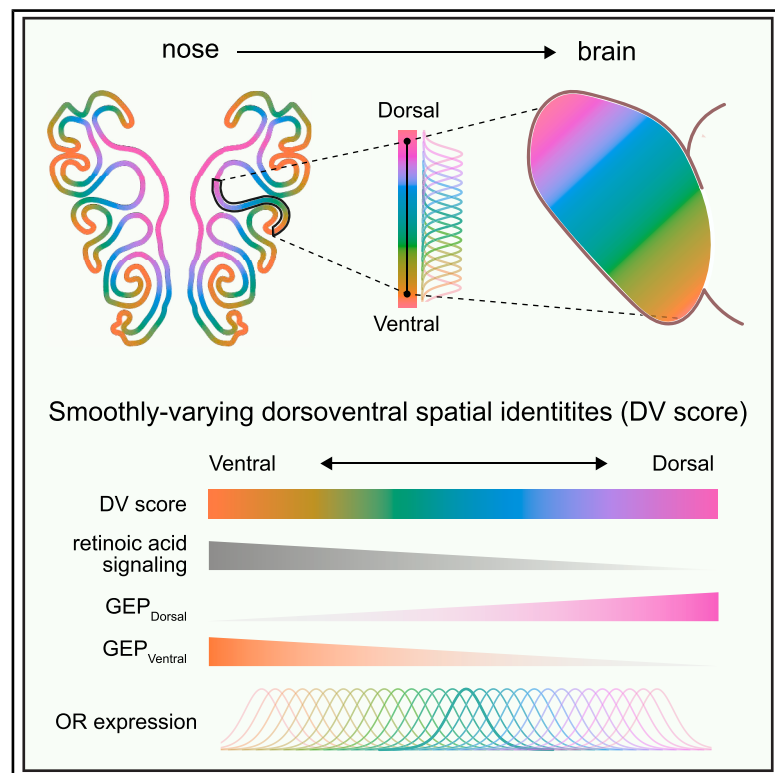


A spatial code governs olfactory receptor choice and aligns sensory maps in the nose and brain

Graphical abstract



Authors

David H. Brann, Tatsuya Tsukahara, Cyrus Tau, ..., Alexander Fleischmann, Thomas Bozza, Sandeep Robert Datta

Correspondence

srdatta@hms.harvard.edu

In brief

Dorsoventral epithelial position induces graded expression of a transcriptional program that maps each of the 1,100 olfactory sensory neuron subtypes to stereotyped spatial distributions in the epithelium, thereby building a precise receptor map in the nose that aligns with the glomerular map in the olfactory bulb.

Highlights

- ~1,100 olfactory receptors adopt stereotyped spatial distributions in the epithelium
- Epithelial space coherently regulates the graded expression of ~250 genes
- Precursor spatial identities bias olfactory receptor choice
- Receptor positions in the nose are aligned with their axonal targets in the brain

Article

A spatial code governs olfactory receptor choice and aligns sensory maps in the nose and brain

David H. Brann,¹ Tatsuya Tsukahara,^{2,3} Cyrus Tau,¹ Dennis Kalloor,¹ Rylin Lubash,¹ Lakshanyaa Thamarai Kannan,¹ Nell Klimpert,^{4,5} Mihaly Kollo,⁶ Martín Escamilla-Del-Arenal,^{7,8} Bogdan Bintu,⁹ Andreas Schaefer,⁶ Alexander Fleischmann,^{4,5} Thomas Bozza,¹⁰ and Sandeep Robert Datta^{1,11,*}

¹Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA

²Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON M5G 1X5, Canada

³Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

⁴Department of Neuroscience, Division of Biology and Medicine, Brown University, Providence, RI 02912, USA

⁵Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI 02912, USA

⁶Sensory Circuits and Neurotechnology Laboratory, The Francis Crick Institute, London NW1 1AT, UK

⁷Department of Biochemistry and Molecular Biophysics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

⁸Mortimer B. Zuckerman Mind, Brain and Behavior Institute, Columbia University, New York, NY 10027, USA

⁹Department of Cellular and Molecular Medicine, University of California, San Diego, San Diego, CA 92093, USA

¹⁰Department of Neurobiology, Northwestern University, Evanston, IL 60208, USA

¹¹Lead contact

*Correspondence: srdatta@hms.harvard.edu

<https://doi.org/10.1016/j.cell.2026.03.051>

SUMMARY

Although topographical maps organize many peripheral sensory systems, mouse olfactory sensory neurons (OSNs) are thought to randomly choose which one of ~1,100 possible olfactory receptors (ORs) to express, with spatial organization in the olfactory epithelium limited to a handful of broad anatomical “zones” that modestly restrict OR choice. Here, we reveal that each OR is instead expressed at a unique mean dorsoventral position, thereby instantiating a stereotyped receptor map in the olfactory epithelium. OSN dorsoventral identities are encoded by a coherent gene expression program, which includes key transcription factors and axon guidance molecules; use of this program reflects a dorsoventral gradient in retinoic acid signaling, translates each physical location into a spatially appropriate distribution of potential OR choices, and aligns receptor maps in the nose and brain. Spatial order in the olfactory system, therefore, arises from a continuously varying transcriptional code that precisely organizes the many discrete channels responsible for smell.

INTRODUCTION

The nervous system uses space to map sensation.^{1,2} The cochlea, for example, encodes different frequencies at distinct positions along its unrolled length; this tonotopic map is serially propagated through patterned projections to the brainstem, thalamus, and cortex.³ Like these auditory maps, a similar cascade of maps—originating in the sensory periphery and hierarchically running through higher brain centers—supports touch and vision.^{4,5} In the olfactory system, odors are detected by mature olfactory sensory neurons (OSNs) in the nasal main olfactory epithelium (MOE), each of which expresses only one of the 1,172 functional olfactory receptors (ORs) encoded in the mouse genome.^{6,7} However, the extent to which OR expression is spatially organized in the MOE into a stereotyped peripheral sensory map remains unclear.^{8–10} Addressing this foundational question is critical for our understanding of the sense of smell, as the answer has significant implications for how OSNs choose

their receptors, encode information about odors, and find their targets in the higher brain.

Two obstacles have vexed our understanding of whether OSNs are spatially patterned in the epithelium. First, characterizing the relationship between physical location and gene expression is difficult in the MOE because the underlying cartilaginous turbinates transform an otherwise smooth and continuous epithelial sensory sheet into a complex labyrinth,¹¹ challenging analysis of spatial order. Second, most of what we know about spatial patterning of OR expression arises from *in situ* hybridization and related techniques, which typically can query only a few genes at once and are difficult to align across samples^{12–15}; although microdissection has been recently used to assign ORs to particular epithelial locations, the resulting positional estimates are indirect and the degree of inter-individual stereotypy remains unclear.^{16,17}

Despite these constraints, current models propose that individual receptors are assigned to one of between four and nine

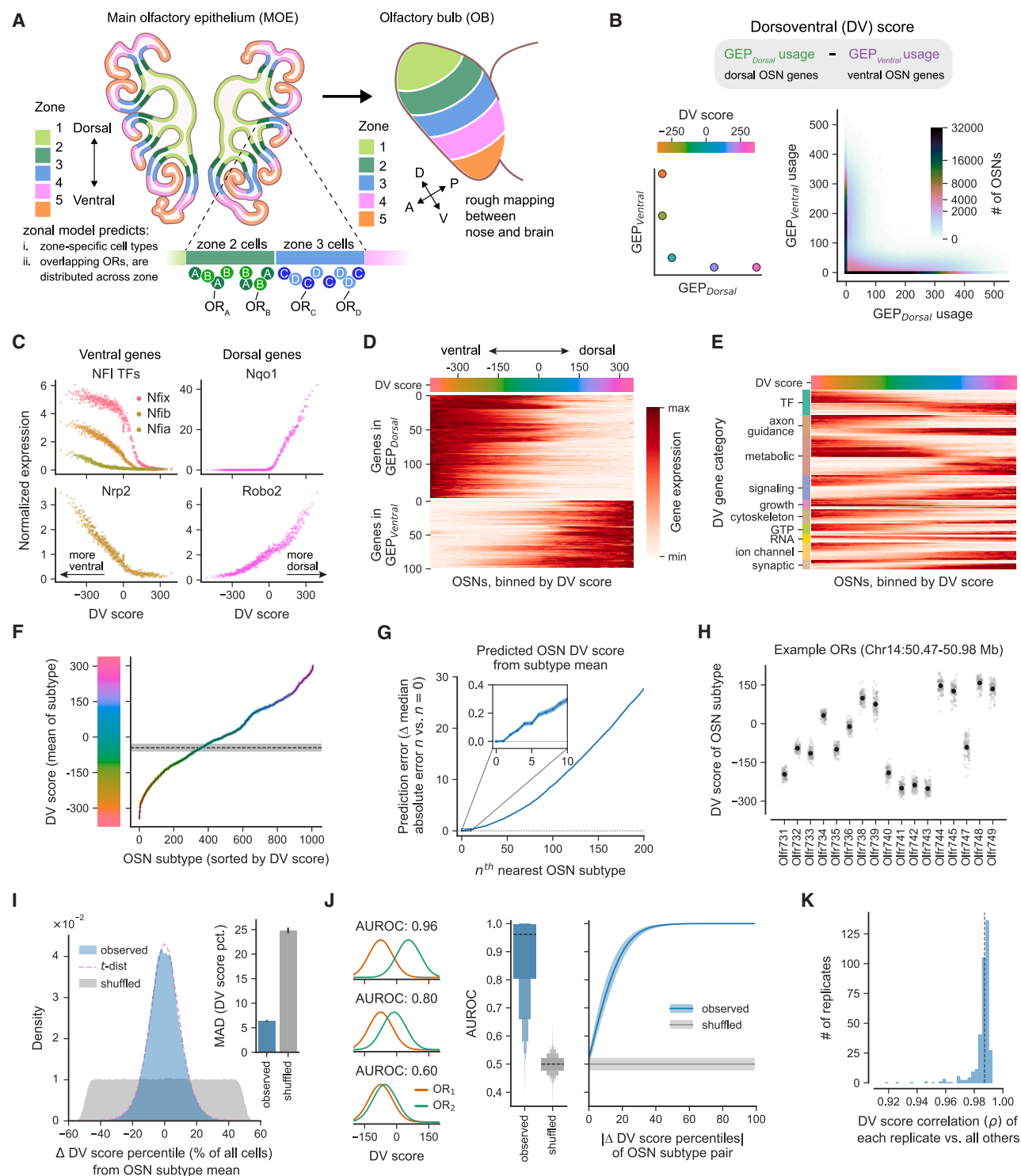


Figure 1. Each OSN subtype is associated with a unique DV score

(A) Schematic depicting the zonal model for OR choice.

(B) (Left) Measurement of the DV score for each OSN via scRNA-seq by subtracting GEP_{Ventral} from GEP_{Dorsal}. (Right) DV scores for all OSNs, with power-transformed colormap.

(C) Normalized expression of the indicated genes, averaged in equal-frequency bins as a function of cellular DV scores.

(D) Expression of 248 DV-related genes across OSNs, binned by DV scores and separated by associated GEP.

(E) Similar to (D) but separated by gene category.

(legend continued on next page)

dorsoventral (DV) “zones” (with some models positing additional subzones) but are randomly chosen by the OSNs within each zone, leading to spatially distributed expression of each OR across a broad swath of the epithelium (Figure 1A).^{12–16,18} Because the MOE continuously regenerates during adulthood from a population of basally localized stem cells, this zonal model implies that OSN precursors acquire zone-specific cell identities that limit OR choice to an appropriate subset.¹⁹ Consistent with this possibility, OSNs are known to be segregated into a small number of fate-restricted lineages that each express a different receptor gene class.^{20,21} Late-stage OSN precursors also appear fated to choose from restricted subsets of ORs, as the OR “switches” occasionally observed during OSN differentiation are largely among ORs from the same epithelial zone.^{22–24} However, only a handful of marker genes uniquely identify the dorsal-most zone, and no genes have been identified that demarcate the different ventral zones.^{25,26}

It also remains unclear how OSN precursors might restrict OR choice to the spatially appropriate set of ~150–250 ORs per zone.²⁷ Such restrictions must ultimately influence the mechanisms through which precursors choose a single OR, which are not yet fully understood but include: low-level co-expression of ORs in OSN precursors, which may license ORs for potential choice; heterochromatin-mediated silencing of ORs that are not ultimately chosen; interchromosomal interactions between OR genes, which facilitate both OR repression and choice; and, finally, high-level expression of a single OR allele.^{7,28–33} These choice mechanisms may be sensitive to zonal position, as recent results suggest that zonal positioning may influence the set of ORs available for co-expression in precursor cells.²⁸

Information from OSNs is projected to insular structures called glomeruli in the olfactory bulb (OB), the first waystation for olfactory processing in the brain. All OSNs of a given subtype (i.e., those expressing the same OR) innervate one of a small number of glomeruli whose spatial location is largely invariant across animals, thereby creating a stereotyped spatial map of odor identity.^{34–37} This architecture demands a means to align the proposed zones in the nose and the precise, spatially organized map in the OB.^{38,39} This mechanism likely involves the axon guidance genes *Robo2* and *Nrp2*, which are expressed in complementary gradients in the epithelium and causally regulate the targeting of OSN axons along the DV axis in the bulb.^{40–43} However, it is not known whether these gradients systematically vary across OSN subtypes, or whether this variation is sufficient to ensure proper glomerular targeting.^{44,45} Furthermore, the

disconnect between the broad spatial distribution of ORs in the MOE and the precision of their axonal targeting in the bulb has led to the suggestion that the OR itself determines key aspects of OSN maturation by regulating the expression of axon guidance genes and by facilitating homophilic and heterophobic interactions among axons^{38,46–49}; in this model, OSN cell identities are largely derived from their late-stage expression of ORs, rather than from spatial position within the epithelium.

Here, we show that the nose harbors a receptor map, one built from ~1,100 distinguishable peaks of OR expression whose mean positions are reliably stereotyped across mice. This patterning is the consequence of a coherent transcriptional program comprised of ~250 genes, including key transcription factors, retinoic acid (RA) signaling components, and axon guidance molecules. The relative use of this program distinguishes all mature OSN subtypes, varies systematically as a function of DV spatial position in the epithelium, and reflects RA gradients in the mesenchyme underneath the OE proper. OSN stem cells and precursors take advantage of this program—which both influences many of the known mechanistic processes that underlie OR choice and predicts the spatial location of target glomeruli—to encode information about spatial position before receptors are chosen for expression and before the process of axon guidance has begun. Thus, the transcriptional identity of OSN precursors is defined by spatially organized gradients of gene expression, which are then transformed into discrete and predictable OR choices and glomerular targets; this mechanism coordinately organizes the first two stages of the peripheral olfactory system by constructing a detailed receptor map in the nose aligned with the precise receptor map apparent in the OB.

RESULTS

We hypothesized that OSNs contain transcriptional signatures that reflect their spatial positions within the epithelium. To explore this idea, we used single-cell RNA sequencing (scRNA-seq) to characterize ~5 million MOE cells across hundreds of mice, which yielded a dataset of ~2.3 million OSNs that collectively expressed ~1,100 ORs (Figure S1A). Although we expected to identify specific marker genes for each zone,⁵⁰ our analysis failed to reveal such genes (Figure S1B).

However, matrix factorization identified several OSN-specific gene expression programs (GEPs), each composed of ~100–200 individual genes whose expression levels co-vary across different OSNs.^{51,52} Two such GEPs (provisionally named

(F) DV scores, sorted by the mean for each OSN subtype (defined as the set of OSNs singularly expressing a given OR).

(G) Error in predicting the DV score for each OSN based upon the mean DV score of OSNs of the same subtype or those with neighboring DV scores.

(H) DV scores of OSNs expressing ORs from a genomic cluster on chromosome 14. Black dots indicate the mean, and gray dots depict the means when cells were split by replicate into 100 bins.

(I) The percent-normalized DV score of each OSN relative to its respective subtype mean, with *t*-distribution fit (dashed line). Gray, shuffled data. (Inset) Median absolute deviation (MAD) of each distribution, with mean and 95% confidence intervals (CIs) across restarts.

(J) (Left to right) Area under the receiver operating characteristic curve (AUROC) values for example pairs of DV score distributions; the AUROCs for all pairs of OSN subtypes (black dashed line indicates median, boxes represent 25th/75th percentile, 12.5th/87.5th percentile, and so forth across $_{1007}C_2 = 506,521$ pairs); and the median and shaded interquartile range of AUROC values for pairs versus their difference in DV scores (as a percentile of all subtypes). Gray, shuffled OR labels.

(K) Histogram of the correlation of the mean DV scores for each OSN subtype from each replicate versus those from all other replicates. Dashed line depicts the median across replicates ($n = 363$).

See also Figure S1, Tables S1 and S2, and Data S1 and S2.

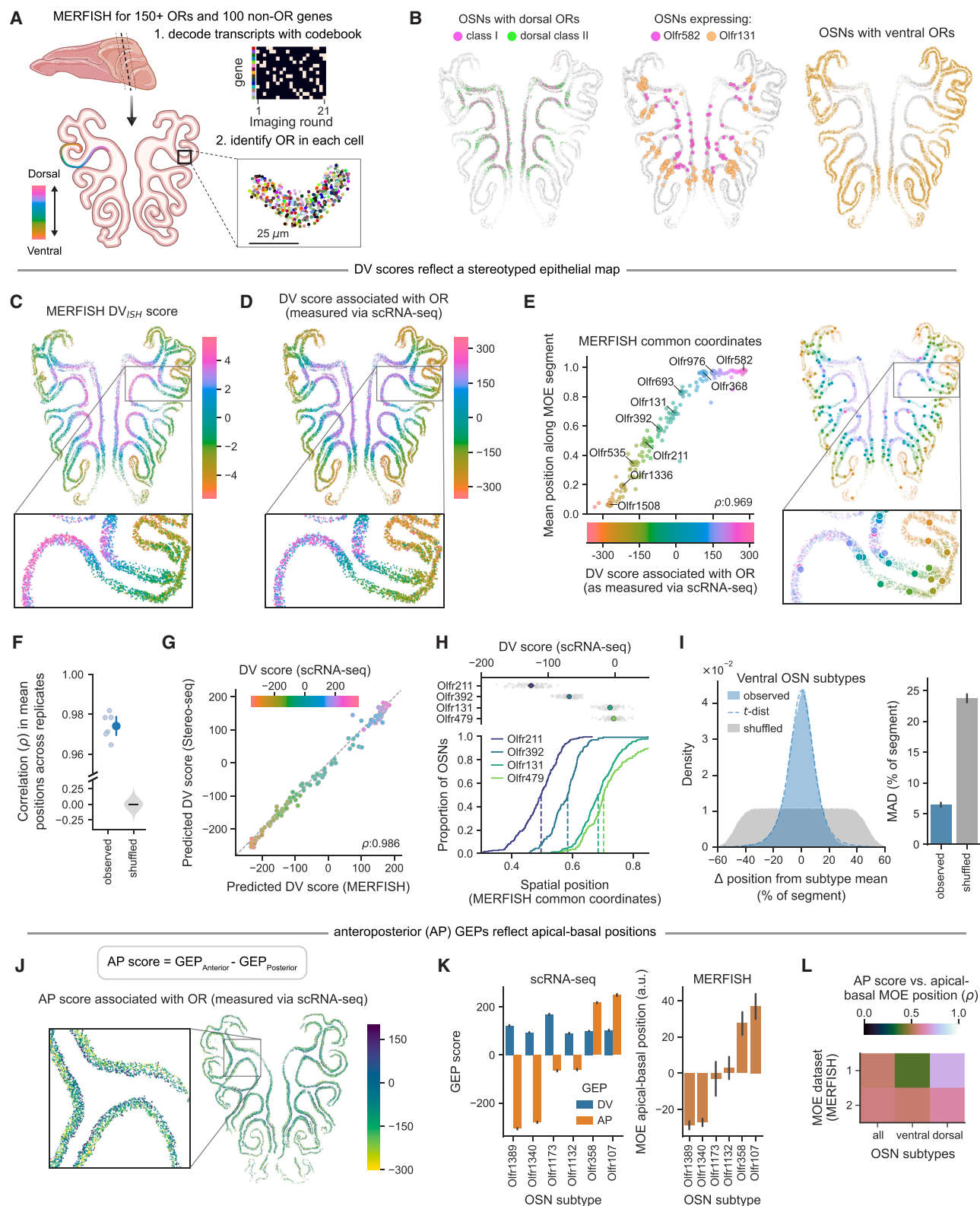


Figure 2. Each OSN subtype occupies a unique region of the epithelium

(A) Schematic depicting multiplexed detection of ORs and non-OR genes (including DV genes) in the MOE via MERFISH, with the DV axis colored along an example epithelial segment.

(legend continued on next page)

GEP_{Dorsal} and GEP_{Ventral}) included genes like the known dorsal marker *Nqo1* and the ventral markers *Nfia/b/x*, which encode the nuclear factor I (NFI) family of transcription factors; these two GEPs also included many additional transcription factors, axon guidance molecules, and cytoskeletal regulatory genes (Figure S1A; Table S1). For simplicity, we distilled the expression of the genes comprising each GEP in each OSN into a single number, the GEP “usage.” Because individual OSNs used either GEP_{Dorsal} or GEP_{Ventral}, but not both, we could further summarize the use of these two GEPs by computing a DV score (i.e., the usage of GEP_{Dorsal} subtracted from that of GEP_{Ventral}) (Figure 1B). Unlike what would be expected if OSNs were organized into homogeneous and discrete “zones,” *Nqo1* expression tonically and smoothly increased as a function of each OSN’s DV score, as did the ~100 other genes participating in GEP_{Dorsal} (including e.g., *Robo2*); conversely, the expression of *Nfia/b/x* and the ~150 other genes participating in GEP_{Ventral} (including e.g., *Nrp2*) tonically decreased as a function of each OSN’s DV score (Figures 1C–1E).^{40–43} Thus, the expression of the ~250 genes that make up GEP_{Dorsal} and GEP_{Ventral} systematically varies in a graded fashion across OSNs.

The use of GEP_{Dorsal} and GEP_{Ventral} by a given OSN reflected its expressed odor receptor, as each OSN subtype (defined as the collection of OSNs expressing the same OR) was associated with a unique average DV score; across subtypes, average DV scores varied smoothly without obvious clustering (Figures 1F–1H and S1C–S1F; Table S2). The relationship between OSN subtypes and DV scores reflected structured covariation across many genes, as DV scores were well-approximated by random subsets of 10–20 DV genes and changed little after excluding dorsal and ventral marker genes (Figures S1G and S1H). The individual OSNs belonging to each subtype exhibited closely related (but not identical) DV scores, such that OSNs belonging to different subtypes could be effectively distinguished based upon their DV scores alone (median auROC = 0.96) (Figures 1I, 1J, and S1I). The ability to distinguish OSN subtypes depended upon the difference in the DV scores of each pair, both within and across proposed zones (Figures 1J, S1J, and S1K). Similarly, expression levels of nearly every DV gene were better predicted by DV scores than by each subtype’s putative zonal identity (Figure S1L).

