Elife* **6**, (2017).
957 70. A. D. Rygg, B. Van Valkenburgh, B. A. Craven, The Influence of Sniffing on
958 Airflow and Odorant Deposition in the Canine Nasal Cavity. *Chem Senses* **42**,
959 683-698 (2017).
960 71. J. W. Scott, P. E. Scott-Johnson, The electroolfactogram: a review of its history
961 and uses. *Microsc Res Tech* **58**, 152-160 (2002).
962 72. L. Secundo *et al.*, Individual olfactory perception reveals meaningful nonolfactory
963 genetic information. *Proceedings of the National Academy of Sciences of the*
964 *United States of America* **112**, 8750-8755 (2015).
965 73. Daniel M. Bear, J.-M. Lassance, Hopi E. Hoekstra, Sandeep R. Datta, The
966 Evolving Neural and Genetic Architecture of Vertebrate Olfaction. *Current*
967 *Biology* **26**, R1039-R1049 (2016).
968 74. L. R. Saraiva *et al.*, A transcriptomic atlas of mammalian olfactory mucosae
969 reveals an evolutionary influence on food odor detection in humans. *Sci Adv* **5**,
970 eaax0396 (2019).
971 75. L. F. Horowitz, L. R. Saraiva, D. Kuang, K. H. Yoon, L. B. Buck, Olfactory
972 receptor patterning in a higher primate. *J Neurosci* **34**, 12241-12252 (2014).
973 76. J. E. R. Octura, K. I. Maeda, Y. Wakabayashi, Structure and zonal expression of
974 olfactory receptors in the olfactory epithelium of the goat, Capra hircus. *J Vet*
975 *Med Sci* **80**, 913-920 (2018).
976 77. D. Manoel *et al.*, Deconstructing the mouse olfactory percept through an
977 ethological atlas. *Curr Biol*, (2021).
978 78. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-
979 21 (2013).
980 79. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
981 **25**, 2078-2079 (2009).
982 80. S. Anders, P. T. Pyl, W. Huber, *HTSeq - A Python framework to work with high-*
983 *throughput sequencing data*. (2014).
984 81. A. T. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level
985 analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122
986 (2016).
987 82. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and
988 Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
989 83. R. J. Simes, An Improved Bonferroni Procedure for Multiple Tests of
990 Significance. *Biometrika* **73**, 751-754 (1986).
991 84. C. Loader. (2007).
992 85. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and
993 Projection for Dimension Reduction. 2018.

994 86. D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation. *J. Mach. Learn. Res.*
995 **3**, 993–1022 (2003).
996 87. K. K. Dey, C. J. Hsiao, M. Stephens, Visualizing the structure of RNA-seq
997 expression data using grade of membership models. *PLoS Genet* **13**, e1006599
998 (2017).
999 88. M. Taddy, paper presented at the Proceedings of the Fifteenth International
1000 Conference on Artificial Intelligence and Statistics, Proceedings of Machine
1001 Learning Research, 2012.
1002 89. R. Rehurek, P. Sojka, paper presented at the Proceedings of the LREC 2010
1003 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, 2010.
1004 90. M. Aldinucci *et al.*, OCCAM: a flexible, multi-purpose and extendable HPC
1005 cluster. *Journal of Physics: Conference Series* **898**, 082039 (2017).
1006 91. P. Angerer *et al.*, destiny: diffusion maps for large-scale single-cell data in R.
1007 *Bioinformatics* **32**, 1241-1243 (2016).
1008 92. R. J. Kinsella *et al.*, Ensembl BioMarts: a hub for data retrieval across taxonomic
1009 space. *Database (Oxford)* **2011**, bar030 (2011).
1010 93. A. Liaw, M. Wiener, Classification and Regression by randomForest. *R News* **2**,
1011 18--22 (2002).
1012 94. T. A. Rather, S. Kumar, J. A. Khan, Multi-scale habitat modelling and predicting
1013 change in the distribution of tiger and leopard using random forest algorithm.
1014 *Scientific reports* **10**, 11473 (2020).
1015 95. X. Liu *et al.*, ODORactor: a web server for deciphering olfactory coding.
1016 *Bioinformatics* **27**, 2302-2303 (2011).
1017 96. B. Malnic, P. A. Godfrey, L. B. Buck, The human olfactory receptor gene family.
1018 *Proceedings of the National Academy of Sciences of the United States of*
1019 *America* **101**, 2584-2589 (2004).
1020 97. T. Bozza, P. Feinstein, C. Zheng, P. Mombaerts, Odorant receptor expression
1021 defines functional units in the mouse olfactory system. *The Journal of*
1022 *neuroscience : the official journal of the Society for Neuroscience* **22**, 3033-3043
1023 (2002).
1024 98. R. C. Araneda, Z. Peterlin, X. Zhang, A. Chesler, S. Firestein, A pharmacological
1025 profile of the aldehyde receptor repertoire in rat olfactory epithelium. *J Physiol*
1026 **555**, 743-756 (2004).
1027 99. P. A. Godfrey, B. Malnic, L. B. Buck, The mouse olfactory receptor gene family.
1028 *Proceedings of the National Academy of Sciences of the United States of*
1029 *America* **101**, 2156-2161 (2004).
1030 100. E. Shirokova *et al.*, Identification of specific ligands for orphan olfactory
1031 receptors. G protein-dependent agonism and antagonism of odorants. *J Biol*
1032 *Chem* **280**, 11807-11815 (2005).
1033 101. K. Kajiya *et al.*, Molecular bases of odor discrimination: Reconstitution of
1034 olfactory receptors that recognize overlapping sets of odorants. *The Journal of*
1035 *neuroscience : the official journal of the Society for Neuroscience* **21**, 6018-6025
1036 (2001).
1037 102. Y. Oka, M. Omura, H. Kataoka, K. Touhara, Olfactory receptor antagonism
1038 between odorants. *Embo J* **23**, 120-126 (2004).