The relationship between each OSN subtype and its associated mean DV score was remarkably stable, as the rank ordering of OSN subtypes based on their DV scores was nearly invariant ($\rho = 0.987$) across hundreds of independent samples (Figures 1K and S1M–S1P). Thus, despite variation in DV scores among individual OSNs expressing the same OR, the mean DV score associated with each OSN subtype—which summarizes the coordinated expression of several hundred genes—is unique and stereotyped across mice and is sufficient to differentiate each OSN subtype from all others.

DV scores predict the unique mean spatial location of each OR

We hypothesized that the precise DV location adopted by each OSN in the epithelium is encoded by its DV score. If so, the mean physical location of each OSN subtype—and therefore OR—along the DV axis should be unique, predictable, and stereotyped across mice. To test this alternative hypothesis to the discrete zonal model, we used multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) to measure the spatial positions of hundreds of OR and DV genes throughout the epithelium (Figures 2A, 2B, S2A, and S2B).⁵³

As expected, individual ORs were expressed in a spatially restricted manner, with the most-dorsal receptors being more broadly expressed than more ventral receptors (Figure 2B). However, simply coloring each OSN based upon its associated patterns of DV gene expression (the DV_{ISH} score, determined directly via MERFISH) revealed a previously unappreciated fine organization to the epithelium: along each segment, DV_{ISH} scores smoothly varied from dorsal to ventral (Figures 2C and S2C). Furthermore, identifying the OR expressed within each OSN and coloring it based upon its DV score (as measured via scRNA-seq) also yielded a similar smooth and continuous spatial gradient (Figures 2D and S2C). Physical space also organized OSN DV scores when the epithelium was “unrolled,” by aligning all segments to a common spatial axis (Figures 2E and S2D). Given that each OSN subtype is associated with a unique mean DV score, these observations reveal that each OR occupies a unique mean position along the DV axis of the epithelium.

(B) OSNs expressing (left to right) all dorsal ORs; the dorsal OR Olfr582 and the ventral OR Olfr131 (with dots ~100× normal size for visualization purposes); and all ventral class II ORs.

(C) OSNs colored by their DV_{ISH} score, the first principal component of the MERFISH-measured expression of DV genes.

(D) OSNs colored by the DV score associated with their expressed OR, as measured via scRNA-seq.

(E) (Left) Mean common coordinate spatial position for each OSN subtype versus its associated DV score. (Right) Mean positions in each epithelial segment for the subtypes highlighted on the left.

(F) Correlation of the mean common coordinates of OSN subtypes across pairs of individual sections (blue, observed data; gray, shuffled data).

(G) DV scores for each OSN subtype from MERFISH or Stereo-seq data, as predicted from neighboring OSNs.

(H) Cumulative distribution of common coordinate positions for OSNs expressing each example OR (bottom), along with their associated DV scores (via scRNA-seq, top).

(I) (Left) Spatial positions of OSNs belonging to ventral subtypes, plotted as the difference (as a percentile of all positions) from their subtype mean, with *t*-distribution fit (dashed line). Gray, shuffled data. (Right) The MAD of the respective distributions, with mean and 95% CIs across restarts.

(J) Similar to (D) but colored by the anteroposterior (AP) score, using MERFISH data from Bintu et al.⁵⁴

(K) (Left) DV and AP GEP usages for example dorsal OSN subtypes. (Right) Apical-basal positions of the OSNs expressing each OR, with mean and 95% CIs.

(L) Correlation of AP score versus apical-basal positions in two MERFISH datasets (1, this manuscript; 2, Bintu et al.⁵⁴), for all detected OSN subtypes or for those identified as dorsal or ventral based on their DV scores.

See also Figures S2 and S3.

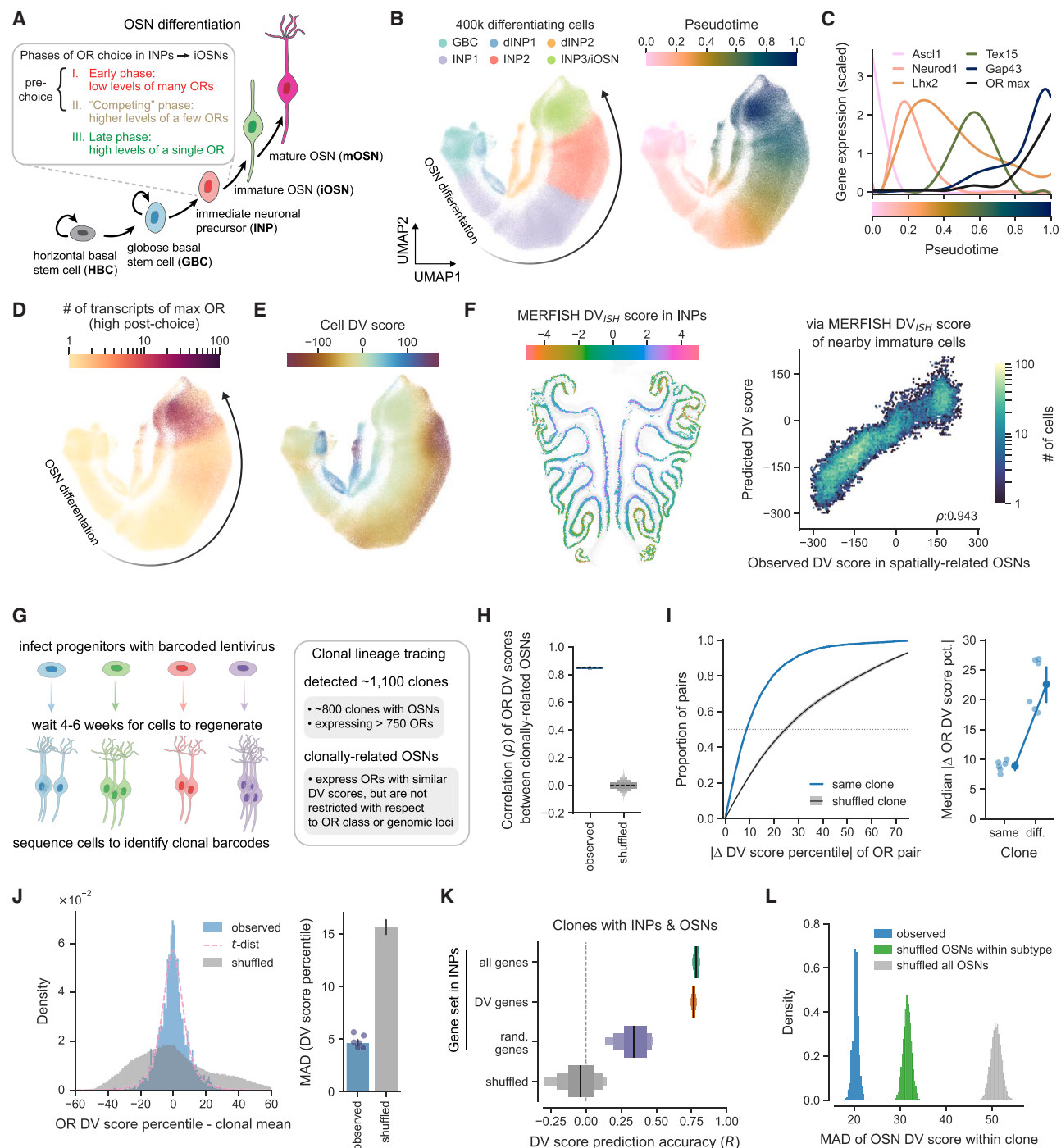


Figure 3. Epithelial position restricts OSN precursor fates and OR choice

(A) Schematic depicting OSN differentiation and OR co-expression stages that precede singular expression.

(B) UMAP plots depicting ~400,000 precursors and differentiating OSNs (excluding mature cells), colored either by cell cluster (left) or inferred pseudotime (right). dINP, dorsal INP.

(C) Expression across pseudotime of genes related to OSN differentiation, scaled by their standard deviation.

(D) Expression of the highest-expressed OR in each cell.

(E) DV scores for each differentiating cell (see STAR Methods).

(F) (Left) INPs, colored by their DV_{ISH} score, as measured via MERFISH. (Right) Predictions of the DV score associated with each OSN using the DV_{ISH} score of nearby immature cells.

(G) Schematic of *in vivo* clonal lineage tracing.

(legend continued on next page)

The smoothly varying relationships between epithelial position, DV scores, and OR expression were conserved across animals and enabled accurate predictions of the spatial location of each OSN subtype (and hence OR) in our MERFISH data, in a separate spatial transcriptomics dataset generated via Stereo-seq, and in five independently obtained datasets in which the position of OSN subtypes could be precisely or approximately assigned (Figures 2F, 2G, and S2E–S2H).^{16,17,28,54} In contrast, we did not observe any discrete spatial clustering of ORs into clear zones, and DV scores outperformed zonal models at predicting OSN spatial positions (Figures S3A and S3B).

Whereas ventral OSN subtypes had distinct and predictable mean locations, their constituent OSNs occupied a distribution of locations (with a sharp peak and long tails encompassing ~5%–15% of the epithelium), with the degree of overlap in DV score distributions between OSN subtypes predicting their extent of spatial overlap (Figures 2H, 2I, and S3C). The relationship between spatial positions and DV scores was less apparent in the dorsal-most region (Figures 2D, 2E, and S3D), likely reflecting the broader spatial distributions observed for dorsal receptors. Importantly, DV_{ISH} scores accurately predicted spatial positions within OSNs of a given subtype (Figures S3E and S3F), suggesting that DV score variation largely reflects differences in physical location rather than noise.

Together, these data reveal that the precise location of each OSN is well captured by its DV score and that each OSN subtype occupies a unique and restricted spatial distribution in the olfactory epithelium, thereby creating a receptor map with ~1,100 overlapping but distinguishable peaks. This pattern is the consequence of each OR having a preferred epithelial spatial position, and of those OSNs residing at each spatial position having a preferred OR.

AP scores reflect apical-basal position in the epithelium

Our factorization approach identified an additional pair of GEPs (together comprising ~300 total genes) (see Table S3), which we provisionally named GEP_{Anterior} and GEP_{Posterior}, as they included genes (such as *Nrp1* and *Plxna1*) known to mediate axon targeting along the bulb anteroposterior (AP) axis.^{46,47,51,55} Similar to the DV GEPs, OSNs used either GEP_{Anterior} or GEP_{Posterior} (Figure S3G), enabling the generation of an “AP score” that varied across subtypes. We wondered whether this variation was also driven by epithelial location, and we observed via MERFISH that differences in AP scores across OSN subtypes corresponded to their apical-basal position within the OSN layer (Figures 2J, 2K, and S3H). This relation-

ship was strongest for dorsal OSN subtypes (Figure 2L), which are notably the subset of OSNs in which DV scores are least informative about spatial position.

DV identity is independent of the expressed OR and downstream activity

Prior work suggests that OR-specific spontaneous activity drives the expression of some AP GEP genes.⁵¹ We therefore wondered whether the DV score of each OSN subtype was similarly determined by the expressed OR. To test this idea, we performed scRNA-seq on OMP-IRES-tTA; tetO-P2 (OMP-P2) mice, in which the P2 receptor is ectopically expressed across the DV axis; this process frequently silences the previously chosen OR in each OSN (Figure S4A).⁵⁶ P2 OSNs from OMP-P2 mice adopted a much wider distribution of DV scores than did P2-expressing OSNs from wild-type mice (Figure S4B); this result demonstrates that the identity (and associated activity) of the expressed OR does not determine OSN DV scores, suggesting instead that DV scores are determined by OSN spatial position. Consistent with this alternative hypothesis, the ~15% of OSNs in which P2 was co-expressed with another OR (i.e., those switching from their original OR to P2) exhibited DV scores that matched the originally chosen OR rather than P2 (Figure S4C and S4D). Odor-evoked activity downstream of the OR also appeared unlikely to determine DV scores, which were essentially unchanged after three distinct activity manipulations: naris occlusion, exposure to novel olfactory environments, and knocking out CNGA2, the key ion channel that transforms OR activation into calcium entry and spiking activity (Figures S4E and S4F).^{51,57} Thus, the OR expressed by each OSN does not influence its DV score, and yet under normal circumstances, the identity of the expressed OR correlates with each OSN's spatial position and DV score.

Transcriptional signatures of DV identity are present in OSN precursors

These observations suggest a model in which spatially organized patterns of DV gene expression are initiated prior to OR expression, in either the stem cells or immediate neuronal precursors (INPs) that reside in organized layers beneath their mature progeny (Figure 3A).^{58,59} To test this hypothesis, we analyzed DV gene expression in 400,000 differentiating cells extracted from our scRNA-seq dataset (Figures 3B–3D and S5A). Many DV genes were expressed during OSN differentiation, which allowed us to compute DV scores for all precursor cells (Table S1). This revealed a gradient of DV scores across OSN precursors that was present throughout all stages of differentiation and which spanned all three OSN lineages (corresponding to expression

(H) Correlation in associated DV scores of clonally related ORs. Gray, shuffled clonal labels.

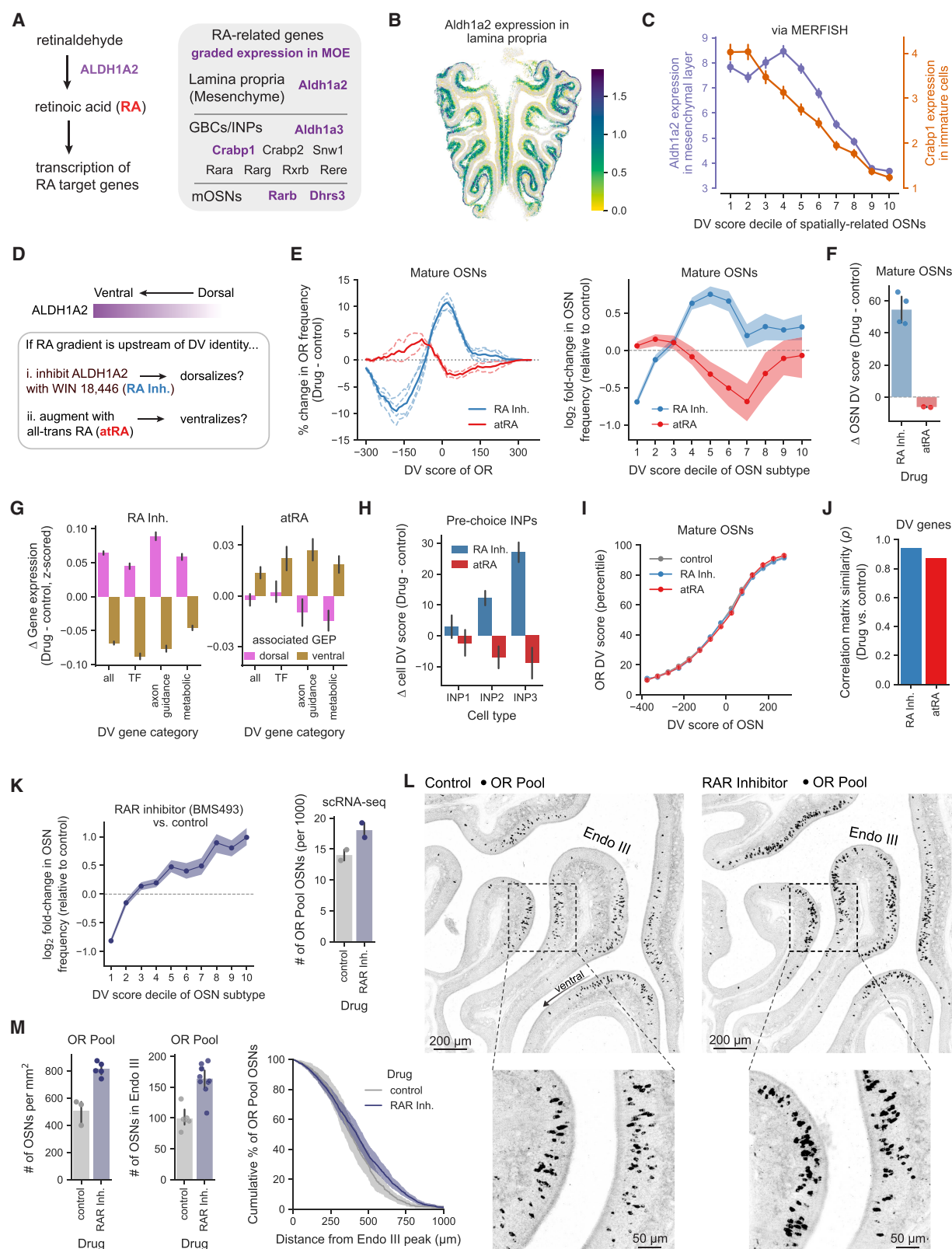
(I) (Left) Cumulative difference in DV scores of clonally related ORs. Gray, shuffled clonal labels, with mean and 95% CIs across restarts. (Right) Median DV score distances for OSN pairs from the same or different clones, with mean and 95% CIs across mice ($n = 6$).

(J) DV scores associated with clonally related ORs, plotted as the difference from their clonal mean, with t -distribution fit (dashed line). Gray, shuffled clones. (Right) MAD of the respective distributions, with mean and 95% CIs across restarts.

(K) Performance (across 100 restarts) of models trained on gene sets in INPs to predict the OR-associated DV scores of clonally related OSNs. Gray, DV genes with shuffled OSN DV scores.

(L) MAD of DV scores for clonally related OSNs (blue), or for OSNs with clonal labels shuffled across OSNs within a subtype (green) or shuffled across all OSNs (gray).

See also Figures S4, S5, and S6.



(legend on next page)

of class I ORs, dorsal class II ORs, and ventral class II ORs), apparent in our dataset (Figures 3E and S5B–S5E).

Principal components analysis revealed that DV score variation was a major driver of transcriptional variation among precursors within each developmental stage, though expression of individual DV genes peaked at different stages (Figures S5F and S5G). Differences in precursor DV gene expression reflected differences in spatial positioning within the epithelium, while at each DV location, precursors and overlying OSNs exhibited similar DV scores (Figure 3F). Spatial position, therefore, determines the expression levels of GEP_{Dorsal} and $GEP_{Ventral}$ genes across all stages of OSN differentiation, thereby transcriptionally diversifying OSN precursors before they commit to expressing a single OR.

Clonally related cells express restricted sets of ORs with similar DV identities

The observation that precursors and nearby OSNs have similar DV scores suggests that stem cell DV position might restrict OR choice. To test this idea, we performed *in vivo* clonal lineage tracing by chemically ablating OSNs with methimazole and infecting basal stem cells intranasally with lentiviral DNA barcodes. After allowing the stem cells to regenerate the epithelium for several weeks,⁶⁰ we performed scRNA-seq to identify the set of OSN subtypes (and therefore ORs) generated by each marked stem cell (Figures 3G and S6A–S6D).

Most identified clones (~800 out of ~1,100) included OSNs, which collectively expressed more than 750 different ORs (Figure S6E). Individual OSNs derived from the same clone often expressed different ORs, demonstrating that there is not a deterministic 1:1 relationship between stem cell positional identity and OR choice. However, clonally related OSNs had much more similar DV scores than did OSNs from different clones, and as a result expressed the same OR at above-chance rates (Figures 3H–3J and S6F–S6I). Clonally related precursors also had similar DV scores; importantly, precursor DV gene expression could be used to predict the ORs expressed by the OSNs found in the same clone (as determined by their associated DV scores) (Figures 3K and S6J). Interestingly, the DV scores of

clonally related OSNs were even more restricted than expected based on their chosen ORs, suggesting a tighter mapping between epithelial space and DV identity than between epithelial space and OR choice (Figure 3L). These observations are consistent with the process of OR choice in OSN progenitors being biased toward a spatially optimal receptor—thereby generating a patterned map of OR identity in the nose—but also being sufficiently stochastic such that OSN progenitors often pick ORs associated with OSN subtypes nearby in space.

RA signaling varies spatially and influences OR choice

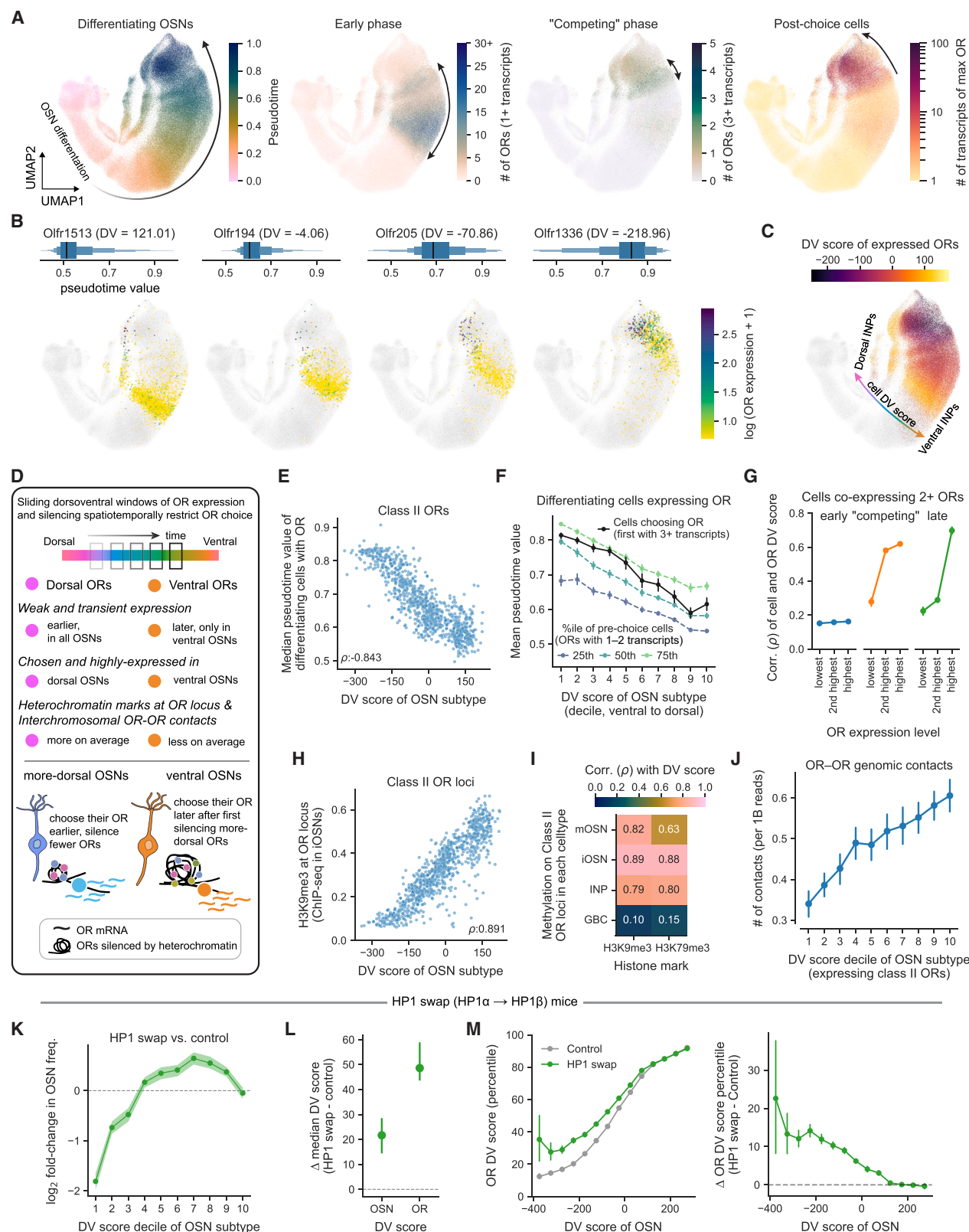
Our results demonstrate that OSN precursors somehow “know” where they are located physically along the DV axis, enabling location-aware OR choice. Consistent with this idea, OSNs regenerated following methimazole treatment exhibited the same relationships between DV scores and OR choice observed in untreated epithelia (Figure S6C).¹⁹ We searched for candidate signaling programs that might inform stem cells about their DV position, and observed several genes related to RA signaling whose expression varied spatially across different epithelial cell types (Figure 4A).^{61–63} Consistent with prior reports,^{40,63} retinaldehyde dehydrogenase 2 (ALDH1A2, the enzyme that converts retinaldehyde into RA) is expressed in lamina propria mesenchymal cells—below the stem cell layer—in a gradient that negatively correlated with the DV scores of nearby OSN subtypes (Figures 4B, 4C, and S7A). OSN precursors expressed RA receptors, such as *Rara*, *Rarg*, *Rarb*, and *Rere*, and both precursors and OSNs expressed additional genes capable of producing or responding to RA in a graded manner, including *Aldh1a3*, *Rarb*, *Dhrs3*, and the cellular RA binding protein 1 (*Crabp1*) (Figures 4A, 4C, S6B, and S6C).

Mesenchymal RA might therefore signal DV position and influence OSN fates. To test this hypothesis, we systemically delivered either an inhibitor of ALDH1A2 (WIN 18,446) or an activator of RA signaling (all-trans retinoic acid [atRA]) to adult mice during methimazole-induced regeneration and evaluated their effects via scRNA-seq (Figures 4D and S7D). Compared with control mice, we observed bidirectional shifts in the number of regenerated “dorsal” or “ventral” OSNs: mice given the RA inhibitor had

Figure 4. Manipulating retinoic acid signaling biases OSN fates and OR choice

- (A) Schematic of enzymes involved in retinoic acid (RA) signaling, and RA-related gene expression in the MOE.
 (B) *Aldh1a2* expression (via MERFISH) in the mesenchymal lamina propria (LP).
 (C) Expression of *Aldh1a2* in the LP and the cellular retinoic acid binding protein *Crabp1* in immature cells, binned based on the DV scores associated with ORs expressed in spatially related OSNs.
 (D) Schematic depicting consequences of RA manipulation during OSN regeneration.
 (E) (Left) Post-regeneration changes in OR frequency (drug versus vehicle), for OSN subtypes with given DV scores, with mean and individual replicates (dashed line). (Right) $\log_2$ fold-changes in OSN frequency, with 95% CIs.
 (F) RA drug-induced change in median cellular DV score. Dots and error bars depict individual replicates and 95% CIs, respectively.
 (G) RA drug-induced changes in z-scored DV gene expression, split by associated GEP and gene category.
 (H) Changes in cellular DV scores for pre-choice INP cells.
 (I) Percentile-normalized OR-associated DV scores versus OSN DV scores.
 (J) DV gene similarity between control and drug-treated animals.
 (K) (Left) As in (E) but for mice given an RA receptor (RAR) inhibitor. (Right) Via scRNA-seq, OSN frequency for OR set queried using a pooled *in situ* hybridization probe (see STAR Methods). Dots, individual replicates.
 (L) *In situ* hybridizations with the pooled OR probe in control and RAR animals. Endo III, Endoturbinate III.
 (M) In control and RAR-treated animals, the number of OSNs labeled by the pooled OR probe and cumulative OSN density as a function of dorsoventral position along Endo III.

See also Figure S7.



(legend on next page)

relatively fewer OSNs with more ventral DV scores (and therefore fewer cells expressing ventral ORs), whereas mice given atRA exhibited the opposite phenotype (Figures 4E and S7E); these bidirectional, RA-dependent shifts in DV scores were observed in both precursors and OSNs (Figures 4F–4H). Importantly, these manipulations neither altered the relationship between DV scores and the chosen ORs *per se*, nor changed the covariation structure of the genes that make up the DV score (Figures 4I and 4J). Off-target drug effects are unlikely to explain our observations, as treating mice with a pan-retinoic acid receptor (RAR) inverse agonist (BMS493) also led to dorsal shifts in OR choices and DV scores (Figure 4K).

These results predict that inhibiting RA signaling should shift the relative positions of OSN subtypes along the DV axis of the epithelium, which we tested by performing *in situ* hybridizations for a pooled set of ORs expressed in the dorsal part of the ventral epithelium (see STAR Methods). Consistent with our scRNA-seq results, BMS493 treatment caused more cells to express relatively dorsal ORs; furthermore, the spatial distribution of these cells was shifted ventrally (Figures 4L and 4M), consistent with the ventral epithelium being dorsalized under conditions of RA inhibition.

Linking DV location to OR choice

Our data suggest that a sub-epithelial RA gradient informs overlying stem cells and precursors about their DV position, thereby translating spatial location into a particular pattern of DV gene expression and ultimately into the selection of spatially appropriate ORs. We therefore wondered how precursor DV gene expression might influence the multiple mechanisms that support singular OR choice, which in broad strokes involve the low-level co-expression of multiple receptors, the repression of unchosen receptors through targeted heterochromatinization and genomic compartmentalization of OR genes, and the selective high-level expression of a single OR gene.^{28–31,64–67}

Receptor co-expression is organized along a DV gradient

Pseudotemporal ordering of our scRNA-seq data confirmed that OSN precursors often co-express 10–20 ORs at low levels before they ultimately choose a single OR to express at high levels. We wondered whether dorsal precursors preferentially co-expressed dorsal ORs and conversely whether ventral precursors co-expressed ventral ORs. Surprisingly, all precursors, regardless of spatial position (as inferred by the DV score of each cell), initially express a limited subset of the dorsal-most receptors (Figures 5A–5C). The dorsal-most precursors subsequently express dorsal ORs at higher levels, such that dorsal ORs like *Olf1513* are exclusively expressed at high levels in dorsal cells. In contrast, more ventral precursors instead silence the weakly co-expressed dorsal ORs and start to co-express slightly more ventral ORs like *Olf194* (Figure 5B). As differentiation proceeds, these precursors then similarly resolve their co-expression into a singular choice, and the process of co-expression, selection (if spatially appropriate), and down-regulation (if not) progressively repeats itself in a dorsal-to-ventral pattern until the ventral-most ORs are finally co-expressed and subsequently chosen in the ventral-most cells.

As a result of this process, there is a sliding window of OR co-expression that systematically evolves as OSN precursors differentiate, with dorsal-most receptors being co-expressed and chosen first, and ventral-most receptors being co-expressed and chosen last (Figures 5C–5F). The delay between dorsal and ventral choice likely reflects the fact that ventral precursor cells must first co-express and silence more dorsal ORs before they can express a spatially appropriate ventral OR. Note that class I receptors located in the dorsal-most zone violate this pattern of co-expression and repression, consistent with class I OR-expressing OSNs having a distinct developmental trajectory (Figure S5D).

Do the spatial identities of each OSN precursor predict which of the co-expressed ORs is chosen for singular expression? We

Figure 5. Space, time, and heterochromatin translate DV scores into restricted OR choice

- (A) UMAP plots coloring each differentiating OSN by (left to right) pseudotime; the number of ORs expressed with 1+ or 3+ transcripts; and the expression of the highest-expressed OR.
- (B) Log-transformed expression (bottom) and pseudotime distribution (top) of cells expressing the indicated ORs (gray, no expression), and their associated DV scores in mOSNs.
- (C) The mean DV score associated with each cell's co-expressed ORs (as measured in the respective mOSN subtype), weighted by OR expression for cells co-expressing ORs (gray, no OR expression).
- (D) Summary of how OR choice is spatiotemporally restricted, and how DV scores relate to OR co-expression, heterochromatin, and interchromosomal contacts.
- (E) Median pseudotime of differentiating cells expressing class II ORs versus their associated DV score in mOSNs.
- (F) Mean pseudotime for differentiating cells expressing class II ORs, binned based on their associated DV scores in mOSNs. Dashed lines, the 25th, 50th, and 75th percentiles of pseudotime values for cells with 1–2 OR transcripts; solid lines, the 10th percentile of cells with 3+ transcripts.
- (G) For each stage of OR co-expression, the correlation of cellular DV scores and the mOSN DV scores associated with the OR expressed at the highest, second-highest, and lowest levels.
- (H) Heterochromatin levels at each class II OR locus (via H3K9me3 ChIP-seq^{28,65}), versus the DV score associated with each OSN subtype.
- (I) As in (H), for either H3K9me3 or H3K79me3 levels on OR loci in the indicated cell types.
- (J) Interchromosomal OR–OR contact frequency (via the chromatin conformation method Micro-C) for class II OR loci, binned based on their associated DV scores, with means and 95% CIs.
- (K) Log₂ fold-changes in OSN frequency (HP1 swap versus control mice, with 95% CIs) for OSN subtypes with given DV scores.
- (L) Changes (HP1 swap versus control mice, with 95% CIs) in the median DV scores at the cell level or the DV scores associated with each OR (as measured in control mOSNs).
- (M) (Left) Percentile-normalized OR-associated DV scores versus cellular DV scores, for control and HP1 swap mice. (Right) Residuals from predicting OR-associated DV scores using cellular DV scores from control mice.
- See also Figures S5 and S7.

find that at earlier stages of differentiation, there is only weak alignment between the weakly co-expressed ORs and cellular DV scores, but that once precursors start expressing ORs at higher levels, the single highest-expressed OR becomes well aligned with cellular DV scores (Figures 5G, S5H, and S5I). Thus, as OSN precursors transition from expressing multiple to single ORs, expression levels of “incorrect” ORs systematically decline, yielding an OSN that, on average, expresses the most appropriate OR for its given spatial location (although OSNs still frequently choose ORs associated with nearby spatial positions). Consistent with RA gradients informing OSN precursors about spatial location, RA manipulations systematically shifted the set of ORs that were co-expressed in pre-choice cells (Figure S7F). DV positioning, therefore, not only influences which of the ~1,100 different ORs a given OSN chooses to express but also shapes how this process unfolds during differentiation.

A heterochromatin gradient represses inappropriate OR expression

OR genes not chosen for singular expression are repressed by heterochromatin, which is deposited onto OR loci during differentiation. We wondered whether regulators of this process might participate in establishing the relationships between precursor DV scores, the timing of OR choice during differentiation, and the set of ORs available for singular choice. Indeed, recent work identified the NFI transcription factors as being causally responsible for tuning the levels of OR heterochromatinization in more ventral zones, as knocking out *Nfi* genes in OSN precursors reduces heterochromatin levels on ventral OR loci and causes OSNs in putative ventral zones to express more dorsal ORs.²⁸ The *Nfi* genes also plausibly couple spatial identity to OR choice, as they are part of $GEP_{Ventral}$, are expressed in a DV gradient in both precursors and OSNs, and are bidirectionally sensitive to RA manipulations (Figures 1C, S2B, S5B, and S7G). The expression of *Nfi* and potentially other regulatory DV genes—which reflect DV position—may therefore directly bias the choice process in OSN precursors by regulating heterochromatin deposition in a spatially sensitive manner.

Given these observations, we hypothesized that OR heterochromatin deposition may not vary by zones but instead be organized into a continuous DV gradient, one capable of repressing suboptimal ORs to allow each precursor to express spatially appropriate ORs. To explore this idea, we reanalyzed datasets in which OSNs from the entire epithelium were pooled and the levels of heterochromatin marks on each OR locus were quantified using chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq).^{28,31} If OSN differentiation requires heterochromatinization of co-expressed ORs—and the most-dorsal ORs are transiently expressed and subsequently silenced in all precursors, whereas ventral ORs are expressed only in ventral precursors—then dorsal OR loci should have relatively higher levels of heterochromatin marks in this type of pooled analysis.

Indeed, we observe a strong monotonic relationship ($\rho = 0.89$) between the DV score associated with each OR and the level of heterochromatin observed at its locus (Figures 5H and 5I). In contrast, relatively uniform levels of heterochromatin were observed across all OR loci in ChIP-seq data from OSNs ex-

pressing the ventral Olfr1507 receptor (Figure S7H), consistent with ventral OSNs silencing all more dorsal OR loci. We also observe a continuous DV gradient of interchromosomal OR-OR genomic contacts, in which dorsal OR loci systematically interact more frequently with other OR loci than do ventral OR loci when analyzed in bulk across all OSNs (Figures 5J and S7I). This result implies that more ventral OSNs have nuclei with relatively more interchromosomal OR-OR genomic contacts than do more dorsal OSNs, likely because they have silenced more OR loci. Thus—like OR expression patterns—both heterochromatin levels on OR loci and genomic contacts between them are not organized zonally but systematically vary along a DV gradient (Figure 5D).

If heterochromatin levels regulate which OR is chosen for expression at a given spatial position, then disrupting heterochromatin might alter the relationship between spatial position and OR choice. One strategy for doing so is suggested by the finding that heterochromatinized OR loci are bound by the heterochromatin protein 1 β (HP1 β), which helps to compact DNA into its repressed form.^{68,69} Because HP1 β loss leads to neonatal lethality, we performed scRNA-seq on “HP1 swap” mice, in which HP1 β loss is rescued via HP1 α , which also interacts with chromatin but plays a less prominent role in transcriptional silencing (Figure S7J).^{70–72}

OSNs from HP1 swap mice were dorsalized relative to control mice and had fewer OSNs expressing ventral receptors, similar to what we previously observed after treating mice with drugs that decrease RA signaling (Figures 5K, S7K, and S7L). However, unlike what we observed after RA manipulations, OSNs and precursors derived from HP1 swap mice of a given DV score expressed even more dorsal receptors than would otherwise be expected (Figures 5L, 5M, S7M, and S7N). The fact that these manipulations alter the typical relationship between DV gene expression and OR choice demonstrates that OR heterochromatinization—the result of DV genes, such as *Nfi*—causally translates spatial position into a spatially appropriate distribution of ORs.

Genomic context relates spatial positions to appropriate OR choices

While variation in heterochromatin levels on OR loci provides a compelling means for suppressing the expression of spatially inappropriate receptors, it remains unclear how each OR is linked to its specific spatial location. One possibility is that differences in the binding and/or activity of transcription factors upstream of OR genes organize OR expression in space. Motif enrichment analysis confirmed the presence of putative binding sites for the LHX and EBF transcription factors within most OR promoters.^{73,74} Consistent with the early expression of dorsal ORs during OSN differentiation, ORs with an increased number of binding sites for EBF (which is required for OR expression) were more likely to be dorsal (Figure S8A). Furthermore, motifs for NFI and RAR transcription factors were enriched in the promoters of more ventral ORs, consistent with the graded expression of both *Nfi* and RA-related genes observed in ventral OSNs (Figure S8A).

These observations suggest that local genomic context renders each OR competent for expression at a particular DV

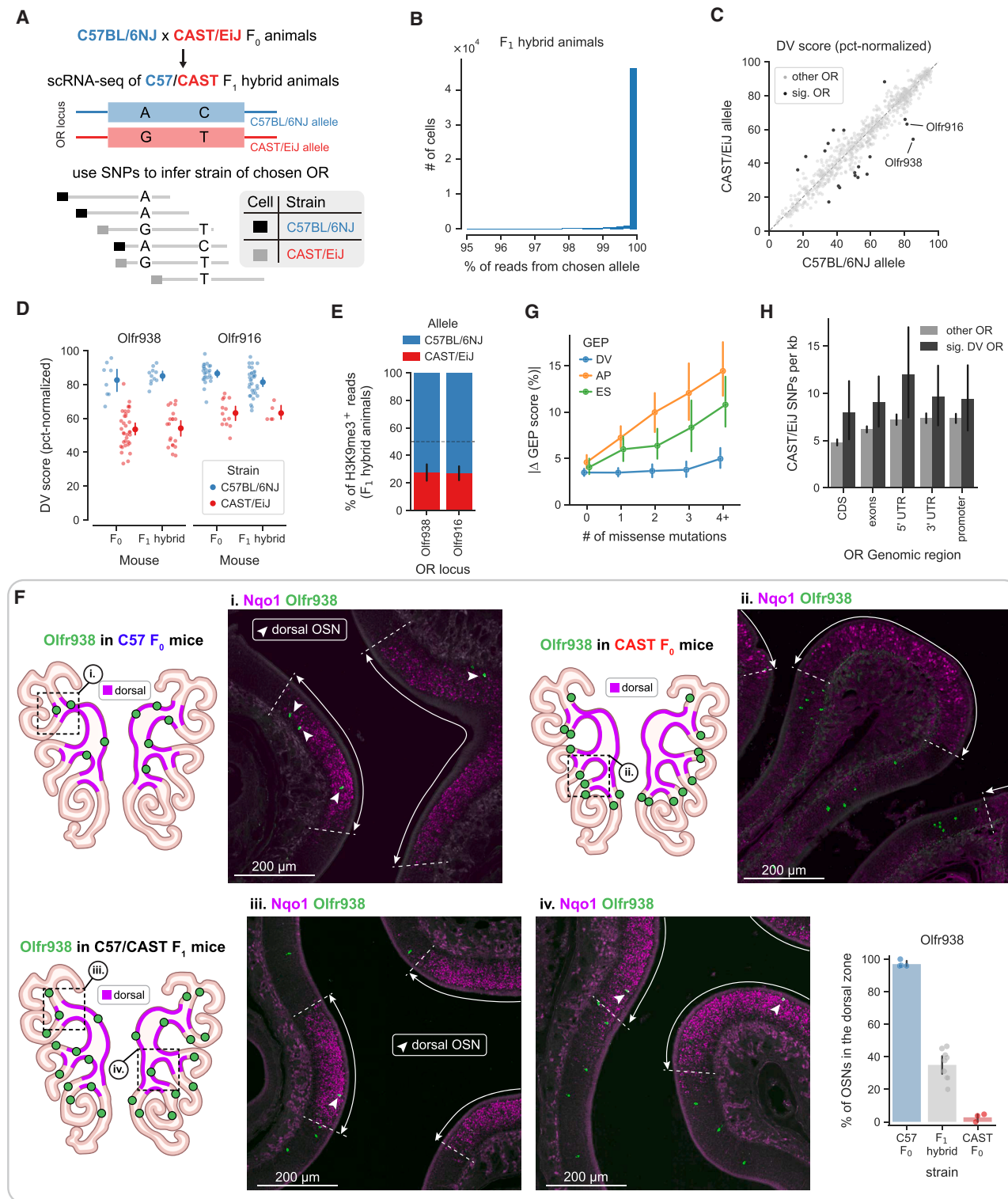


Figure 6. OSN spatial locations are controlled by genomic context

(A) Schematic of F₁ hybrid scRNA-seq experiment.

(B) For OSNs singularly expressing ORs from F₁ animals, the percentage of reads from the chosen allele.

(C) DV scores (percent-normalized across all OSNs) for OSN subtypes in F₁ animals, separated based on the allele of the chosen OR. In black, OSN subtypes with significant DV score changes across strains (see STAR Methods).

(legend continued on next page)

location in the epithelium. Indeed, analysis of “receptor swap” mice, in which the coding sequence of one receptor replaces that of another receptor at a distinct genomic location,²² revealed that OSN DV scores (and expression levels of *Nfi* and other DV genes) reflect the genomic locus in which the expressed OR resides (Figures S8B–S8E).

To further explore the sufficiency of *cis*-regulatory elements in determining OR DV positioning,⁷⁵ we characterized DV scores in OSNs from F_1 hybrid crosses between C57BL/6J (C57) mice and wild-derived CAST/EiJ (CAST) mice (Figure 6A). Consistent with OR genes being expressed mono-allelically, each OSN exclusively expressed either the C57 or CAST allele of its chosen OR (Figure 6B). Whereas these two mouse strains have largely homologous sets of ORs, their genomes also harbor many single-nucleotide polymorphisms (SNPs) that in principle could impact *cis*-regulatory elements and thereby alter OR spatial positions.⁷⁶

Consistent with this prediction, a subset of OSN subtypes (and therefore a subset of OR alleles) exhibited significant across-strain differences in their DV scores in both F_1 hybrid mice and in the parental F_0 C57 and CAST strains (Figures 6C, S9A, and S9B). We wondered whether these strain-specific differences in DV scores reliably predicted differences in the epithelial positions of individual ORs. We therefore performed *in situ* hybridizations for *Olfr938* and *Olfr916*, two ORs with higher DV scores and correspondingly higher levels of heterochromatin in C57 mice compared with those in CAST mice (Figures 6D and 6E). Consistent with the observed differences in their DV scores, OSNs expressing *Olfr938* and *Olfr916* were located exclusively in the *Nqo1*⁺ dorsal region of the epithelium in C57 mice and in the *Nqo1*[−] ventral region in CAST mice (Figures 6F and S9C). Furthermore, *in situ* for *Olfr938* and *Olfr916* in the F_1 hybrids revealed that the same OR was found both within and outside of the *Nqo1*⁺ dorsal region—and thus effectively resided in two “zones” simultaneously—matching the spatial positions of the respective F_0 animals (Figures 6F and S9C).

How might OR genomic variation lead to spatial variation between strains? Genomic SNPs are located both in the upstream and coding regions of OR genes. However, neither strain-specific mutations in OR coding regions nor the number of SNPs in upstream regions predicted whether the position of a given OR varied across strains (Figures 6G and 6H), suggesting that the specific set of upstream SNPs determines their influence on spatial position. Indeed, comparing the spatial pattern of expression of *Olfr938* in CAST, C57, and two additional wild-derived strains (MOLF/EiJ and PWK/PhJ) with the set of SNPs present in each strain nominated a single upstream SNP as a potentially causal determinant of its ventral expression in CAST

mice (Figures S9D–S9F).^{76,77} This analysis, taken with our genetic manipulation and F_1 hybrid experiments, argues that the spatial position of each OR is governed by non-coding elements in its genomic locus.

Phylogeny, function, and spatial location are weakly related

ORs are the product of gene duplication events, which can generate OR genes with similar promoters and primary sequences at nearby genomic locations.^{21,78} This phenomenon raises the possibility that OR genomic position, phylogenetic relationships, and/or chemical tuning properties might correlate with DV scores and therefore spatial positioning. Consistent with prior work that suggested that homologous ORs within a genomic cluster may have similar zonal identities,^{16,79} we find that genomically adjacent OR loci have, on average, more similar protein sequences (as measured by their phylogenetic similarity) and more similar epithelial spatial positions (as captured by the DV score) than do more distant loci (Figure S10A). However, the overall relationship between phylogenetic distance and DV scores was weak (Figures S10B and S10C), suggesting that these effects may only hold for a small subset of ORs, such as those that were the recent products of gene duplication.

Although OR phylogeny only modestly relates to spatial position, it remains possible that OR tuning properties themselves are spatially organized in the epithelium. To address this possibility, we used Act-seq to identify the set of OSNs activated by various monomolecular odorants and to reveal their associated ORs and DV scores.⁵¹ Monomolecular odors delivered at concentrations that activate many receptors triggered patterns of neural activation that were spatially distributed across the entire epithelium (Figure S10D). In contrast, acids largely activated class I ORs in the dorsal region of the epithelium, and odors delivered at concentrations that activated fewer ORs tended to recruit ORs that had, on average, more similar DV scores (Figure S10E). However, this spatial clustering of odor tuning properties was weak and appeared to be largely driven by close phylogenetic relationships among a subset of ORs. These results indicate that ORs with highly related sequences can sometimes be both similarly tuned and located near each other in the epithelium, but that, in general, odors activate OSNs that are spatially distributed across the entire epithelium.