1039 103. I. Gaillard *et al.*, A single olfactory receptor specifically binds a set of odorant
1040 molecules. *Eur J Neurosci* **15**, 409-418 (2002).
1041 104. M. Q. Nguyen, Z. Zhou, C. A. Marks, N. J. Ryba, L. Belluscio, Prominent roles for
1042 odorant receptor coding sequences in allelic exclusion. *Cell* **131**, 1009-1017
1043 (2007).
1044 105. W. B. Floriano, N. Vaidehi, W. A. Goddard, 3rd, M. S. Singer, G. M. Shepherd,
1045 Molecular mechanisms underlying differential odor responses of a mouse
1046 olfactory receptor. *Proceedings of the National Academy of Sciences of the*
1047 *United States of America* **97**, 10712-10716 (2000).
1048 106. Y. Oka *et al.*, Odorant receptor map in the mouse olfactory bulb: in vivo
1049 sensitivity and specificity of receptor-defined glomeruli. *Neuron* **52**, 857-869
1050 (2006).
1051 107. S. E. Repicky, C. W. Luetje, Molecular receptive range variation among mouse
1052 odorant receptors for aliphatic carboxylic acids. *J Neurochem* **109**, 193-202
1053 (2009).
1054 108. X. Grosmaitre *et al.*, SR1, a mouse odorant receptor with an unusually broad
1055 response profile. *The Journal of neuroscience : the official journal of the Society*
1056 *for Neuroscience* **29**, 14545-14552 (2009).
1057 109. X. Grosmaitre, A. Vassalli, P. Mombaerts, G. M. Shepherd, M. Ma, Odorant
1058 responses of olfactory sensory neurons expressing the odorant receptor MOR23:
1059 a patch clamp analysis in gene-targeted mice. *Proceedings of the National*
1060 *Academy of Sciences of the United States of America* **103**, 1970-1975 (2006).
1061 110. Y. Oka, Y. Takai, K. Touhara, Nasal airflow rate affects the sensitivity and pattern
1062 of glomerular odorant responses in the mouse olfactory bulb. *The Journal of*
1063 *neuroscience : the official journal of the Society for Neuroscience* **29**, 12070-
1064 12078 (2009).
1065 111. H. Saito, M. Kubota, R. W. Roberts, Q. Chi, H. Matsunami, RTP family members
1066 induce functional expression of mammalian odorant receptors. *Cell* **119**, 679-691
1067 (2004).
1068 112. H. Zhuang, H. Matsunami, Synergism of accessory factors in functional
1069 expression of mammalian odorant receptors. *J Biol Chem* **282**, 15284-15293
1070 (2007).
1071 113. K. Yoshikawa, K. Touhara, Myr-Ric-8A enhances G(alpha15)-mediated Ca2+
1072 response of vertebrate olfactory receptors. *Chemical senses* **34**, 15-23 (2009).
1073 114. T. Abaffy, H. Matsunami, C. W. Luetje, Functional analysis of a mammalian
1074 odorant receptor subfamily. *J Neurochem* **97**, 1506-1518 (2006).
1075 115. A. Dunkel *et al.*, Nature's chemical signatures in human olfaction: a foodborne
1076 perspective for future biotechnology. *Angew Chem Int Ed Engl* **53**, 7124-7143
1077 (2014).
1078 116. K. Yoshikawa, H. Nakagawa, N. Mori, H. Watanabe, K. Touhara, An unsaturated
1079 aliphatic alcohol as a natural ligand for a mouse odorant receptor. *Nat Chem Biol*
1080 **9**, 160-162 (2013).
1081 117. Y. Jiang *et al.*, Molecular profiling of activated olfactory neurons identifies odorant
1082 receptors for odors in vivo. *Nat Neurosci* **18**, 1446-1454 (2015).

1083 118. H. Saito *et al.*, Immobility responses are induced by photoactivation of single
1084 glomerular species responsive to fox odour TMT. *Nature communications* **8**,
1085 16011 (2017).
1086 119. B. von der Weid *et al.*, Large-scale transcriptional profiling of chemosensory
1087 neurons identifies receptor-ligand pairs in vivo. *Nat Neurosci* **18**, 1455-1463
1088 (2015).
1089 120. M. Shirasu *et al.*, Olfactory receptor and neural pathway responsible for highly
1090 selective sensing of musk odors. *Neuron* **81**, 165-178 (2014).
1091 121. Y. Yu *et al.*, Responsiveness of G protein-coupled odorant receptors is partially
1092 attributed to the activation mechanism. *Proceedings of the National Academy of*
1093 *Sciences of the United States of America* **112**, 14966-14971 (2015).
1094 122. E. M. Jones *et al.*, A Scalable, Multiplexed Assay for Decoding GPCR-Ligand
1095 Interactions with RNA Sequencing. *Cell Syst* **8**, 254-260 e256 (2019).
1096 123. P. Pfister *et al.*, Odorant Receptor Inhibition Is Fundamental to Odor Encoding.
1097 *Curr Biol* **30**, 2574-2587 e2576 (2020).
1098