Epithelial spatial positions coordinate axon targeting in the OB

The OB contains a precise and stereotyped map in which each OSN subtype typically projects to one medial and one lateral glomerulus. In addition to transcription factor genes, such as

(D) DV scores, normalized as in (C) with 95% CIs, for OSNs expressing *Olfr938* or *Olfr916* alleles, from either F_0 or F_1 animals.

(E) Relative H3K9me3 levels (via CUT&RUN) at the *Olfr938* and *Olfr916* loci in F_1 animals; note, decreased heterochromatin on CAST alleles is consistent with a more ventral epithelial location.

(F) *In situ* hybridizations for *Olfr938* (green) and the dorsal marker *Nqo1* (magenta, with white outline) in F_0 and F_1 animals. Arrowheads, dorsal *Olfr938*⁺ OSNs. (Lower right) The percentage of dorsal OSNs in each replicate (dots, with 95% CIs).

(G) Change in percent-normalized GEP usage in CAST as a function of OR missense mutations. Environmental state (ES) and AP GEP usages reflect differences in ligand-dependent and ligand-independent OR activity, respectively.⁵¹

(H) SNPs per kb in the indicated genomic regions of each CAST OR locus.

See also Figures S8, S9, and 10.

Nfi and *Rar*, that likely influence OR choice, the DV score includes known regulators of OSN axonal targeting, such as *Robo2* and *Nrp2*, as well as over thirty additional cell surface and cytoskeletal regulatory molecules that could facilitate axon guidance (e.g., *Ncam1*, *Plxna4*, *Plxnb3*, *Cdh3*) (see Table S1; Figures 1C and S5B).^{40–43} Because the expression of DV score genes is independent of OR activity, often precedes OR choice, and is bidirectionally altered upon manipulations of RA signaling, DV GEPs represent a potentially parsimonious mechanism to couple epithelial space to both the process of OR choice and axonal guidance.

To address the possibility that DV genes might link epithelial spatial positions to axonal targets in the OB, we took advantage of a recent spatial transcriptomics dataset that identified the glomeruli for many OSN subtypes.⁸⁰ Coloring identified glomeruli based on the DV scores of their associated receptors revealed that DV scores varied smoothly across the OB and that neighboring glomeruli had similar DV scores (Figures 7A–7C). As a result, the 3D position of each glomerulus predicted the DV score of the innervating OSN subtype with high accuracy ($\rho = 0.95$) (Figure 7D), and conversely, the DV score of each OSN subtype predicted the 3D location of its glomerulus (median error $\approx 400 \mu\text{m}$, 3–4 glomerular widths) (Figure 7E). Thus, variation in DV gene expression (and by extension, epithelial location) across OSN subtypes can at least in part explain the spatial organization of their axonal projections to the bulb.

Notably, the AP GEPs also include genes (such as *Nrp1* and *Plxna1*) known to mediate axon targeting along the bulb AP axis,^{46,47,55} as well as additional axon guidance and cytoskeletal genes (e.g., *Plxna3*, *Sema6c*, *Pcdh11x*, *Cdh15*) (see Table S3). Similar to the DV GEPs, the use of the AP GEPs also smoothly varied across OB glomeruli, although along an orthogonal spatial axis to that used by the DV GEPs and in a manner that was somewhat less predictive of glomerular position (Figures 7A–7E). Glomerular 3D positions for each OSN subtype were best predicted by combining DV and AP GEP usage; the error of these predictions ($\sim 300 \mu\text{m}$) was similar to that observed using models trained with all $\sim 1,300$ OSN variable genes (Figure 7E), indicating these two GEPs together effectively summarize the relevant transcriptional axes that map OSN subtypes onto bulb position. Thus, the two cardinal axes (DV and AP) that organize the OB are topographically and transcriptionally represented in the epithelium itself.

DISCUSSION

The peripheral olfactory system has long been thought to balance determinism and randomness by limiting the expression of individual ORs to one of a handful of DV zones, while randomly distributing ORs within each zone.^{13,14} Here we show instead that the olfactory epithelium harbors an organized map in which each of the $\sim 1,100$ receptors predictably occupies a unique and distinguishable distribution along the DV axis of the epithelium. This pattern reflects a developmental program in which RA signaling transforms spatial position into transcriptional gradients across ~ 250 co-regulated genes. This coherent transcriptional axis—summarized as the DV score—allows us to demonstrate that precursors and OSNs harbor spatially specific

transcriptional identities that restrict OR choice and coordinate the construction of spatial maps in the epithelium and bulb (Figure 7F).

These findings reveal that development can reliably organize $\sim 1,100$ different sensory channels in space. DV position influences many key steps that precede OR choice during OSN differentiation, including OR co-expression, heterochromatin deposition, and interchromosomal genomic interactions (Figure 5D). However, whereas the mean relationship between spatial position and OR choice is essentially invariant at the population level, precursors do not always pick their most preferred OR, resulting in each OR occupying a tight distribution of positions. We speculate that variability in the timing of the steps that translate space into OR choice might allow noise to accumulate during differentiation; consistent with this possibility, DV score distributions are wider for ventral OSN subtypes (Figure S1P).

Our work is consistent with transplantation experiments that have suggested that spatial determinants influence OR choice.⁸¹ RA-related genes like *Aldh1a2* have also been previously proposed to signal information about physical space, but testing the role of RA on OSN differentiation has been challenging given the pleiotropic roles of RA on neuronal development.^{40,63,82–84} Here we implicate RA in organizing the olfactory epithelial receptor map via pharmacological manipulations, which demonstrate that RA controls the coordinated expression of DV genes and spatially restricts OR choice. The ability of a continuous RA gradient to generate discrete variation in OR expression along the DV axis of the olfactory epithelium is reminiscent of its role in development, where, for example, a continuous RA gradient ultimately supports the ontogeny of different somites along the rostrocaudal axis.^{85,86}

What is the meaning or purpose of the receptor map in the nose? Whereas at high concentrations monomolecular odors activate spatially distributed sets of ORs, at lower concentrations the sparser set of activated ORs has, on average, more similar associated DV scores. These activated ORs are nevertheless separated by ~ 100 – 200 intervening ORs, and as such, the observed spatial clustering is modest at best. We hypothesize that the “weak” chemotopy observed here (and by others^{17,87}) is driven by recently duplicated ORs that are more likely to respond to the same or similar odorants. However, as different OR genes accumulate genetic variation, their spatial positions likely diverge due to variation in upstream promoter regions, which are less conserved than coding sequences. Such a mechanism could support robustness, as spatially distributing tuning properties renders the epithelium resilient to local insults. Note that our analyses do not rule out the possibility that chemical odor features not queried here—or other odor features related to meaning or valence—are spatially organized in the epithelium.

Recent results suggest that odor experience can influence the number of OSNs that respond to an experienced odor.^{88–91} Our observation that OSN precursors of a given DV score and therefore spatial identity sit immediately underneath the OSNs to which they give rise might also allow for long-term adaptation in OSN frequency. We speculate that odor responses in OSNs generate signals that increase proliferation and differentiation of underlying stem cells by either diffusing locally or by

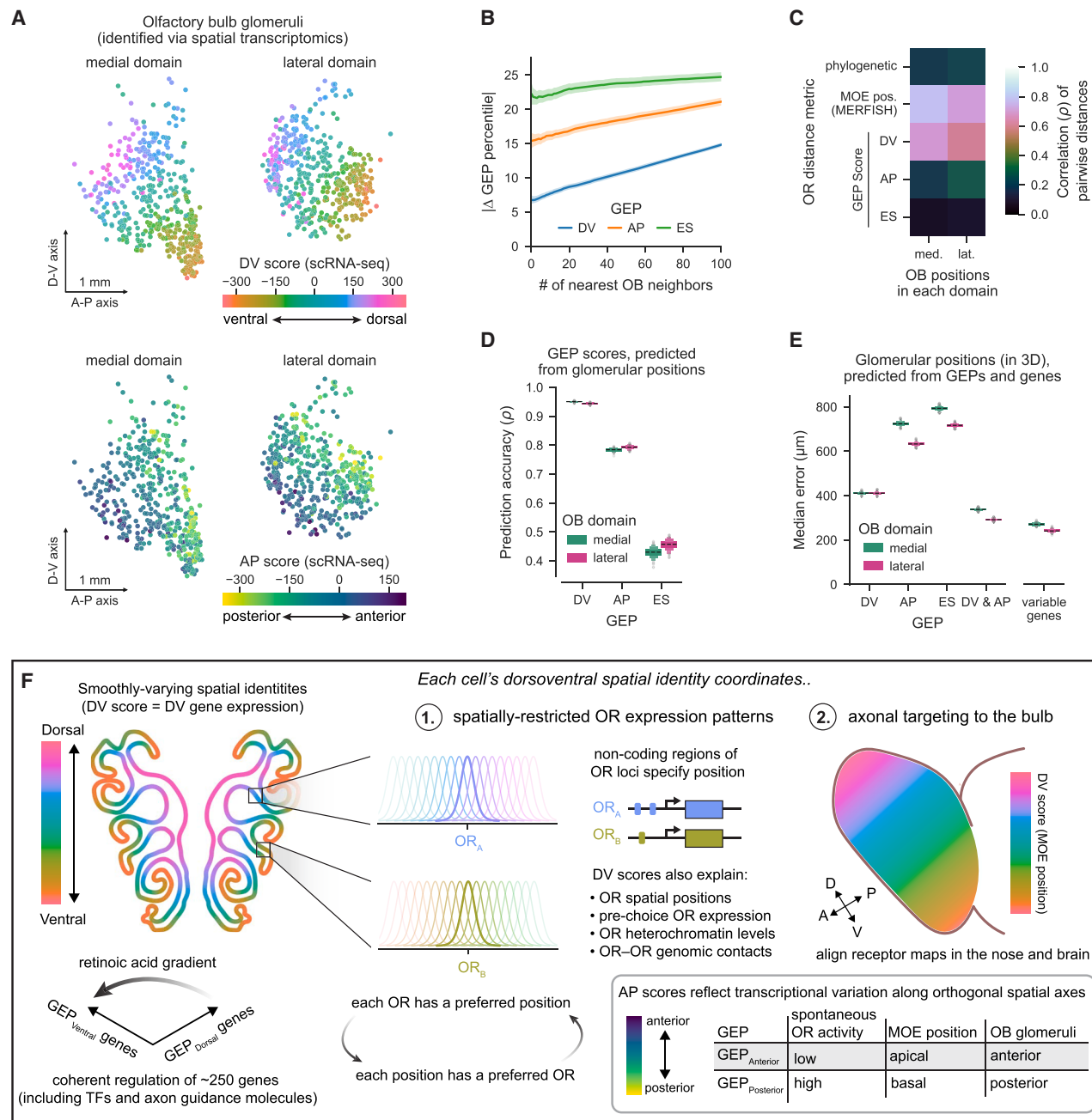


Figure 7. Olfactory bulb glomerular positions reflect epithelial DV scores

(A) Glomerular positions in the medial and lateral OB domain, as identified via spatial transcriptomics,⁸⁰ colored by the DV and AP scores associated with each OSN subtype, as measured via scRNA-seq.

(B) The mean pairwise distance in GEP usage (with 95% CIs) between OSN subtypes whose glomeruli are within given distances.

(C) Correlation (across all OSN subtypes) between pairwise distances of OB glomerular positions and of GEP usages; MOE epithelial positions; or OR protein phylogenetic distances.

(D) The accuracy of regression models (across 1,000 restarts) at predicting GEP scores using glomerular positions.

(E) Similar to (D), but the error in glomerular positions predicted using either GEP scores or all 1,300 OSN variable genes.

(F) Summary schematic depicting how each cell's DV identity coordinates both OR choice during OSN differentiation as well as axonal targeting to the OB, and how AP scores relate to orthogonal spatial axes in the MOE and OB.

stimulating adjacent sustentacular cells. However, because single OSN precursors give rise to multiple OSN subtypes, a main prediction from our results is that the odor experience will recruit not only OSNs whose receptors respond to the experienced odors but also OSNs with similar DV scores; future work will be required to test this hypothesis.

Spatial variation also helps OSNs to simultaneously know both which OR they are to express and where in the brain they should send their axons. In this model, axonal targets are largely determined by the physical location of each OSN in the epithelium rather than by the OR chosen for expression.^{12,48} Nevertheless, ORs likely still play causal roles in glomerular targeting and refinement,^{48,49} as the mapping between spatial positions, DV scores, and OR choice is not fully deterministic. Furthermore, AP genes are receptor-dependent, include key causal genes for guiding axons along the AP bulb axis, improve predictions of glomerular location, and likely play important roles in glomerular targeting.^{47,51,92} Thus, while OSN identities reflect continuous variation in DV gene expression, this continuity is segmented by the expression of single ORs, which ultimately enables accurate targeting to discrete glomeruli.

Limitations of the study

First, much remains to be discovered about how RA patterns the MOE and translates space into coherent patterns of DV gene expression. Additional diffusible factors may also convey spatial information to epithelial cells.^{93,94} This is likely true in the dorsal-most region of the epithelium, which appears less sensitive to RA manipulations than the ventral epithelium and harbors distinct precursor types that generate OSNs expressing class I and trace amine-associated receptors (TAAR)-type receptors^{95–97}; future work is required to understand how space relates to choice for these receptors, and for those few ORs that are organized into epithelial patches.^{98–100} Second, although we find that upstream genomic sequences largely determine the epithelial positioning of each OR (extending past work showing OR promoters can confer spatially restricted transgene expression^{75,101–103}), it remains unclear how a transcriptional code might license individual ORs for choice. Third, while DV genes like *Robo2* and *Nrp2* are known to be required for appropriate glomerular targeting, dissecting the causal role of additional DV genes in OSN differentiation, fate determination, and axon guidance may be challenging due to either redundancy or pleiotropy. Finally, how the OR-dependent AP GEPs influence OSN positioning along the epithelial apical-basal axis remains a mystery.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sandeep Robert Datta (srdatta@hms.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw and processed next-generation sequencing data have been deposited at the NCBI Sequence Read Archive (SRA; accessions SRA: SRP583999 and SRP318630) and Gene Expression Omnibus

(GEO; accessions GEO: GSE297068 and GSE173947) databases and are publicly available as of the date of publication. This paper also analyzes existing, publicly available data (accessible at SRA: SRP096660 and SRP285789).

- Original code for calculating GEP usages of OSNs and for analyzing DV scores can be found at https://github.com/dattalab/Tsukahara_Brann_OSN and https://github.com/dattalab/Brann_olfactory_dorsoventral. They have been deposited at Zenodo ([10.5281/zenodo.5550454](https://doi.org/10.5281/zenodo.5550454) and [10.5281/zenodo.18941308](https://doi.org/10.5281/zenodo.18941308)) and are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank Catherine Dulac for CNGA2 knockout mice and members of the Datta lab for helpful feedback on the manuscript. We thank Jingwen Araki and Julia Martz for laboratory assistance and Euphonia Yoon for illustrations. We acknowledge the Bauer Core Facility at Harvard University, the Flow Cytometry Core and Spatial Technologies Unit (STU) at Beth Israel Deaconess Medical Center, and the Neurobiology Imaging Facility (NIF) and Microscopy Resources on the North Quad (MicRoN) cores at Harvard Medical School. Portions of this research were conducted on the O2 High Performance Compute Cluster at Harvard Medical School. S.R.D. is supported by NIH R01DC021669, R01DC021422, and R01DC021965 and received funding from the Tan-Yang Center. D.H.B. was supported by an NSF Graduate Research Fellowship and by the NIH grant F31DC019017.

AUTHOR CONTRIBUTIONS

D.H.B., T.T., and S.R.D. conceived and designed the study. D.H.B. and T.T. performed the experiments. C.T. and D.K. helped perform the F₁ hybrid experiments, and R.L. and L.T.K. helped perform the clonal lineage tracing and RA experiments, respectively. M.E.-D.-A. provided the HP1 swap mice, and B.B. generated an additional MERFISH dataset. N.K. and M.K. provided the olfactory bulb spatial transcriptomics data and were supervised by A.F. and A.S. S.R.D. acquired funding and supervised the study. T.B. generated the swap mice and provided additional supervision. D.H.B. and S.R.D. wrote the manuscript with input from the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Animals
- [METHOD DETAILS](#)
 - OR activity manipulations
 - Lentiviral and retinoic acid experiments
 - Single-cell RNA-seq and data preprocessing
 - Clonal lineage tracing
 - Spatial profiling via MERFISH
 - Spatial transcriptomics via Stereo-seq
 - Chromatin conformation via Hi-C
 - CUT&RUN
 - In situ hybridizations
 - Data integration of scRNA-seq replicates
 - Identification of zonally-restricted genes
 - Identification of the dorsoventral (DV) score
 - MERFISH analyses
 - Stereo-seq analyses

- DV scores versus other spatial metrics
- Comparison of AP scores with MERFISH data
- DV scores in activity manipulations
- DV scores in mice with OR manipulations
- Gene and OR expression in differentiating OSNs
- Clonal lineage tracing
- Retinoic acid manipulations
- ChIP-seq data analyses
- Hi-C data analyses
- HP1 swap experiments
- Allelic expression in C57/CAST F₁ hybrids
- OR promoter analyses
- Olfactory bulb glomerular positions
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Software used for data analysis
 - Statistical testing

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2026.03.051>.

Received: April 30, 2025

Revised: December 23, 2025

Accepted: March 27, 2026

REFERENCES

1. Kaas, J.H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. [https://doi.org/10.1016/s0361-9230\(97\)00094-4](https://doi.org/10.1016/s0361-9230(97)00094-4).
2. Knudsen, E.I., du Lac, S., and Esterly, S.D. (1987). Computational maps in the brain. *Annu. Rev. Neurosci.* 10, 41–65. <https://doi.org/10.1146/annurev.ne.10.030187.000353>.
3. Brewer, A.A., and Barton, B. (2016). Maps of the Auditory Cortex. *Annu. Rev. Neurosci.* 39, 385–407. <https://doi.org/10.1146/annurev-neuro-070815-014045>.
4. Nauhaus, I., and Nielsen, K.J. (2014). Building maps from maps in primary visual cortex. *Curr. Opin. Neurobiol.* 24, 1–6. <https://doi.org/10.1016/j.conb.2013.08.007>.
5. Abaira, V.E., and Ginty, D.D. (2013). The Sensory Neurons of Touch. *Neuron* 79, 618–639. <https://doi.org/10.1016/j.neuron.2013.07.051>.
6. Barnes, I.H.A., Ibarra-Soria, X., Fitzgerald, S., Gonzalez, J.M., Davidson, C., Hardy, M.P., Manthavadi, D., Van Gerven, L., Jorissen, M., Zeng, Z., et al. (2020). Expert curation of the human and mouse olfactory receptor gene repertoires identifies conserved coding regions split across two exons. *BMC Genomics* 21, 196. <https://doi.org/10.1186/s12864-020-6583-3>.
7. Monahan, K., and Lomvardas, S. (2015). Monoallelic Expression of Olfactory Receptors. *Annu. Rev. Cell Dev. Biol.* 31, 721–740. <https://doi.org/10.1146/annurev-cellbio-100814-125308>.
8. Murthy, V.N. (2011). Olfactory Maps in the Brain. *Annu. Rev. Neurosci.* 34, 233–258. <https://doi.org/10.1146/annurev-neuro-061010-113738>.
9. Imai, T., Sakano, H., and Vossahl, L.B. (2010). Topographic mapping—the olfactory system. *Cold Spring Harb. Perspect. Biol.* 2, a001776. <https://doi.org/10.1101/cshperspect.a001776>.
10. Brann, D.H., and Datta, S.R. (2020). Finding the Brain in the Nose. *Annu. Rev. Neurosci.* 43, 277–295. <https://doi.org/10.1146/annurev-neuro-102119-103452>.
11. Schwob, J.E., and Gottlieb, D.I. (1986). The primary olfactory projection has two chemically distinct zones. *J. Neurosci.* 6, 3393–3404. <https://doi.org/10.1523/jneurosci.06-11-03393.1986>.
12. Zapiec, B., and Mombaerts, P. (2020). The Zonal Organization of Odorant Receptor Gene Choice in the Main Olfactory Epithelium of the Mouse. *Cell Rep.* 30, 4220–4234.e5. <https://doi.org/10.1016/j.celrep.2020.02.110>.
13. Ressler, K.J., Sullivan, S.L., and Buck, L.B. (1993). A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* 73, 597–609. [https://doi.org/10.1016/0092-8674\(93\)90145-G](https://doi.org/10.1016/0092-8674(93)90145-G).
14. Vassar, R., Ngai, J., and Axel, R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* 74, 309–318. [https://doi.org/10.1016/0092-8674\(93\)90422-M](https://doi.org/10.1016/0092-8674(93)90422-M).
15. Miyamichi, K., Serizawa, S., Kimura, H.M., and Sakano, H. (2005). Continuous and Overlapping Expression Domains of Odorant Receptor Genes in the Olfactory Epithelium Determine the Dorsal/Ventral Positioning of Glomeruli in the Olfactory Bulb. *J. Neurosci.* 25, 3586–3592. <https://doi.org/10.1523/jneurosci.0324-05.2005>.
16. Tan, L., and Xie, X.S. (2018). A Near-Complete Spatial Map of Olfactory Receptors in the Mouse Main Olfactory Epithelium. *Chem. Senses* 43, 427–432. <https://doi.org/10.1093/chemse/bjy030>.
17. Ruiz Tejada Segura, M.L., Abou Moussa, E., Garaballo, E., Nakahara, T.S., Makhoul, M., Mathew, L.S., Wang, L., Valle, F., Huang, S.S.Y., Mainland, J.D., et al. (2022). A 3D transcriptomics atlas of the mouse nose sheds light on the anatomical logic of smell. *Cell Rep.* 38, 110547. <https://doi.org/10.1016/j.celrep.2022.110547>.
18. Sullivan, S.L., Ressler, K.J., and Buck, L.B. (1995). Spatial patterning and information coding in the olfactory system. *Curr. Opin. Genet. Dev.* 5, 516–523. [https://doi.org/10.1016/0959-437X\(95\)90057-N](https://doi.org/10.1016/0959-437X(95)90057-N).
19. Iwema, C.L., Fang, H., Kurtz, D.B., Youngentob, S.L., and Schwob, J.E. (2004). Odorant Receptor Expression Patterns Are Restored in Lesion-Recovered Rat Olfactory Epithelium. *J. Neurosci.* 24, 356–369. <https://doi.org/10.1523/jneurosci.1219-03.2004>.
20. Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* 5, 124–133. <https://doi.org/10.1038/nn800>.
21. Bear, D.M., Lassance, J.-M., Hoekstra, H.E., and Datta, S.R. (2016). The Evolving Neural and Genetic Architecture of Vertebrate Olfaction. *Curr. Biol.* 26, R1039–R1049. <https://doi.org/10.1016/j.cub.2016.09.011>.
22. Bozza, T., Vassalli, A., Fuss, S., Zhang, J.-J., Weiland, B., Pacifico, R., Feinstein, P., and Mombaerts, P. (2009). Mapping of Class I and Class II Odorant Receptors to Glomerular Domains by Two Distinct Types of Olfactory Sensory Neurons in the Mouse. *Neuron* 61, 220–233. <https://doi.org/10.1016/j.neuron.2008.11.010>.
23. Pacifico, R., Dewan, A., Cawley, D., Guo, C., and Bozza, T. (2012). An Olfactory Subsystem that Mediates High-Sensitivity Detection of Volatile Amines. *Cell Rep.* 2, 76–88. <https://doi.org/10.1016/j.celrep.2012.06.006>.
24. Shykind, B.M., Rohani, S.C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., and Barnea, G. (2004). Gene Switching and the Stability of Odorant Receptor Gene Choice. *Cell* 117, 801–815. <https://doi.org/10.1016/j.cell.2004.05.015>.
25. Oka, Y., Kobayakawa, K., Nishizumi, H., Miyamichi, K., Hirose, S., Tsu-boi, A., and Sakano, H. (2003). O-MACS, a novel member of the medium-chain acyl-CoA synthetase family, specifically expressed in the olfactory epithelium in a zone-specific manner. *Eur. J. Biochem.* 270, 1995–2004. <https://doi.org/10.1046/j.1432-1033.2003.03571.x>.
26. Yoshihara, Y., Kawasaki, M., Tamada, A., Fujita, H., Hayashi, H., Kagamiyama, H., and Mori, K. (1997). OCAM: A New Member of the Neural Cell Adhesion Molecule Family Related to Zone-to-Zone Projection of Olfactory and Vomeronasal Axons. *J. Neurosci.* 17, 5830–5842. <https://doi.org/10.1523/jneurosci.17-15-05830.1997>.
27. Tan, L., Zong, C., and Xie, X.S. (2013). Rare event of histone demethylation can initiate singular gene expression of olfactory receptors. *Proc. Natl. Acad. Sci. USA* 110, 21148–21152. <https://doi.org/10.1073/pnas.1321511111>.

28. Bashkirova, E.V., Klimpert, N., Monahan, K., Campbell, C.E., Osinski, J., Tan, L., Schieren, I., Pourmorady, A., Stecky, B., Barnea, G., et al. (2023). Opposing, spatially-determined epigenetic forces impose restrictions on stochastic olfactory receptor choice. *eLife* 12, RP87445. <https://doi.org/10.7554/eLife.87445>.
29. Tan, L., Li, Q., and Xie, X.S. (2015). Olfactory sensory neurons transiently express multiple olfactory receptors during development. *Mol. Syst. Biol.* 11, 844. <https://doi.org/10.15252/msb.20156639>.
30. Hanchate, N.K., Kondoh, K., Lu, Z., Kuang, D., Ye, X., Qiu, X., Pachter, L., Trapnell, C., and Buck, L.B. (2015). Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* 350, 1251–1255. <https://doi.org/10.1126/science.aad2456>.
31. Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* 565, 448–453. <https://doi.org/10.1038/s41586-018-0845-0>.
32. Yusuf, N., and Monahan, K. (2024). Epigenetic programming of stochastic olfactory receptor choice. *Genesis* 62, e23593. <https://doi.org/10.1002/dvg.23593>.
33. Bashkirova, E., and Lomvardas, S. (2019). Olfactory receptor genes make the case for inter-chromosomal interactions. *Curr. Opin. Genet. Dev.* 55, 106–113. <https://doi.org/10.1016/j.gde.2019.07.004>.
34. Zapiec, B., and Mombaerts, P. (2015). Multiplex assessment of the positions of odorant receptor-specific glomeruli in the mouse olfactory bulb by serial two-photon tomography. *Proc. Natl. Acad. Sci. USA* 112, E5873–E5882. <https://doi.org/10.1073/pnas.1512135112>.
35. Soucy, E.R., Albeanu, D.F., Fantana, A.L., Murthy, V.N., and Meister, M. (2009). Precision and diversity in an odor map on the olfactory bulb. *Nat. Neurosci.* 12, 210–220. <https://doi.org/10.1038/nn.2262>.
36. Strotmann, J., Conzelmann, S., Beck, A., Feinstein, P., Breer, H., and Mombaerts, P. (2000). Local Permutations in the Glomerular Array of the Mouse Olfactory Bulb. *J. Neurosci.* 20, 6927–6938. <https://doi.org/10.1523/jneurosci.20-18-06927.2000>.
37. Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J., and Axel, R. (1996). Visualizing an Olfactory Sensory Map. *Cell* 87, 675–686. [https://doi.org/10.1016/S0092-8674\(00\)81387-2](https://doi.org/10.1016/S0092-8674(00)81387-2).
38. Mori, K., and Sakano, H. (2011). How Is the Olfactory Map Formed and Interpreted in the Mammalian Brain? *Annu. Rev. Neurosci.* 34, 467–499. <https://doi.org/10.1146/annurev-neuro-112210-112917>.
39. Ma, M., and Shepherd, G.M. (2000). Functional mosaic organization of mouse olfactory receptor neurons. *Proc. Natl. Acad. Sci. USA* 97, 12869–12874. <https://doi.org/10.1073/pnas.220301797>.
40. Norlin, E.M., Alenius, M., Gussing, F., Hägglund, M., Vedin, V., and Böhm, S. (2001). Evidence for Gradients of Gene Expression Correlating with Zonal Topography of the Olfactory Sensory Map. *Mol. Cell. Neurosci.* 18, 283–295. <https://doi.org/10.1006/mcne.2001.1019>.
41. Takahashi, H., Yoshihara, S.-I., Nishizumi, H., and Tsuboi, A. (2010). Neuropeptide-2 is required for the proper targeting of ventral glomeruli in the mouse olfactory bulb. *Mol. Cell. Neurosci.* 44, 233–245. <https://doi.org/10.1016/j.mcn.2010.03.010>.
42. Cho, J.H., Prince, J.E.A., Cutforth, T., and Cloutier, J.-F. (2011). The Pattern of Glomerular Map Formation Defines Responsiveness to Aversive Odorants in Mice. *J. Neurosci.* 31, 7920–7926. <https://doi.org/10.1523/jneurosci.2460-10.2011>.
43. Takeuchi, H., Inokuchi, K., Aoki, M., Suto, F., Tsuboi, A., Matsuda, I., Suzuki, M., Aiba, A., Serizawa, S., Yoshihara, Y., et al. (2010). Sequential Arrival and Graded Secretion of Sema3F by Olfactory Neuron Axons Specify Map Topography at the Bulb. *Cell* 141, 1056–1067. <https://doi.org/10.1016/j.cell.2010.04.041>.
44. Bozza, T., Feinstein, P., Zheng, C., and Mombaerts, P. (2002). Odorant Receptor Expression Defines Functional Units in the Mouse Olfactory System. *J. Neurosci.* 22, 3033–3043. <https://doi.org/10.1523/jneurosci.22-08-03033.2002>.
45. Wang, F., Nemes, A., Mendelsohn, M., and Axel, R. (1998). Odorant Receptors Govern the Formation of a Precise Topographic Map. *Cell* 93, 47–60. [https://doi.org/10.1016/S0092-8674\(00\)81145-9](https://doi.org/10.1016/S0092-8674(00)81145-9).
46. Sakano, H. (2020). Developmental regulation of olfactory circuit formation in mice. *Dev. Growth Differ.* 62, 199–213. <https://doi.org/10.1111/dgd.12657>.
47. Nakashima, A., Takeuchi, H., Imai, T., Saito, H., Kiyonari, H., Abe, T., Chen, M., Weinstein, L.S., Yu, C.R., Storm, D.R., et al. (2013). Agonist-Independent GPCR Activity Regulates Anterior-Posterior Targeting of Olfactory Sensory Neurons. *Cell* 154, 1314–1325. <https://doi.org/10.1016/j.cell.2013.08.033>.
48. Feinstein, P., and Mombaerts, P. (2004). A Contextual Model for Axonal Sorting into Glomeruli in the Mouse Olfactory System. *Cell* 117, 817–831. <https://doi.org/10.1016/j.cell.2004.05.011>.
49. Serizawa, S., Miyamichi, K., Takeuchi, H., Yamagishi, Y., Suzuki, M., and Sakano, H. (2006). A Neuronal Identity Code for the Odorant Receptor-Specific and Activity-Dependent Axon Sorting. *Cell* 127, 1057–1069. <https://doi.org/10.1016/j.cell.2006.10.031>.
50. Gussing, F., and Böhm, S. (2004). NQO1 activity in the main and the accessory olfactory systems correlates with the zonal topography of projection maps. *Eur. J. Neurosci.* 19, 2511–2518. <https://doi.org/10.1111/j.0953-816x.2004.03331.x>.
51. Tsukahara, T., Brann, D.H., Pashkovski, S.L., Guitchounts, G., Bozza, T., and Datta, S.R. (2021). A transcriptional rheostat couples past activity to future sensory responses. *Cell* 184, 6326–6343.e32. <https://doi.org/10.1016/j.cell.2021.11.022>.
52. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* 8, e43803. <https://doi.org/10.7554/eLife.43803>.
53. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090. <https://doi.org/10.1126/science.aaa6090>.
54. Bintu, B., Isogai, Y., Zhuang, X., and Dulac, C. (2026). Molecular and Spatial Organization of the Primary Olfactory System and its Responses to Social Odors. *Cell* 189, "bxs. <https://doi.org/10.1016/j.cell.2026.03.053>.
55. Imai, T., Suzuki, M., and Sakano, H. (2006). Odorant receptor-derived cAMP signals direct axonal targeting. *Science* 314, 657–661. <https://doi.org/10.1126/science.1131794>.
56. Fleischmann, A., Abdus-Saboor, I., Sayed, A., and Shykind, B. (2013). Functional Interrogation of an Odorant Receptor Locus Reveals Multiple Axes of Transcriptional Regulation. *PLoS Biol.* 11, e1001568. <https://doi.org/10.1371/journal.pbio.1001568>.
57. Zhao, H., and Reed, R.R. (2001). X Inactivation of the *OCNC1* Channel Gene Reveals a Role for Activity-Dependent Competition in the Olfactory System. *Cell* 104, 651–660. [https://doi.org/10.1016/S0092-8674\(01\)00262-8](https://doi.org/10.1016/S0092-8674(01)00262-8).
58. Fletcher, R.B., Das, D., Gadye, L., Street, K.N., Baudhuin, A., Wagner, A., Cole, M.B., Flores, Q., Choi, Y.G., Yosef, N., et al. (2017). Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell* 20, 817–830.e8. <https://doi.org/10.1016/j.stem.2017.04.003>.
59. Hussaini, M., Korsching, S.I., and Tresch, A. (2022). Pseudotime analysis reveals novel regulatory factors for multigenic onset and monogenic transition of odorant receptor expression. *Sci. Rep.* 12, 16183. <https://doi.org/10.1038/s41598-022-20106-w>.
60. Bergström, U., Giovanetti, A., Piras, E., and Brittebo, E.B. (2003). Methimazole-Induced Damage in the Olfactory Mucosa: Effects on Ultrastructure and Glutathione Levels. *Toxicol. Pathol.* 31, 379–387. <https://doi.org/10.1080/01926230390201101>.
61. Whitesides, J., Hall, M., Anchan, R., and LaMantia, A.S. (1998). Retinoid signaling distinguishes a subpopulation of olfactory receptor neurons in the developing and adult mouse. *J. Comp. Neurol.* 394, 445–461.

[https://doi.org/10.1002/\(sici\)1096-9861\(19980518\)394:4<445::aid-cne4>3.0.co;2-1](https://doi.org/10.1002/(sici)1096-9861(19980518)394:4<445::aid-cne4>3.0.co;2-1).

62. Asson-Batres, M.A., and Smith, W.B. (2006). Localization of retinaldehyde dehydrogenases and retinoid binding proteins to sustentacular cells, glia, Bowman's gland cells, and stroma: Potential sites of retinoic acid synthesis in the postnatal rat olfactory organ. *J. Comp. Neurol.* 496, 149–171. <https://doi.org/10.1002/cne.20904>.
63. Peluso, C.E., Jang, W., Dräger, U.C., and Schwob, J.E. (2012). Differential expression of components of the retinoic acid signaling pathway in the adult mouse olfactory epithelium. *J. Comp. Neurol.* 520, 3707–3726. <https://doi.org/10.1002/cne.23124>.
64. Pourmorady, A.D., Bashkirova, E.V., Chiariello, A.M., Belagzhal, H., Kodra, A., Duffié, R., Kahiapo, J., Monahan, K., Pulupa, J., Schieren, I., et al. (2024). RNA-mediated symmetry breaking enables singular olfactory receptor choice. *Nature* 625, 181–188. <https://doi.org/10.1038/s41586-023-06845-4>.
65. Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E.S., and Lomvardas, S. (2017). Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *eLife* 6, e28620. <https://doi.org/10.7554/eLife.28620>.
66. Wu, H., Zhang, J., Jian, F., Chen, J.P., Zheng, Y., Tan, L., and Sunney Xie, X. (2024). Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat. Methods* 21, 974–982. <https://doi.org/10.1038/s41592-024-02239-0>.
67. Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C., et al. (2011). An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell* 145, 555–570. <https://doi.org/10.1016/j.cell.2011.03.040>.
68. Clowney, E.J., LeGros, M.A., Mosley, C.P., Clowney, F.G., Markenscoff-Papadimitriou, E.C., Myllys, M., Barnea, G., Larabell, C.A., and Lomvardas, S. (2012). Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* 151, 724–737. <https://doi.org/10.1016/j.cell.2012.09.043>.
69. Le Gros, M.A., Clowney, E.J., Magklara, A., Yen, A., Markenscoff-Papadimitriou, E., Colquitt, B., Myllys, M., Kellis, M., Lomvardas, S., and Larabell, C.A. (2016). Soft X-Ray Tomography Reveals Gradual Chromatin Compaction and Reorganization during Neurogenesis In Vivo. *Cell Rep.* 17, 2125–2136. <https://doi.org/10.1016/j.celrep.2016.10.060>.
70. Bosch-Presegué, L., Raurell-Vila, H., Thackray, J.K., González, J., Casal, C., Kane-Goldsmith, N., Vizoso, M., Brown, J.P., Gómez, A., Ausió, J., et al. (2017). Mammalian HP1 Isoforms Have Specific Roles in Heterochromatin Structure and Organization. *Cell Rep.* 21, 2048–2057. <https://doi.org/10.1016/j.celrep.2017.10.092>.
71. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I.A., Zenn, H.M., Kost, N., Pohl, W., et al. (2016). Dynamic and flexible H3K9me3 bridging via HP1 β dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.* 7, 11310. <https://doi.org/10.1038/ncomms11310>.
72. Escamilla-del-Arenal, M., Duffié, R., Shayya, H., Loconte, V., Ekman, A., Street, L.A., Horta, A., Canzio, D., Monahan, K., Larabell, C., et al. (2026). HP1 β and H3K9me3 Regulate Olfactory Receptor Choice and Transcriptional Identity. *Int. J. Mol. Sci.* 27, 2958. <https://doi.org/10.3390/ijms27072958>.
73. Clowney, E.J., Magklara, A., Colquitt, B.M., Pathak, N., Lane, R.P., and Lomvardas, S. (2011). High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res.* 21, 1249–1259. <https://doi.org/10.1101/gr.120162.110>.
74. Plessy, C., Pascarella, G., Bertin, N., Akalin, A., Carrieri, C., Vassalli, A., Lazarevic, D., Severin, J., Vlachouli, C., Simone, R., et al. (2012). Promoter architecture of mouse olfactory receptor genes. *Genome Res.* 22, 486–497. <https://doi.org/10.1101/gr.126201.111>.
75. Qasba, P., and Reed, R.R. (1998). Tissue and Zonal-Specific Expression of an Olfactory Receptor Transgene. *J. Neurosci.* 18, 227–236. <https://doi.org/10.1523/jneurosci.18-01-00227.1998>.
76. Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294. <https://doi.org/10.1038/nature10413>.
77. Doran, A.G., Wong, K., Flint, J., Adams, D.J., Hunter, K.W., and Keane, T.M. (2016). Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. *Genome Biol.* 17, 167. <https://doi.org/10.1186/s13059-016-1024-y>.
78. Niimura, Y. (2012). Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. *Curr. Genomics* 13, 103–114. <https://doi.org/10.2174/138920212799860706>.
79. Tsuboi, A., Yoshihara, S.-I., Yamazaki, N., Kasai, H., Asai-Tsuboi, H., Komatsu, M., Serizawa, S., Ishii, T., Matsuda, Y., Nagawa, F., Sakano, H., et al. (1999). Olfactory Neurons Expressing Closely Linked and Homologous Odorant Receptor Genes Tend to Project Their Axons to Neighboring Glomeruli on the Olfactory Bulb. *J. Neurosci.* 19, 8409–8418. <https://doi.org/10.1523/jneurosci.19-19-08409.1999>.
80. Klimpert, N., Kollo, M., Brann, D.H., Tan, C., Barry, D., Ma, Y., Schaefer, A.T., and Fleischmann, A. (2025). 3D spatial transcriptomics reveals the molecular structure of input and output pathways in the mouse olfactory bulb. Preprint at bioRxiv. <https://doi.org/10.1101/2025.02.19.639192>.
81. Coleman, J.H., Lin, B., Louie, J.D., Peterson, J., Lane, R.P., and Schwob, J.E. (2019). Spatial Determination of Neuronal Diversification in the Olfactory Epithelium. *J. Neurosci.* 39, 814–832. <https://doi.org/10.1523/jneurosci.3594-17.2018>.
82. Hägglund, M., Berghard, A., Strotmann, J.R., and Böhm, S. (2006). Retinoic Acid Receptor-Dependent Survival of Olfactory Sensory Neurons in Postnatal and Adult Mice. *J. Neurosci.* 26, 3281–3291. <https://doi.org/10.1523/jneurosci.4955-05.2006>.
83. Login, H., Hägglund, S., Berghard, A., and Böhm, S. (2015). The Stimulus-Dependent Gradient of Cyp26B1+ Olfactory Sensory Neurons Is Necessary for the Functional Integrity of the Olfactory Sensory Map. *J. Neurosci.* 35, 13807–13818. <https://doi.org/10.1523/jneurosci.2247-15.2015>.
84. Paschaki, M., Cammas, L., Muta, Y., Matsuoka, Y., Mak, S.-S., Rataj-Baniowska, M., Fraulob, V., Dollé, P., and Ladher, R.K. (2013). Retinoic acid regulates olfactory progenitor cell fate and differentiation. *Neural Dev* 8, 13. <https://doi.org/10.1186/1749-8104-8-13>.
85. Shimozone, S., Imura, T., Kitaguchi, T., Higashijima, S.-I., and Miyawaki, A. (2013). Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* 496, 363–366. <https://doi.org/10.1038/nature12037>.
86. Cunningham, T.J., and Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* 16, 110–123. <https://doi.org/10.1038/nrm3932>.
87. Mori, K., Takahashi, Y.K., Igarashi, K.M., and Yamaguchi, M. (2006). Maps of Odorant Molecular Features in the Mammalian Olfactory Bulb. *Physiol. Rev.* 86, 409–433. <https://doi.org/10.1152/physrev.00021.2005>.
88. Liff, C.W., Ayman, Y.R., Jaeger, E.C.B., Lee, H.S., Kim, A., Albarracín, A.V., and Marlin, B.J. (2023). Fear conditioning biases olfactory stem cell receptor fate. *eLife* 12, RP92882. <https://doi.org/10.7554/eLife.92882.1>.
89. Vihani, A., Hu, X.S., Gundala, S., Koyama, S., Block, E., and Matsunami, H. (2020). Semiochemical responsive olfactory sensory neurons are sexually dimorphic and plastic. *eLife* 9, e54501. <https://doi.org/10.7554/eLife.54501>.
90. van der Linden, C.J., Gupta, P., Bhuiya, A.I., Riddick, K.R., Hossain, K., and Santoro, S.W. (2020). Olfactory Stimulation Regulates the Birth of

- Neurons That Express Specific Odorant Receptors. *Cell Rep.* 33, 108210. <https://doi.org/10.1016/j.celrep.2020.108210>.
91. Hossain, K., Smith, M., Rufenacht, K.E., O'Rourke, R., and Santoro, S.W. (2025). In mice, discrete odors can selectively promote the neurogenesis of sensory neuron subtypes that they stimulate. *eLife* 13, RP96152. <https://doi.org/10.7554/eLife.96152.3>.
92. Imai, T., Yamazaki, T., Kobayakawa, R., Kobayakawa, K., Abe, T., Suzuki, M., and Sakano, H. (2009). Pre-Target Axon Sorting Establishes the Neural Map Topography. *Science* 325, 585–590. <https://doi.org/10.1126/science.1173596>.
93. Yang, L.M., Huh, S.-H., and Ornitz, D.M. (2018). FGF20-Expressing, Wnt-Responsive Olfactory Epithelial Progenitors Regulate Underlying Turbinate Growth to Optimize Surface Area. *Dev. Cell* 46, 564–580.e5. <https://doi.org/10.1016/j.devcel.2018.07.010>.
94. Herrick, D.B., Lin, B., Peterson, J., Schnitke, N., and Schwob, J.E. (2017). Notch1 maintains dormancy of olfactory horizontal basal cells, a reserve neural stem cell. *Proc. Natl. Acad. Sci. USA* 114, E5589–E5598. <https://doi.org/10.1073/pnas.1701333114>.
95. Enomoto, T., Nishida, H., Iwata, T., Fujita, A., Nakayama, K., Kashiwagi, T., Hatanaka, Y., Kondo, H., Kajitani, R., Itoh, T., et al. (2019). Bcl11b controls odorant receptor class choice in mice. *Commun. Biol.* 2, 296. <https://doi.org/10.1038/s42003-019-0536-x>.
96. Iwata, T., Tomeoka, S., and Hirota, J. (2021). A class I odorant receptor enhancer shares a functional motif with class II enhancers. *Sci. Rep.* 11, 510. <https://doi.org/10.1038/s41598-020-79980-x>.
97. Cichy, A., Shah, A., Dewan, A., Kaye, S., and Bozza, T. (2019). Genetic Depletion of Class I Odorant Receptors Impacts Perception of Carboxylic Acids. *Curr. Biol.* 29, 2687–2697.e4. <https://doi.org/10.1016/j.cub.2019.06.085>.
98. Shah, A., Ratkowski, M., Rosa, A., Feinstein, P., and Bozza, T. (2021). Olfactory expression of trace amine-associated receptors requires cooperative cis-acting enhancers. *Nat. Commun.* 12, 3797. <https://doi.org/10.1038/s41467-021-23824-3>.
99. Fei, A., Wu, W., Tan, L., Tang, C., Xu, Z., Huo, X., Bao, H., Kong, Y., Johnson, M., Hartmann, G., et al. (2021). Coordination of two enhancers drives expression of olfactory trace amine-associated receptors. *Nat. Commun.* 12, 3798. <https://doi.org/10.1038/s41467-021-23823-4>.
100. Strotmann, J., Wanner, I., Helfrich, T., Beck, A., Meinken, C., Kubick, S., and Breer, H. (1994). Olfactory neurones expressing distinct odorant receptor subtypes are spatially segregated in the nasal neuroepithelium. *Cell Tissue Res.* 276, 429–438. <https://doi.org/10.1007/BF00343941>.
101. Rothman, A., Feinstein, P., Hirota, J., and Mombaerts, P. (2005). The promoter of the mouse odorant receptor gene M71. *Mol. Cell. Neurosci.* 28, 535–546. <https://doi.org/10.1016/j.mcn.2004.11.006>.
102. Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., and Mombaerts, P. (2002). Minigenes Impart Odorant Receptor-Specific Axon Guidance in the Olfactory Bulb. *Neuron* 35, 681–696. [https://doi.org/10.1016/S0896-6273\(02\)00793-6](https://doi.org/10.1016/S0896-6273(02)00793-6).
103. Zhang, Y.-Q., Breer, H., and Strotmann, J. (2007). Promotor elements governing the clustered expression pattern of odorant receptor genes. *Mol. Cell. Neurosci.* 36, 95–107. <https://doi.org/10.1016/j.mcn.2007.06.005>.
104. Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons. *Nature* 428, 44–49. <https://doi.org/10.1038/nature02375>.
105. Yu, C.R., Power, J., Barnea, G., O'Donnell, S., Brown, H.E.V., Osborne, J., Axel, R., and Gogos, J.A. (2004). Spontaneous Neural Activity Is Required for the Establishment and Maintenance of the Olfactory Sensory Map. *Neuron* 42, 553–566. [https://doi.org/10.1016/S0896-6273\(04\)00224-7](https://doi.org/10.1016/S0896-6273(04)00224-7).
106. Hébert, J.M., and McConnell, S.K. (2000). Targeting of cre to the Foxg1 (BF-1) Locus Mediates loxP Recombination in the Telencephalon and Other Developing Head Structures. *Dev. Biol.* 222, 296–306. <https://doi.org/10.1006/dbio.2000.9732>.
107. Gong, C., Li, S., Wang, L., Zhao, F., Fang, S., Yuan, D., Zhao, Z., He, Q., Li, M., Liu, W., et al. (2024). SAW: an efficient and accurate data analysis workflow for Stereo-seq spatial transcriptomics. *GigaByte* 2024, gigabyte111. <https://doi.org/10.46471/gigabyte.111>.
108. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
109. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., et al. (2022). A Python library for probabilistic analysis of single-cell omics data. *Nature Biotechnology* 40, 163–166. <https://doi.org/10.1038/s41587-021-01206-w>.
110. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
111. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35, 316–319. <https://doi.org/10.1038/nbt.3820>.
112. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* 38, 276–278. <https://doi.org/10.1038/s41587-020-0439-x>.
113. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
114. Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1303.3997>.
115. Abdennur, N., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., and Venev, S.V. (2024). Pairtools: From sequencing data to chromosome contacts. *PLoS Comput. Biol.* 20, e1012164. <https://doi.org/10.1371/journal.pcbi.1012164>.
116. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
117. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* 7, 16878. <https://doi.org/10.1038/s41598-017-17204-5>.
118. Suzukawa, K., Kondo, K., Kanaya, K., Sakamoto, T., Watanabe, K., Ushio, M., Kaga, K., and Yamasoba, T. (2011). Age-related changes of the regeneration mode in the mouse peripheral olfactory system following olfactotoxic drug methimazole-induced damage. *J. Comp. Neurol.* 519, 2154–2174. <https://doi.org/10.1002/cne.22611>.
119. Hogarth, C.A., Evanoff, R., Mitchell, D., Kent, T., Small, C., Amory, J.K., and Griswold, M.D. (2013). Turning a Spermatogenic Wave into a Tsunami: Synchronizing Murine Spermatogenesis Using WIN 18,446. *Biol. Reprod.* 88, 40. <https://doi.org/10.1095/biolreprod.112.105346>.
120. Finlay, J.B., Brann, D.H., Abi Hachem, R., Jang, D.W., Oliva, A.D., Ko, T., Gupta, R., Wellford, S.A., Moseman, E.A., Jang, S.S., et al. (2022). Persistent post-COVID-19 smell loss is associated with immune cell infiltration and altered gene expression in olfactory epithelium. *Sci. Transl. Med.* 14, eadd0484. <https://doi.org/10.1126/scitranslmed.add0484>.
121. Hsieh, T.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of

- Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* 78, 539–553.e8. <https://doi.org/10.1016/j.molcel.2020.03.002>.
122. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753. <https://doi.org/10.1242/dev.165753>.
 123. Barrios, A.W., Núñez, G., Sánchez Quinteiro, P., and Salazar, I. (2014). Anatomy, histochemistry, and immunohistochemistry of the olfactory subsystems in mice. *Front. Neuroanat.* 8, 63. <https://doi.org/10.3389/fnana.2014.00063>.
 124. Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., et al. (2022). Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* 40, 121–130. <https://doi.org/10.1038/s41587-021-01001-7>.
 125. Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madisson, E., Markov, N.S., Zaragosi, L.-E., Ji, Y., Ansari, M., et al. (2023). An integrated cell atlas of the lung in health and disease. *Nat. Med.* 29, 1563–1577. <https://doi.org/10.1038/s41591-023-02327-2>.
 126. Clifton, K., Anant, M., Aihara, G., Atta, L., Aimuwu, O.K., Kebschull, J.M., Miller, M.I., Tward, D., and Fan, J. (2023). STalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat. Commun.* 14, 8123. <https://doi.org/10.1038/s41467-023-43915-7>.
 127. Petukhov, V., Xu, R.J., Soldatov, R.A., Cadinu, P., Khodosevich, K., Mofitt, J.R., and Kharchenko, P.V. (2022). Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* 40, 345–354. <https://doi.org/10.1038/s41587-021-01044-w>.
 128. McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1802.03426>.
 129. Jones, D.C., Elz, A.E., Hadadianpour, A., Ryu, H., Glass, D.R., and Newell, E.W. (2025). Cell simulation as cell segmentation. *Nat. Methods* 22, 1331–1342. <https://doi.org/10.1038/s41592-025-02697-0>.
 130. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
 131. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>.
 132. Kenney, M., Vasylieva, I., Hood, G., Cao-Berg, I., Tuite, L., Laghaei, R., Smith, M.C., Watson, A.M., and Ropelewski, A.J. (2024). The Brain Image Library: A Community-Contributed Microscopy Resource for Neuroscientists. *Sci. Data* 11, 1212. <https://doi.org/10.1038/s41597-024-03761-8>.
 133. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rothen, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. <https://doi.org/10.1126/science.aad0501>.
 134. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 367, eaaw3381. <https://doi.org/10.1126/science.aaw3381>.
 135. Lilue, J., Doran, A.G., Fiddes, I.T., Abrudan, M., Armstrong, J., Bennett, R., Chow, W., Collins, J., Collins, S., Czechanski, A., et al. (2018). Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat. Genet.* 50, 1574–1583. <https://doi.org/10.1038/s41588-018-0223-8>.
 136. Rogozhnikov, A., Ramkumar, P., Shah, K., Bedi, R., Kato, S., and Escala, G.S. (2021). Demuxalot: scaled up genetic demultiplexing for single-cell sequencing. Preprint at bioRxiv. <https://doi.org/10.1101/2021.05.22.443646>.
 137. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
 138. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat Polyclonal anti-Nqo1	Abcam	Cat# ab2346; RRID: AB_302995
Rabbit polyclonal H3K9me3	Abcam	Cat # ab8898; RRID: AB_306848
Bacterial and virus strains		
CloneTracker XP 10M Barcode-3' Lentiviral Library	Cellecta	Cat# BCXP10M3VP-V
Biological samples		
Mouse olfactory epithelium	This paper	N/A
Chemicals, peptides, and recombinant proteins		
WIN 18,446	Santa Cruz Biotechnology	Cat# sc-295819; CAS 1477-57-2
All-trans Retinoic Acid	Santa Cruz Biotechnology	Cat# sc-200898; CAS 302-79-4
BMS493	MedChemExpress	Cat# HY-108529; CAS 215030-90-3
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat# D8418; CAS 67-68-5
Methimazole	Sigma-Aldrich	Cat # M8506; CAS 60-56-0
Dipropylene glycol (DPG)	Sigma-Aldrich	Cat# D215554; CAS 25265-71-8
Actinomycin D	Sigma-Aldrich	Cat# A9415; CAS 50-76-0
Triptolide	Sigma-Aldrich	Cat# T3652; CAS 38748-32-2
Anisomycin	Sigma-Aldrich	Cat# A9789; CAS 22862-76-6
37% Formaldehyde solution	Sigma-Aldrich	Cat# 252549
Papain	Worthington	Cat# LK003178
DNase I	Worthington	Cat# LK003172
EBSS	Worthington	Cat# LK003188
Hibernate-A	Thermo Fisher	Cat# A1247501
Heat-inactivated Fetal Bovine Serum	Thermo Fisher	Cat# 16140071
Propidium iodide	Thermo Fisher	Cat# P1304MP
Disuccinimidyl glutarate (DSG)	Thermo Fisher	Cat# A35392
HCR buffers	Molecular Instruments	N/A
Critical commercial assays		
Chromium Next GEM Single Cell 3' Kit v3.1	10x Genomics	Cat# 1000268
GEM-X Universal 3' Gene Expression v4	10x Genomics	Cat# 1000691
Dual Index Kit TT Set A	10x Genomics	Cat# 1000215
MERSCOPE 300-gene imaging kit	Vizgen	Cat# 10400005
MERSCOPE Cell Boundary Staining Kit	Vizgen	Cat# 10400118
Stereo-seq Transcriptomics T kit v1.3	STOmics	Cat# 211KT13114-CG
Stereo-seq 16 Barcode Library Preparation Kit	STOmics	Cat# 111KL160-CG
Dovetail Epigenetics Micro-C Full Solution	Dovetail Genomics	Cat# 20101E
CUTANA ChIC/CUT&RUN Kit	EpiCypher	Cat# 14-1048
CUTANA CUT&RUN Library Prep Kit	EpiCypher	Cat# 14-1002
Agilent High Sensitivity DNA kit	Agilent Technologies	Cat# 5067-4626
KAPA Library Quantification Kit	Roche	Cat# KK4824
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat# Q32851
Qubit ssDNA Assay Kit	Thermo Fisher Scientific	Cat# Q10212
Deposited data		
Raw and analyzed sequencing data	This paper	GEO: GSE297068
scRNA-seq data	Tsukahara et al. ⁵¹	GEO: GSE173947

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ChIP-seq data	Monahan et al. ⁶⁵	SRA: SRP096660
RNA-seq and ChIP-seq data	Bashkirova et al. ²⁸	SRA: SRP285789
Experimental models: Organisms/strains		
Mouse: C57BL/6J	The Jackson Laboratory	JAX: 000664; RRID:IMSR_JAX:000664
Mouse: CAST/EiJ	The Jackson Laboratory	JAX: 000928; RRID:IMSR_JAX:000928
Mouse: MOLF/EiJ	The Jackson Laboratory	JAX: 000550; RRID:IMSR_JAX:000550
Mouse: PWK/PhJ	The Jackson Laboratory	JAX: 003715; RRID:IMSR_JAX:003715
Mouse: Ai3: B6.Cg-Gt(ROSA)26Sor ^{tm3(CAG-EYFP)Hze/J}	The Jackson Laboratory	JAX: 007903; RRID:IMSR_JAX:007903
Mouse: OMP-IRES-Cre Omp ^{tm1(cre)Jae}	Eggan et al. ¹⁰⁴	MGI: 3045687
Mouse: OMP-IRES-tTA B6;129-Omp ^{tm1(tTA)Gogo/J}	Yu et al. ¹⁰⁵	JAX: 017754; RRID:IMSR_JAX: 017754
Mouse: tetO-P2-GFP Or10a4 ^{tm1.1Besk}	Fleischmann et al. ⁵⁶	MGI:5520984
Mouse: Cnga2 Cnga2 ^{tm1Reed}	Zhao and Reed ⁵⁷	MGI: 3525590
Mouse: Cbx1(Cbx5) cKI	Escamilla-del-Arenal et al. ⁷²	N/A
Mouse: Foxg1-Cre B6.129P2(Cg)-Foxg1 ^{tm1(cre)Skm/J}	Hebert et al. ¹⁰⁶	JAX: 006084; RRID: IMSR_JAX:006084
Mouse: S50 ^(M72 locus) B6;129P2-Or8a1b ^{tm11(Or55b10)Mom} /MomJ	Bozza et al. ²²	JAX: 006714; RRID: IMSR_JAX:006714
Mouse: M72 ^(S50 locus) B6;129P2-Or55b10 ^{tm3(Or8a1b)Mom} /MomJ	Bozza et al. ²²	JAX: 006715; RRID: IMSR_JAX:006715
Oligonucleotides		
Custom 300-gene panel	Vizgen	Cat# 10400002
Lentiviral barcode primer: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACCGAACGCAA CGCACGCA-3'	This paper	N/A
Primers for Hybridization Chain Reaction (HCR)	This paper	Table S4
HCR Amplifier B1-546	Molecular Instruments	N/A
HCR Amplifier B3-647	Molecular Instruments	N/A
Software and algorithms		
Python (v3.8–3.12)	Python Software Foundation	https://www.python.org/
GEP usages code	Tsukahara et al. ⁵¹	GitHub: https://github.com/dattalab/Tsukahara_Brann_OSN ; DOI: 10.5281/zenodo.5550454
DV score analysis code	This paper	GitHub: https://github.com/dattalab/Brann_olfactory_dorsoventral ; DOI: 10.5281/zenodo.18941308
Cell Ranger (v6.1.2 & v9.0.1)	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/downloads
SAW	Gong et al. ¹⁰⁷	https://stereotoolss-organization.gitbook.io/stomics-software-download/saw-previous-versions
Scanpy	Wolf et al. ¹⁰⁸	https://scanpy.readthedocs.io/en/stable/
scvi-tools	Gayoso et al. ¹⁰⁹	https://docs.scvi-tools.org/en/stable/index.html
cNMF	Kotliar et al. ⁵²	https://github.com/dylkot/cNMF
Samtools	Li et al. ¹¹⁰	https://www.htslib.org/
Nextflow	Di Tommaso et al. ¹¹¹	https://nextflow.io/docs/latest/index.html
nf-core/chipseq (v2.0.0)	Ewels et al. ¹¹²	https://nf-co.re/chipseq/2.0.0/
Bowtie2	Langmead and Salzberg ¹¹³	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
bwa	Li ¹¹⁴	https://github.com/lh3/bwa
pairtools	Abdennur et al. ¹¹⁵	https://pairtools.readthedocs.io/en/latest/
Fiji	Schindelin et al. ¹¹⁶	https://imagej.net/software/fiji/
QuPath	Bankhead et al. ¹¹⁷	qupath.readthedocs.io

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Chromium Next GEM Chip G Single Cell Kit	10x Genomics	Cat# 1000120
Chromium GEM-X Single Cell 3' Chip Kit v4	10x Genomics	Cat# 1000690
MERSCOPE Slide Box	Vizgen	Cat# 10500001
Stereo-seq Chip T Slide	STOmics	Cat# 210CT13114
SPRIselect Beads	Beckman Coulter	Cat# B23318

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

C57BL/6J mice were obtained from Jackson Laboratory (Stock No. 000664). Mice expressing OMP-IRES-tTA and tetO-P2-IRES-GFP (OMP-P2 mice) were obtained from the Lomvardas lab and maintained in the Datta Lab.^{56,105} Cnga2 knockout animals were obtained from the Dulac lab and maintained by breeding heterozygous females with C57BL/6J males.⁵⁷ HP-1 swap mice (mice with loxP-HP1 β -STOP-loxP-HP1 α alleles at the HP1 β locus crossed to Foxg1-Cre [Jackson Laboratory Stock No. 006084]¹⁰⁶ such that knockout of HP1 β is rescued via the expression of HP1 α) were generated by Martín Escamilla-Del-Arenal using standard homologous recombination, as recently described.⁷² Wild-derived CAST/EiJ (Stock No. 000928), MOLF/EiJ (Stock No. 000550), and PWK/PhJ (Stock No. 003715) mice were obtained from Jackson Laboratory, and CAST/C57 F₁ hybrids were generated in lab by crossing CAST males with C57 females. As described, we found little effect of sex on the DV score (Figure S1N), and thus mice of either sex between 6–16 weeks old were used for experiments. Mouse husbandry and experiments were performed following institutional and federal guidelines and were approved by Harvard Medical School's Institutional Animal Care and Use Committee.

METHOD DETAILS

OR activity manipulations

Data from wild-type mice housed in home-cage environments, or from mice that either underwent transient naris occlusion for a week or were housed in novel olfactory environments, were all previously generated.⁵¹ Act-seq was performed as previously described, by exposing mice to monomolecular odorants on filter paper for two hours prior to OSN dissociation and scRNA-seq.⁵¹ Odorants were diluted in dipropylene glycol (DPG) based on their vapor pressures to give a nominal concentration of 500 ppm at their highest concentrations.

Lentiviral and retinoic acid experiments

The main olfactory epithelium (MOE) was ablated and induced to regenerate via a single intraperitoneal injection of methimazole (50 mg/kg) in adult mice (>6 weeks old), as described elsewhere.¹¹⁸ Intranasal lentiviral injections were performed 36–48 hours after methimazole injection, at timepoints in which the OSNs have sloughed off and the epithelium is a single monolayer of activated horizontal basal stem cells (HBCs). Mice were anaesthetized with Ketamine/Xylazine (100/10 mg/kg), and intranasal lentiviral injections were performed via a Hamilton syringe connected to thin PE-10 tubing; 15–25 μ L of virus (at a titer of 1–10 $\times 10^8$ TU/mL) was injected over the course of 2–3 minutes. scRNA-seq was performed 4–6 weeks after methimazole injection, at timepoints under which much of the epithelium has regenerated and the infected progenitors have differentiated to give rise to clones of newly-born cells, many of which are OSNs.

Manipulations of retinoic acid signaling were performed in adult mice that received a single dose of methimazole (50 mg/kg). Starting 72 hours after methimazole injection, mice received either the ALDH1A2 inhibitor WIN 18,446 (5 mg/kg), all trans-retinoic acid (atRA, 5 mg/kg), or the RA receptor inverse agonist BMS493 (10 mg/kg). WIN 18,446 was dissolved in DMSO, atRA was dissolved in 5 % DMSO (dissolved in corn oil), BMS493 was dissolved in 5 % DMSO (dissolved in corn oil), and drugs were delivered subcutaneously. Littermates within each cohort received either drug or vehicle daily for 3–4 weeks after methimazole-induced regeneration, and scRNA-seq was performed after 4–5 weeks; animals receiving each drug or vehicle control were dissociated and sequenced or processed for in situ in parallel. The effects of the RA inhibitors WIN 18,446 and BMS493 were verified via observation of the testes,¹¹⁹ which were typically 30–50% the weight of littermate controls.

Single-cell RNA-seq and data preprocessing

Isolation of single cells from the MOE and fluorescence-activated cell sorting (FACS) was performed as described.⁵¹ In brief, mice were euthanized via carbon dioxide inhalation, and the epithelial tissue of the MOE was dissected. Individual mice were used for each replicate. Cells were dissociated with Papain and DNase-I (Worthington), incubated at 37°C for 60 minutes, triturated and then washed and resuspended in Hibernate-A medium supplemented with fetal bovine serum; these solutions also contained

transcriptional and translation inhibitors (5 mg/mL of Actinomycin D, 10 mg/mL of Anisomycin and 10 mM of Triptolide, all obtained from Sigma). FACS was performed to enrich for OSNs expressing fluorescent reporters in the OMP-P2 and lentiviral experiments. In all other experiments, FACS was performed to enrich for live singlets.

scRNA-seq was performed using the 10x genomics platform using the default protocols (CG000315 Rev E) for the Chromium Next GEM Single Cell 3' Kit v3.1 and the default protocols (CG000731 Rev B) for the Chromium GEM-X Single Cell 3' Reagent Kits v4. For each replicate, cells were loaded at a concentration predicted to maximize yield (10,000 cells for v3.1 kits and 20,000 for v4 kits), or to obtain the maximum number of cells given the measured concentrations when the total number of cells was less than 10,000. For the OMP-P2 experiments, GFP-positive and GFP-negative cells were mixed and loaded together in a single lane of the Next GEM chip. In the lentiviral experiments, Venus-positive cells and negative cells from each mouse were not mixed and were loaded into separate lanes. For experiments with genetic or experimental controls (like the CNGA2, RA, and HP1 swap experiments), control and experiment animals were processed together, loaded onto the same chips, and were sequenced together.

The resulting libraries were quantified via qPCR and via the Qubit and were sequenced with paired-end sequencing on the Illumina NovaSeq or NextSeq platforms (with Read1 = 28 cycle, Index (i7) = 10 cycle, Index (i5) = 10 cycle, Read2 = remaining number of cycles). Demultiplexed fastq files were processed in a manner similar to that previously described,⁵¹ using a custom Nextflow pipeline that ran the Cell Ranger (version 6.1.2 for v3.1 and v9.0.1 with `--include-introns=false` for v4) count pipeline, as well as additional post-processing steps to identify and remove multi-mapped reads and regenerate the resulting cell by gene matrix. The Ensembl GRCm39 mouse genomic index (version 105) was used with a custom GTF file containing extended 3' UTR annotations for some ORs and TAARs, to facilitate their identification in the 3'-biased scRNA-seq data. Cell doublets and low-quality cells were excluded, and OSNs were identified via iterative subclustering of the latent spaces of scVI models, as previously described,^{51,120} based on their expression of known marker genes and ORs.

Clonal lineage tracing

Clonal lineage tracing was performed using a lentiviral library with ~10M DNA barcodes at the 3' UTR side of transcripts for Venus fluorescent protein and upstream of the polyA sequence (Collecta CloneTracker XP). To amplify the lentiviral barcodes from the RNA transcripts in each cell, a primer (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACCACCGAACGCAACGC ACGCA-3') whose 5' end overlapped with the Truseq Read 2 (underlined) and whose 3' end was complementary to a constant region upstream of the barcode was spiked in (final concentration 0.5 μ M) during the initial RT and cDNA amplification steps; this approach increased the yield of the lentiviral barcodes compared to targeted amplification from the already-amplified cDNA libraries. Amplified barcode fragments were purified from the cDNA library and barcode sequencing libraries were constructed by performing PCR on these fragments with 10x Library Index PCR primers (which bind to Truseq); the resulting barcode libraries were mixed and sequenced together with the final 10x libraries and were subsequently demultiplexed computationally.

Spatial profiling via MERFISH

MERFISH was performed using the Vizgen MERSCOPE platform. A custom 300-plex codebook, which contained nearly 200 ORs and 100 non-OR genes consisting of cell type marker genes, DV-related genes, and other highly-variable OSN genes, was designed. Bulk sequencing and scRNA-seq was used to identify suitable ORs and genes, and the total abundance of the gene panel was ~7600 FPKM to avoid optical crowding. ORs were chosen based on their DV scores and frequency of expression to obtain a set of ORs predicted to uniformly span the MOE at relatively-higher frequencies. Due to their high abundance, additional probes for OSNs (*Omp*, *Calm2*), and sustentacular cells (*Cbr2*, *Cyp2g1*) were imaged separately and sequentially (along with the staining for DAPI, polyT, and cell boundaries).

Fresh frozen 10 μ m coronal samples from the MOE from young (4-5 week) mice were processed according to the manufacturer's user guides (Vizgen 91600002 Rev E and 91600112 Rev C). In brief, tissue sectioning, fixation, permeabilization, and autofluorescence quenching were performed using the guidelines for fresh frozen tissue. However, the RNA anchoring steps from the formalin-fixed, paraffin-embedded (FFPE) tissue sample preparation guide were added to help the thin MOE tissue adhere to the MERSCOPE slides. Cell boundary staining, gel embedding, clearing, and probe hybridization steps were performed without modification from the fresh-frozen sample preparation guide. Samples were imaged on the MERSCOPE instrument using the 300-plex imaging kit, and transcripts were decoded using the MERLIN pipeline provided by Vizgen.

Spatial transcriptomics via Stereo-seq

Stereo-seq was performed using the STOmics Stereo-seq Transcriptomics T v1.3 kits following the manufacturer's user guide (STUM-TT001). In brief, 10 μ m coronal sections from fresh frozen MOE samples (from young 4-5 week mice) were sectioned onto 1 mm² Stereo-seq Chip T Slides, dried at 37°C for 5 minutes and methanol-fixed for 30 minutes at -20°C. The cells were stained with Qubit ssDNA reagent, and the slide was then rapidly imaged on an Olympus VS200 slide scanner with a 10x objective. Following imaging, slides were permeabilized at 37°C for 12-15 minutes, and the captured RNA was reverse transcribed for 2 hours at 45°C. The resulting cDNA was released at 55°C for 10 minutes and then amplified via PCR and size-selected with SPRIselect beads. 100ng cDNA was used for each sample and sequencing libraries were prepared according to the manufacturer's user guide (STUM-LP002),

using an equimolar mix of 4 index barcodes per sample. Two samples were pooled per sequencing run, and the resulting libraries were sequenced on DNBSEQ-T7 flowcells in the PE75 (50+100) configuration; demultiplexed fastq files were used for downstream analyses.

Chromatin conformation via Hi-C

The Hi-C variant Micro-C, which uses MNase instead of restriction enzymes to digest cross-linked DNA,¹²¹ was used to survey the chromatin conformation of mature olfactory sensory neurons. Micro-C was performed using the Dovetail Genomics Micro-C Kit (PN 20101E) following the instructions in its user guide (v2.1 for cells). In brief, ~1 million EYFP⁺ OSNs were dissociated and FACS-purified from OMP-IRES-Cre;Ai3(RCL-EYFP) animals for each sample, following the approach described above in the scRNA-seq experiments. Sorted cells were pelleted, frozen, and then thawed and cross-linked with DSG and 37% formaldehyde. Next, cells were digested with MNase for 15 minutes at 22°C. The digested lysate was proximity ligated, and DNA was purified. To increase library complexity, three technical replicates were prepared for each sample, using 150 ng of purified DNA per replicate. Fragments with ligation junctions were enriched with streptavidin beads, and index PCR and size selection were performed. Libraries were pooled and sequenced with NovaSeq X Plus 25B flowcells, targeting ~1B reads per sample. The resulting fastq files from each technical replicate were processed in parallel and were pooled for downstream analyses.

CUT&RUN

CUT&RUN was performed to measure the levels of H3K9me3 in C57/CAST F₁ animals, using the reagents and instructions from a commercial kit (EpiCypher CUTANA CUT&RUN Kit, Version 4). In brief, the MOE was dissociated as in the scRNA-seq experiments, and 500k unsorted cells were used for each reaction, and two technical replicates were used for the cells from each animal. Cells were incubated overnight with ConA beads and 0.5 μg of a Rabbit polyclonal H3K9me3 antibody (Abcam ab8898), and on the following day permeabilized cells were incubated with pAG-MNase to target the digestion of DNA, which was purified with SPRIselect beads. *E. coli* spike-in DNA was used to normalize the yields across samples. Sequencing libraries were prepared from 5 ng CUT&RUN DNA, or the entire yield for replicates where the yield was <5 ng. Some experiments also included a H3K4me3 positive control (EpiCypher 13-0041), which yielded the expected enrichment for transcriptional start sites, IgG negative control (EpiCypher 13-0042) antibodies, which showed minimal enrichment, but these were omitted for subsequent experiments given the low background in the H3K9me3 CUT&RUN data. Sequencing libraries were generated using the NEBNext DNA library kit, following the instructions and formulations in the EpiCypher CUT&RUN library prep kit (14-1002). CUT&RUN libraries were pooled and sequenced on the Element AVITI machines with the Cloudbreak Freestyle sequencing kit (2x75 cycles).

In situ hybridizations

In situ hybridization for genes in the MOE was performed via the Hybridization Chain Reaction (HCR) method,¹²² using 16–20 μm cryosections from either fresh or fixed frozen MOE tissue, from 4 to 8-week-old animals. Paraformaldehyde-fixed tissue was post-fixed overnight, decalcified in 0.45M EDTA overnight, and equilibrated in 30% sucrose prior to freezing. In situs were performed according to manufacturer instructions for HCR v3.0 (Molecular Instruments, Rev. 4), except the proteinase K solution was changed to 1 μg/mL or omitted entirely and probe concentration was increased to 1–2 pmol per 100 μL.

Split DNA oligo probes for ORs (~15–20 probes per gene) were manually designed using regions in each OR's cDNA sequence that were homologous between C57BL/6J and CAST/EiJ animals but lacked homology to other ORs. Probes for the dorsal marker *Nqo1* and reference ORs whose positions remained constant across strains were also designed in a similar manner and used as reference landmarks to identify the positions of ORs of interest. For in situs in animals treated with the RAR inhibitor, a pooled OR probe was designed to target the P2 OR (Olfr17) along with eight additional ORs with similar DV scores (Olfr76, Olfr393, Olfr713, Olfr873, Olfr1302, Olfr1341, Olfr1370, and Olfr1490) whose frequency also increased in the RA-inhibited animals. The pooled OR probe was ordered as an oPool oligo pool from IDT (at the 50 pM scale) and was resuspended to a concentration of 1 μM per oligo and used at a concentration of 1 pmol per 100 μL. The resulting probe mix labeled OSNs in the same region of the epithelium as P2 alone, consistent with their similar DV scores. The sequences of all oligo probes can be found in [Table S4](#).

To further validate the in situs, antibody staining was also performed for *Nqo1* in a subset of animals. After the HCR protocol, sections were post-fixed with 4% PFA for 15 minutes, permeabilized and blocked for 30 minutes in PBS with 0.3% Triton X-100 (PBST) and 5% Normal Donkey Serum, and then incubated overnight with Goat anti-*Nqo1* (1:100) in blocking buffer. The following day, sections were washed with 0.1% PBST and then incubated for 2 hours at room temperature with Donkey anti-Goat Alexa Fluor 488 secondary antibodies.

Sections were counterstained with DAPI and confocal imaging was performed using a Nikon Ti inverted spinning disk microscope equipped with a Yokogawa CSU-W1 Spinning disk (50 μm pinhole), a Plan Apo 20x/0.8 NA air objective and a Plan Fluor 40x/1.3 NA oil objective, and 405, 561, and 640 nm laser lines. Tiled images of the MOE were captured using an Andor Zyla 4.2 Plus sCMOS monochrome camera and Nikon Elements Acquisition Software (AR 5.02). For sections probed with individual ORs, the OSNs expressing each OR were manually counted and the fraction of OSNs within the dorsal zone was assessed by comparing to positions of each OSN relative to the *Nqo1*⁺ region of dorsal OSNs. Note, although imaging and quantification was performed across the entire section, images depict OSNs close to the edge of the *Nqo1*⁺ region due to the difficulty in visualizing the expression of single ORs across the entire MOE. Due to their increased density and to avoid experimenter biases, in the RAR inhibitor experiments ORs were

automatically identified and counted using the cell detection algorithms in QuPath.¹¹⁷ To compare across samples, endoturbinate III (using the nomenclature of Barrios et al.¹²³) was chosen because it is readily identifiable, and the number of cells and their distance along an annotated path that spanned the turbinate from its tip were quantified.

Data integration of scRNA-seq replicates

Nearly 5 million cells from over 360 replicates were uniformly processed using a custom Nextflow pipeline to run Cell Ranger and custom postprocessing steps for each replicate. These replicates included the 150 replicates (and 780 thousand mature OSNs) described in a previously published dataset,⁵¹ as well as over 200 additional replicates (containing another 1.4 million OSNs). The additional data consisted of the mice used as part of the experimental conditions described herein (e.g. the lentiviral experiments), as well as additional unpublished datasets that were generated for other work, including from mice that had been exposed to mono-molecular odorants, as well as from smaller numbers of mice housed in different environmental conditions, and control and knock-outs for various genetic manipulations. Mice of both sexes were used for experiments. Note, although some of these lines had not been backcrossed to C57BL/6J backgrounds, few differences in DV scores or gene expression were observed across replicates, even though the entire dataset consisted of 360 replicates from many separate experiments.

A combination of unsupervised and semi-supervised methods was used to map the cells from all replicates into a common latent space to perform cell type identification. The gene set used for this embedding were the top 3000 genes (excluding ORs, mitochondrial and ribosomal genes) identified using the Scanpy highly variable gene function, with the “seurat_v3” flavor. First, unsupervised learning was used to fit a scVI model (n_hidden : 128, n_latent : 30, n_layers : 2, $gene_likelihood$: “nb”, $dropout_rate$: 0.2, with a batch key to identify each replicate) on the raw counts of these genes in a smaller reference dataset of ~250 thousand cells from wild-type animals housed in standard home-cage conditions. The resulting scVI latent embeddings from this dataset were clustered via the Leiden algorithm and cluster labels were manually annotated based on known cell type markers. The trained scVI model and the manually-annotated cluster labels from this subset were then used to train a semi-supervised model using scANVI. The trained scANVI model was then applied to query the entire dataset, resulting in a common embedding of all cells into a 30-dimensional latent space. Cell types in this common latent space were identified by transferring the labels from the manually-annotated reference subset, using previous weighted nearest neighbor approaches for label transfer of scANVI models.^{124,125} For cells from each replicate, their resulting neighbors (in the 30-dimensional latent scANVI space) from the annotated subset were found using PyNNDescent, distances were converted into affinities using a gaussian kernel, and cell type probabilities and matching uncertainty were calculated for each cell.

Non-neuronal cells, as well as low-quality doublets and dying cells were excluded, and mature OSNs were extracted from the cell type labels of the integrated dataset. Due to the dissociation and FACS approaches used, roughly 50% of cells from most replicates were OSNs. OSNs were filtered to retain only cells expressing a single OR, and the resulting dataset of 2.3 million OSNs was used for downstream analyses. Except where noted, only cells expressing single ORs were used for analyses, and OR expression was defined with a threshold of at least 3 transcripts (unique molecular identifiers (UMIs)) that mapped to that OR.

Identification of zonally-restricted genes

Zonal indices for each OR were obtained from prior work,^{15,16} and were discretized into five zones to match prior conventions, with zone 1 being the most-dorsal zone and zone 5 the most-ventral. The mean for each subtype for each variable gene was calculated and genes whose expression was on average ($\log_2$ fold-change > 1) higher among the subtypes in its maximum versus second-highest zone were considered as zonally-restricted genes, and the number of genes for each zone were computed. Note, no genes were enriched in intermediate zone ORs (e.g. zone 2–4), consistent with gradients of gene expression that monotonically increase or decrease with the dorsoventral score. To account for the possibility that genes might be enriched in alternate regions (like 2.5–3.5, or 2.0–2.5, or 2.1–2.6), bin widths of 0.5 and 1.0 were tested, and were slid across the entire zonal range (at 0.1 intervals) to quantify the number of genes with $\log_2$ fold-changes > 1 for each bin location. Such analyses also failed to identify any zonal genes for intermediate zones. Since no genes were identified, no correction for multiple comparisons was performed.

Identification of the dorsoventral (DV) score

Consensus non-negative Matrix Factorization (cNMF) was used to decompose the transcriptomes of each OSN into a smaller set of interpretable factors.^{51,52} The same NMF factors (referred to as gene expression programs (GEPs)) that were previously identified were used, and the gene loadings for these GEPs were applied to all datasets.⁵¹ Importantly, cNMF was performed on a set of ~1,300 highly variable genes (HVGs) that did not include any OR genes and thus the resulting gene loadings and GEP usages do not depend on the expressed ORs. These GEPs included two associated with ligand-dependent activity, (GEP_{High} and GEP_{Low}), two containing dorsal and ventral genes (GEP_{Dorsal} and GEP_{Ventral}), and two relating to anteroposterior positions (GEP_{Anterior} and GEP_{Posterior}). Therefore, even though we did not explicitly attempt to identify GEPs that might reflect spatial position, the fact that this approach yielded spatially-varying GEPs indicated that epithelial positions might be reflected in OSN transcriptomes. These GEPs were used in an orthogonal manner and their differences in usage was summarized into single metrics: the environmental state (ES) score = GEP_{High} - GEP_{Low}, the dorsoventral DV score = GEP_{Dorsal} - GEP_{Ventral}, and the anteroposterior (AP) score = GEP_{Anterior} - GEP_{Posterior} to capture the continuous variation across these three independent axes (ES, DV, and AP). On the other hand, cNMF also identified GEPs that distinguish OSNs that either express CD36, are located in the “patch” region of the MOE, or that have high levels

of genes related to the unfolded protein response. The usage of these GEPs was less graded, suggesting that the orthogonal nature of $GEP_{Dorsal} - GEP_{Ventral}$ is not a necessary consequence of our cNMF approach but rather reflects the continuous transcriptional variation with respect to spatial position (which we validate with subsequent experiments).

The 248 genes (out of the set of highly variable genes) whose expression across OSNs was correlated with the DV scores ($R^2 > 0.5$ for linear regression to predict the DV score from each gene and spearman's $\rho > 0.5$ at the OSN subtype level) were considered as DV genes and were used for downstream analyses. DV genes were categorized into 10 different functional groups (transcription factors, axon guidance and cell adhesion, cellular metabolism, cell signaling, cell growth, cytoskeletal, GTP-activity, RNA-related, ion channels and transporters, and synaptic genes) via a combination of GO pathway analysis, the presence of protein features like DNA binding domains, or via manual annotation. The full list of DV genes and their associated categories (for the 162 belonging to the above categories) can be found in [Table S1](#).

As described above, GEP usages and DV scores were calculated at the cell level using the gene loadings for each GEP. A small fraction (3%) of cells had DV scores that were zero, which can occur for cells that had little expression of either dorsal- or ventral-related genes and because NMF identifies usages for all GEPs at once. For a subset of analyses that required comparisons at the cell level, the DV scores of these cells were imputed using a multilayer perceptron (MLP) trained to predict the DV score of mature OSNs using the expression of the 248 DV genes in each cell. The MLP (with two hidden layers of size 50 and 10) was trained in PyTorch with the Adam optimizer (learning rate 0.01) and gave predictions that were highly correlated ($\rho > 0.99$) at both the cell and OR level for training, testing, and validation data.

Mature OSNs expressing a given OR (i.e. those from the same OSN subtype) had similar DV scores. Therefore, the DV score associated with OSNs expressing each OR (also referred to as an OR DV score) was calculated by taking the mean DV score across all the OSNs expressing a given OR in the entire integrated dataset. Except where noted, for all downstream analyses the DV score of the expressed OR in each cell was considered as the OR DV scores from the entire dataset and thus could be compared to each cell DV's score, especially in experiments that decoupled a given cell's DV score from that of its expressed OR. In some experiments, DV scores were percentile normalized at the cell level via sklearn's QuantileTransformer. This transformer was fit on data from all cells to maintain the distribution of DV scores for cells from each subtype.

OR classes

Only ORs with functional protein activity (as assessed by their singular expression in mature OSNs) were used for analyses. ORs with mean DV scores above 60 were considered as dorsal; this threshold was set such that ORs that had previously been classified as being part of the most-dorsal zone (zone 1) were considered as dorsal with respect to their DV scores. ORs were clustered based on their amino acid sequences and class I ORs were identified as the large cluster of ORs in the resulting phylogenetic tree that share homology with fish ORs and were located in a single genomic cluster on Chromosome 7. Phylogenetic distances were defined as the cophenetic distance of the resulting phylogenetic tree. Most rodent odorant receptors are class II ORs, which are located in multiple genomic clusters of genes that are scattered across the genome. ORs were considered part of the same genomic cluster if they were located on contiguous regions of the chromosome (separated by less than 3 Mb from the start of the previous OR gene). Neighboring ORs (i.e. those at neighboring genomic locations) were more likely to be phylogenetically similar. A small fraction of cells also express trace amine-associated receptors (TAARs). Except where specified, analyses used OSNs expressing either class I, class II, or TAARs though most results reflect properties of class II ORs, which make up the majority (~87%) of all ORs. The set of 1007 OSN subtypes with at least 150 cells (median 1,500 OSNs per subtype) was used for most analyses.

Predicting DV scores and DV genes

The DV genes all had strong positive/negative correlations with the DV score and with each other, suggesting they act as a coherent program. To verify these results, and to ensure that the DV scores weren't being driven by small numbers of genes, a series of validation approaches were performed. First, every single DV gene was individually removed and GEP usages were recalculated (by reweighting the gene loadings for each GEP to sum to 1 and applying the GEPs via `non_negative_factorization` with `update_H=False`). Second, to avoid any confounds in relating the DV score to the expression of known marker genes, the marker genes *Nqo1*, *Nfia*, *Nfib*, *Nfix*, *Ncam2*, *Nrp2*, *Robo2*, and *Acsn4* were all removed and GEP usages were similarly refit. In both cases, the DV scores for each OSN subtype were perfectly correlated ($\rho = 0.999$) with those calculated using all genes, suggesting that individual genes have redundant roles in this larger program. Alternatively, the sufficiency of small numbers of genes at predicting the DV score was tested using regression models that predicted the DV score for each OSN subtype via the mean gene expression for that subtype. Elastic-net-regularized linear regression models were used (with `l1_ratio=0.95`) with cross-validation (training on 80% of OSN subtypes and testing on the held-out 20%). Random sets of 1–248 either random HVGs or DV genes were used across 1000 restarts, and the distribution of prediction accuracies (`r2_score`) for each type of gene set was summarized. These models were also compared to those that attempted to identify the best set of n DV genes for values of n from 1–248 (i.e. using the best DV gene or all DV genes). To do so, the best set of the DV genes in the training data on each restart for each value of n was identified using recursive feature elimination (RFE) with the same Elastic-net model.

Regression analysis was also used to evaluate how well the expression of each DV gene in each OSN subtype could be predicted by either the DV GEPs or a zonal model. Model comparison was performed between six regression models. First, gene expression across OSN subtypes was fit using the default zone indices for the canonical 5-zone model (with zone 1, 1–2, 2–3, 3–4, and 4–5 as categorical features). Five alternative models were evaluated: 1) the zones, but offset by half (≤ 1.5 , 1.5–2.5, 2.5–3.5, 3.5–4.5, > 4.5); 2) a piecewise-linear model (`PiecewiseLinFit` in the `pwlf` library) fit using the DV score with 4 segments; 3) a `LinearRegression` model

that used the usage of the two DV GEPs for each OSN subtype; 4) a KNN regression model (with uniform weights and 20 neighbors) fit on DV GEP usage in the training data; or 5) a LinearRegression model in which the two DV GEPs were expanded into a larger quadratic basis using the PolynomialFeatures(degree=2, include_bias=False) transformation. All models were fit via five-fold cross-validation, and the change in prediction accuracy (Δ r₂ score) for each model for each gene was compared to that from the default zonal model.

Consistency of the DV score

Correlations of the DV scores across replicates were computed at the OSN subtype level, by comparing the mean for each subtype in a given replicate to that from all other data. To evaluate the consistency across data subsets, 150 cells were sampled for each OSN subtype, the mean for each OSN subtype was calculated separately for the first versus second half of cells for each subtype, the spearman's correlation between data halves was evaluated, and this process was repeated 1,000 times. The Spearman's correlation reflects changes in the rank order of OSN subtypes and the absolute change in ranks was also evaluated on each restart, as well as for data in which OR labels were either permuted across cells whose ORs shared the same zonal index (discretized into 5 zones as above) or permuted across all cells. The mean for each OSN subtype was also evaluated separately for data from male and female animals. In an alternative approach, the correlation in DV scores was computed for all pairs ($_{363}C_2$) of mice using, for each pair, the set of ORs with at least 4 cells in both samples. Even with only 4 cells per sample, the median spearman's correlation was still 0.975.

The uniqueness of the DV score for each OSN subtype was determined using linear regression. Across 100 restarts, cross-validation was performed using five stratified folds for training and testing. The mean for each OSN subtype was computed using the training data and in the testing data the observed DV score for each OSN was compared to that of either the mean of its respective OSN subtype in the training data or the mean of the n^{th} nearest OSN subtypes (as determined based on the means of each OSN subtype in the training data); the difference in median absolute errors was computed for models with n ranging from 0–200.

For computing deviations from the mean and area under the receiver operator curve (AUROC), equal numbers of OSNs (150) were sampled for each OSN subtype on each of 1,000 restarts, and AUROC values were calculated empirically for all pairs of OSN subtypes ($_{1007}C_2 = 506,521$ pairs). For comparisons to shuffled data, OSN labels were shuffled across the subsampled cells on each restart. The difference in DV scores for each cell from the respective mean of its corresponding OSN subtype was also summarized via the median absolute deviation (MAD). AUROC values were also evaluated for pairs of ORs found within the same zone, using either the default 5-zone model described above (and plotted as a function of their difference in DV scores), or the “Zolf” zone annotations described in a recent paper.¹²

MERFISH analyses

Cell type identification in MERSCOPE data

Data from different epithelial sections were aligned to a common template using a two-step approach. First, since the sections were from similar regions of the epithelium, they were all roughly aligned to a reference template with a simple rigid transformation. Second, these sections were then aligned STalign,¹²⁶ an approach that relies on diffeomorphic mapping (to help account for slight deformations in the thin epithelial layer). Overlaying the raw images and marker gene expression following alignment confirmed this approach successfully aligned the various tissue layers. MERFISH gene expression was also visualized using Baysor,¹²⁷ in its segmentation-free configuration. In brief, local neighborhoods were evaluated by finding the nearest transcripts (across all genes) for every gene, reducing the dimensionality of this data via principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP),¹²⁸ and converting the resulting low dimensional embedding into colors for visualization purposes. Because OSNs are small and densely packed and difficult to segment via DAPI or cell boundary staining alone, cell segmentation from the Vizgen MERlin pipeline (via cellpose) was further refined using Proseg,¹²⁹ a probabilistic approach that helps to better match transcripts to cells. The resulting log-normalized cell by gene matrix from Proseg was processed in Scanpy. PCA was performed and the top 30 PCs were kept. Leiden clustering was performed on a random subset of 100,000 cells (on a nearest neighbor graph with $n=30$ neighbors and leiden resolution=1.2) and the labels were transferred to all cells using the weighted nearest neighbor approach described above for the scRNA-seq data. Clusters were annotated and merged based on their expression of known marker genes (e.g. for OSNs, immature precursors, and non-neuronal genes); the spatial distribution of these genes and their respective clusters were also evaluated, and matched prior knowledge. As indicated, downstream analyses that used a specific cell type (e.g. INPs or mesenchymal lamina propria cells) used only the cells and/or transcripts found within those cells of a given cluster. Similar spatial segregation was also observed at the transcript level when the Baysor local neighbors for each transcript were clustered using KMeans clustering (with $k=3$ –10 clusters).

OSN subtype identification

OSNs expressing single ORs were identified in an iterative process, using the spatial locations of OR transcripts. First, only transcripts in the OSN layer (identified via the clustering described above) were considered. Next, to avoid false positives from background expression (or errors in spot decoding) OR transcripts were filtered for each gene to only keep those with at least 3 other transcripts within a 10 μm radius (the average size of an OSN), and the resulting transcripts were then clustered into individual cells using the DBSCAN algorithm (eps=5 and min_samples=3). The set of OSNs was further refined by removing 0.1% of cells in which the median pairwise distance between OR transcripts within the cell was greater than 10 μm . OSNs were identified separately for each

sample, using the STaligned coordinates. Four OR genes out of ~200 showed nonspecific binding across the MOE (and were also decoded in damaged areas and areas with high autofluorescence) and were not considered for downstream analyses. Only ORs detected in at least 20 different OSNs were used for analyses.

DV_{ISH} scores and DV scores in MERFISH data

The DV_{ISH} score was computed directly using the MERFISH-measured DV gene expression for each cell. To infer DV_{ISH} scores in OSNs, the total expression of each gene for each cell (all transcripts within a 10 μ m radius of the cell centroid) were total-count normalized. Next, rather than applying cNMF directly (because cNMF was fit on scRNA-seq data and the MERFISH panel only contained a subset of all HVGs), PCA was performed on the total-count normalized expression of the DV genes that were part of the MERFISH panel and the first PC was taken to be the DV_{ISH} scores for each cell. DV_{ISH} scores smoothly varied across the MOE and matched the underlying gradients of individual DV genes. DV_{ISH} scores were calculated in a similar manner for INPs, using the set of INP cells that were located basally to OSNs that did not have detectable OR transcripts and the counts of the nearest 100 transcripts to each cell. The DV score for each cell was taken to be the DV score associated with its expressed OR, as measured via scRNA-seq.

Predictions of the DV score associated with the OR expressed in each cell were made in a cross-validated manner by holding out all the cells of a given OSN subtype at a time using either the DV_{ISH} score (linear regression with the top PC of DV expression, where PCA was fit on the training data) or the DV score associated with the ORs expressed in neighboring cells (with KNeighborsRegressor with $n_neighbors=10$ weighted by distance). Predictions were summarized at the OSN subtype level. Similar predictions were also observed using the nearest neighbor regression approach on a smaller sagittal dataset.

Alignment of MERFISH epithelial segments

The complex 3D geometry of the nose made it difficult to compare data across sections, or across turbinates directly. An alignment procedure was developed to map data into common coordinate systems. Subsections of the epithelium, like the septum, had previously been shown to contain ORs from multiple zones. Using the expression of known dorsal and ventral ORs, the boundaries of all segments of the epithelium that spanned from dorsal to ventral in each turbinate were manually annotated using the coordinates of the aligned sections post-alignment with STalign. Then, for each segment, the ranked distance of each cell from the most dorsal vertex of the polygon boundary of the segment was computed. These positions, expressed in terms of the fractional length along each segment, were considered as the “unaligned” coordinate system for downstream analyses.

If ORs are arranged continuously across the epithelium, then their positional ranks (along the dorsal to ventral axis of each segment) should be a good proxy for distance, assuming OR ranks are preserved across segments on average even if such ranks change with different rates across segments (e.g. due to differences in the total length of segments or their location within the MOE). The mean rank for each OR across all segments was calculated and used to develop a mapping between position (i.e. the fractional length along each segment) and the mean OR rank of the OR expressed at each position and these curves were fit with piecewise linear functions (with at most 5 knots and constrained to increase monotonically) via the *pwlfit* library. The resulting piecewise linear regression thus gives a mapping, for each segment, between the positional coordinates of that segment and the common mean OR ranks; these predicted values were taken to be the new coordinates for each cell, thereby mapping each segment into a common coordinate space. Given the monotonicity constraint of this approach, the rank ordering of cells was relatively preserved by this approach. Accordingly, the overall mean OR ranks changed little by this approach, but the resulting positions in the new coordinate system were less confounded by differences in absolute positions across segments and better captured the underlying latent “position” of each cell, and in the aligned dataset the mean OR rank at a given position was more correlated. While this annotation and alignment procedure was done independently of the DV scores, both the mean unaligned and aligned positions were well-correlated with the DV score.

Analysis for MERFISH common coordinates

The resulting unaligned or aligned positions for the cells from each segment were used for downstream analyses. Common coordinate positions for each OR were averaged across all detected sections, and these mean positions strongly correlated ($\rho = 0.969$) with the DV score. Cross-validated regression models were used to quantitatively compare the predictability of position via the DV GEPs to that predicted by the zonal model. Three formulations of a five-zone model (i.e. linear regression with categorical features) were used: 1) with the standard zone indices (zones 1, 1–2, 2–3, 3–4, and 4–5); 2) the zones, but offset by half (≤ 1.5 , 1.5–2.5, 2.5–3.5, 3.5–4.5, >4.5); or 3) the optimal bins from the training data, found using *sklearn*’s *DecisionTreeRegressor* with *max_leaf_nodes=5*. These models were compared to 1) a *LinearRegression* model that used the usage of the two DV GEPs and DV score for each OSN subtype; 2) via a *KNN* regression model (with uniform weights and 10 neighbors) fit on the same features in the training data, or 3) a *LinearRegression* model in which the two DV GEPs were expanded into a larger quadratic basis using the *PolynomialFeatures* (*degree=2*, *include_bias=False*) transformation. All models were fit via five-fold cross-validation, and the prediction accuracy (*r2_score*) for each model was summarized across 500 restarts.

AUROC analysis was performed between pairs of OSN subtypes, as described above for the scRNA-seq data. For analyses that looked at pairwise distances between cells, only pairs of cells within the same segment were considered, and pairs were accumulated and summarized across all segments. Dorsal ORs had broader and more intermingled spatial positions, which became compressed during our alignment procedure (since the mean ranks in the dorsal part of the epithelium were often flat). To avoid any confounding results, subsequent analyses were performed, where indicated, using only cells expressing ventral class II ORs.

Regression was also used to predict the position of each cell in the aligned coordinate system using each cell's DV_{ISH} score. Predictions were performed using a Histogram-based Gradient Boosting Regression Tree (HistGradientBoostingRegressor with loss="absolute_error", max_leaf_nodes=15, max_depth=15). Equal numbers of cells were subsampled for each OSN subtype (for the set of OSN subtypes with at least 20 cells). The regressor was trained with five-fold cross-validation across 250 restarts and the median absolute error in the predicted positions on each restart was computed. The observed accuracy was compared to models trained on data in which DV_{ISH} scores were shuffled across cells within an OSN subtype or data in which the positions of each cell were shuffled across all cells.

Stereo-seq analyses

Stereo-seq data were processed using the standard SAW pipeline,¹⁰⁷ which maps reads to the Ensembl v105 GRCm39 reference and decodes the spatial position of each barcode using the fiducial markings and tissue masks from the fluorescent ssDNA image. Downstream analyses were performed on the resulting gene expression file. OSNs were identified in data that was binned at 20 μ m, using OR-bin pairs found within the epithelial layers that were detected in at least 3 UMIs. To assess how well DV scores mapped onto space, a similar spatial k-nearest neighbor approach was used, as described above for the MERFISH data. In brief, the DV score for each cell was predicted using the DV score associated with the ORs expressed in neighboring cells (the 10 nearest spatial neighbors, weighted uniformly by distance), using cross-validated KNN models that held out all cells for a given OR (for the 839 ORs detected in at least 6 cells). Predictions were either averaged across all cells from all chips for each OSN subtype or averaged separated for each chip to calculate the correspondence in predictions across animals; predictions were also compared to those made using the MERFISH data, using the set of OSN subtypes identified with both technologies. For quantifying the AP genes via Stereo-seq, the expression of each gene associated with $GEP_{Anterior}$ and $GEP_{Posterior}$ in a 15 μ m radius around each OSN was evaluated. OSNs were grouped into 10 deciles based on their associated AP scores, and the z-scored expression for anterior and posterior-associated genes was averaged for each decile. For visualization purposes, different sections from the same chip were also aligned to each other using affine transformations and STalign, as described above for the MERFISH data. In these aligned coordinates, the DV scores associated with each OR were then averaged in 50 μ m bins.

DV scores versus other spatial metrics

Bulk RNA-seq

DV scores were compared to two previous measurements obtained via bulk RNA-sequencing of the spatial positions of ORs within the epithelium. One approach measured OR RNA levels in micro-dissected subregions of the epithelium,¹⁶ whereas the other performed RNA tomography and measured OR RNA levels in cryosections across all three axes of the epithelium.¹⁷ These spatial indices were compared to the mean DV score for OSNs expressing each OR in wild-type data (as measured via GEP usages, which are calculated using non-OR genes) for the sets of ORs detected both in these datasets and via scRNA-seq.

DV scores were also predicted based on the expression of ORs in manually-dissected "zonal" subregions of the MOE from a third dataset.²⁸ The raw fastq files for these experiments were obtained from the NIH sequence read archive (SRA, dataset SRP285789). Bulk RNA-seq experiments were uniformly processed and mapped to the same mouse index used for scRNA-seq experiments (Ensembl GRCm39 v105) using a Nextflow pipeline (nf-core/rnaseq) that ran STAR and RSEM to quantify the bulk RNA abundance of genes in each sample.^{112,130,131} The DV score for each OSN subtype was predicted as a function of OR abundance in the zonal dissections using support vector machine regressors (C=500 and kernel="rbf"), trained via 5-fold cross validation across 1,000 restarts.

External MOE MERFISH dataset

MERFISH data from the MOE were generated in the Dulac lab and were publicly released (CC BY 4.0) as part of the NIH's Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative - Cell Census Network (BICCN).⁵⁴ Preprocessed data, which indicated the decoded olfactory receptor gene detected in each of ~ 3.3 million cells, were downloaded from the Brain Imaging Library (BIL, dataset ID: ace-bag-tin).¹³² This MERFISH dataset consisted of data from 10 mice in which coronal sections were probed with two OR gene panels that each expressed ~ 500 different ORs (and $\sim 1,100$ ORs in total). Analyses were performed on 49 sections from 8 mice, excluding 2 mice whose sections had lower quality. The epithelial sections were manually segmented as above, and the resulting 704 sections were also aligned to common coordinate system using the global OR ranks and the ranking of the cells in each segment. While the resulting ranks were well-correlated with the DV score overall, some ORs had appeared to be out of place compared to their neighbors, suggesting they might be the result of false-positive detections in this dataset. Lastly, to predict the DV scores of each OR in a segmentation-free manner, the same nearest-neighbor approach described above for the other MERFISH dataset was performed in which all the OSNs of a given subtype were held out and the DV scores associated with the held-out OR was predicted using the DV scores associated with the ORs expressed in the nearest neighbors spatially. Predictions were summarized at the OSN subtype level.

Comparison of AP scores with MERFISH data

AP scores were also evaluated at the OSN subtype level in both MERFISH datasets using the AP scores associated with each OR in the integrated scRNA-seq dataset. To relate AP scores to epithelial positions, the apical and basal axes of each epithelial segment were manually annotated and the difference in the distance to the apical and basal axis (i.e. the side of the epithelium closer to the lumen/sustentacular cell layer and the side closer to the mesenchymal and basal stem cell layer, respectively) was computed for each

OSN. Only OSNs whose distances to both axes were within the bottom 99th percentile were included, to exclude outliers (likely false-positives outside of the epithelium or damaged areas of the tissue). The mean apical-basal distance for each OSN subtype was computed, which were then compared to the associated DV and AP scores for each subtype. The correlation in AP scores and apical-basal distances was also evaluated separately for dorsal and ventral ORs. The apical-basal positions of a subset of ORs were also manually verified via in situ hybridizations.

DV scores in activity manipulations

Naris occlusion and environment swaps

Data from wild-type mice housed in home-cage environments, or from mice that either underwent transient naris occlusion for a week or were housed in novel olfactory environments, were all previously generated.⁵¹ DV scores were evaluated separately for each subtype for each condition (e.g. across open vs. closed nostrils or for mice housed in each environment), for OSN subtypes found in both conditions. In mice housed in home-cage environments, the DV score for each subtype was also compared to each subtype's ES score, which summarizes the chronic activity induced by the odorants in the home-cage.

CNGA2 mice

scRNA-seq was performed on female mice heterozygous for the CNGA2 allele. Because CNGA2 is X-linked its expression is mosaic in female heterozygotes because individual cells inactivate either the wild-type or loss of function allele. Cells with the loss of function allele were identified either based on SNPs in the *Cnga2* cDNA that distinguished it from the C57 allele, or via reads that mapped to lacZ. *Cnga2* loss of function cells had decreased ES scores, consistent with the usage of that GEP being downstream of OR-driven activity.

Act-seq

Identification of activated OSN subtype in the Act-seq experiments was performed as previously described.⁵¹ In brief, activated OSN subtypes were identified via immediate early gene induction, and the activation score for each subtype in the odor condition was calculated using a previous set of acutely-responsive "activation genes." Activation scores were consistent across replicates. The DV scores for each activated OSN subtype were calculated using the entire set of cells in our integrated dataset.

DV scores in mice with OR manipulations

OMP-P2 mice

P2 cells from OMP-P2 mice were identified via the expression of the P2 receptor, which was expressed in a large fraction of cells across the entire epithelium of mature mice. Two-OR cells were defined as those in which both ORs were expressed with at least 3 UMIs; many of these cells contained P2 in OMP-P2 mice, and P2 was often expressed at higher levels than the other non-P2 OR. DV scores in P2-expressing cells in the OMP-P2 mice were compared to those in our integrated dataset expressing P2 or to those expressing the non-P2 OR (for two-OR cells).

DV scores in "OR Swap" mice

Cells expressing each OR in the OR-swap mice were identified based on the 3' UTR associated with each OR transcript, which was unchanged in these mice and therefore identified the genomic locus of the expressed OR for each cell. The crossed design of the OR swap experiment (in which cells express either S50 or M72 receptors from each of the S50 and M72 genomic locus) facilitated the identification of genes that followed the locus or the expressed OR. Classification was performed using cross-validated SVM models (with kernel="rbf") trained on the expression of either the ~1,300 HVGs or the ~250 DV genes in each cell (with a feature selection step that kept the top 100 genes from the training data). Models were trained to predict the genomic locus of each cell using a subset of cells from each mouse and were tested on held-out cells.

Gene and OR expression in differentiating OSNs

Differentiating OSNs were also identified based on the transferred cell type labels of the scANVI latent space. Both mature and immature OSNs were excluded to obtain a dataset of differentiating cells in the process of choosing ORs. Iterative subclustering was performed on the latent space of new scVI models (*n_hidden*: 128, *n_latent*: 10, *n_layers*: 1, *gene_likelihood*: nb, *dropout_rate*: 0.1, with a batch key for each replicate and the number of genes as a continuous covariate) trained on the top variable genes from this dataset. In this process, any remaining doublets, low-quality cells, additional non-neuronal cells that were not excluded by the above filtering steps, and the small subset of cells from the non-canonical *Emx1*⁺ lineage that gives rise to *Gucy1b2*⁺ and *Gucy2d*⁺ sensory neurons were removed and the scVI models were retrained until the dataset consisted of only differentiating cells. The resulting differentiation dataset contained ~400k cells and densely captured cells from the earliest globose basal stem cells to late INP/early immature OSN cells that express single ORs. Note, although differentiating cells are labeled as INP to match prior literature,²⁸ cell cycle analysis suggests that many INPs are likely postmitotic. To remove the effects of the cell cycle on downstream analyses, cells scores for the G2/M and S phases were computed (scanpy's *score_genes_cell_cycle*) using published marker gene sets.¹³³ Any gene whose expression was correlated with either G2/M or S score (pearson's *R* > 0.2) was removed and the remaining 2773 were used to train a new scVI model with the same training parameters as listed above. Cells were clustered based on the nearest neighbor graph of the resulting 10-dimensional scVI latent space and this latent space was also used to reduce the dimensionality of the data to two dimensions for visualization purposes via UMAP (using a KNN-graph with 100 neighbors); these UMAPs are shown in the figures of the differentiating dataset.

Pseudotime analyses

Smooth differentiation trajectories were visible in the resulting UMAP projections, and thus an approach was designed to pseudo-temporally order the cells. An approximate KNN graph was constructed based on the cosine similarity in the scVI latent space, using $k=20$ neighbors. A vector the size of the number of cells in the dataset was constructed. This vector was all zeros except for a single GBC cell, which was given a starting weight of 1. This vector was then multiplied by the KNN graph (weighting each cell by 0.1 and the influence of its nearest 20 neighbors by 0.9), renormalized from 0 to 1, and this procedure was repeated 1500 times until the values at each iteration remained stable. The resulting vector was rank-ordered, and these ranks were considered the pseudotime values for each cell; similar pseudotime values were also observed starting from other cells. A similar KNN-smoothing approach was used to smooth gene expression and GEP usages across neighboring cells, following procedures described previously.¹³⁴ Here the starting vector was the gene expression or GEP usages for each cell and the smoothing procedure was only run for 5 iterations, which results in locally-weighted values for each cell. The smoothed expression of example genes are shown in the UMAP representation, and the locally-smoothed version of each cell's DV score was used for all downstream analyses, to help counteract the fact that fewer DV genes were expressed pre-choice and thus estimates at the cell level were noisier. Gene expression as a function of pseudotime was also fit via Generalized Additive Models using pyGAM. The alignment of DV score with axes of gene expression was assessed for cells of a given pseudotime bin (width = 0.1 and steps of 0.05) by taking the cosine distance of the DV score and the top principal components, where PCA was performed using the log-normalized expression of all OSN highly-variable genes in each pseudotime bin. DV gene expression as a function of pseudotime was also visualized for each gene, using pseudotime bins of width 0.05. Gene expression was z-scored across bins and then smoothed and sorted for visualization purposes using a rolling mean of 3 bins.

OR expression as a function of pseudotime

OR expression was considered with different thresholds. To identify multi-OR cells, a threshold of 1 UMI was used. Such a threshold likely induces some false positives (from e.g. background “soup” expression), given that both non-neuronal cells and GBCs sometimes expressed ORs at this 1 UMI threshold. Therefore, analyses of OR expression were restricted to cells expressing class II ORs with pseudotime values of at least 0.4, the value at which OR expression was consistently observed across cells. At subsequent pseudotime values, cells transiently expressed multiple ORs at the 1 UMI threshold. Cells with “competing” ORs were defined as those expressing multiple ORs at 3+ UMIs each (the threshold used for singular expression in mature OSNs). For cells expressing multiple ORs, the mean DV score of the co-expressed ORs was computed by weighting the DV scores associated with each OR (as measured in mature OSNs) by their expression levels (so an OR expressed at 2 UMIs would be weighted twice as much as one with 1 UMI). For individual ORs the median (and other percentiles) of the pseudotime values of all OSNs expressing that OR was computed, both for OSNs expressing each OR at any level, as well as for those expressing at either low (1–2 UMIs) or moderate levels (3+ UMIs).

To assess the correlation between each cell's DV score and that associated with the OR, the correlation was evaluated both for cells at different differentiation stages as well as for OR expressed in those cells at either the highest, second-highest, or lowest (1-UMI) levels. Across 1,000 restarts, single ORs among the ORs expressed at each level in each cell (e.g. one of the 1-UMI ORs was picked for each cell) and the correlation was evaluated on each restart separately for cells at each differentiation stage (pre-choice with lots of low expression, cells with “competing” ORs, and cells expressing a single OR at high levels and others at lower levels).

Clonal lineage tracing

Identification of clonally-related cells in lentiviral data

To identify cells expressing barcodes and to extract the lentiviral barcode, raw sequencing reads that contained the constant start of the barcode (i.e. those that came from the amplified barcode libraries) were extracted from the BAM files. Only barcode reads with at least 10 reads for a given UMI were evaluated. The barcode library contained two variable regions (that were 14 and 30 bp long, respectively). Barcodes with hamming distances smaller than that observed between the whitelist of barcodes (3 for the 14 bp region and 5 for the 30 bp region) were collapsed and barcodes that were part of the whitelist were included in downstream analyses. Cells that expressed a single barcode (e.g. one barcode accounted for all of the barcode reads/UMIs for that cell) were considered for downstream analyses, and cells from the same replicate that expressed the same lentiviral barcode were considered as clonally-related. The resulting list of barcode-expressing cells from each experiment did not have any overlapping barcodes across mice, suggesting that the lentiviral libraries were sufficiently diverse to uniquely label each clone in each mouse. However, the overall UMIs per barcode were relatively low, and, even with the targeted amplification, many of the cells in the Venus⁺ libraries did not have any detectable barcodes. This might reflect false positives from FACS, or transcriptional silencing of barcode loci, especially given the extensive heterochromatinization of OSNs; the fluorescent signal from many infected cells was also quite weak, which is also consistent with silencing of the integrated cassette. Therefore, the modal number of cells per barcode was one, and the mode across clones with at least two cells was two. Nevertheless, results held true across clone sizes, further indicating that “small” clones are likely undersampled (due to the e.g. difficulties in barcode detection or in capturing all cells during the single-cell dissociated), but not fundamentally different than those with more cells. Additionally, the OSNs identified following methimazole-induced regeneration expressed single ORs at normal levels and had GEP usages that was similar to that of control OSNs that expressed the same OR. Similar results were observed for the lentiviral-infected cells, which expressed a wide array of ORs.

Clonal restrictions

Clones expressing at least two OSNs, with each singularly expressing any OR or only class II ORs, as indicated, were considered for downstream analyses. DV scores for the expressed OR were calculated by using the DV score associated with that OR in the entire integrated dataset. The clonal mean OR DV was calculated by averaging the DV scores across clonally-related OSNs. Differences between the DV scores of clonally related ORs were calculated between all pairs of cells from the same clone and were either summarized across clones or across all pairs. Shuffles for all analyses were performed by shuffling clonal labels across cells. Clonal data was simulated based on the observed clonal restriction in DV scores by sampling, for each of the cells in the observed clones, cells from the wild-type data whose ORs had DV scores that matched the observed *t*-distribution of deltas from the clonal mean (e.g. ORs with DV scores similar to the clonal mean DV score had the highest probability of being chosen); the probabilities were also scaled by the observed frequency of each OR in the wild-type data. The percent of clones that had multiple cells that expressed the same OR (and therefore fewer number of ORs than cells) were then analyzed for both the observed and simulated data.

DV scores were also evaluated at the cell level, in which each cell's DV scores was obtained by taking the difference in GEP_{Dorsal} and GEP_{Ventral} usage irrespective of the expressed OR. To test whether restrictions at the cell level were stronger than that at the OR level, the MAD in cell DV scores within a clone was compared to the MAD obtained when DV scores were shuffled across cells of a given OSN subtype (irrespective of their clonal identity) or shuffled across all cells.

INP gene expression was related to OR DV scores, using the subset of clones that contained both INP cells and OSNs. For the cells in these clones support vector machine regression models (SVR with C=50 and kernel="rbf") were fit on the INP gene expression (where the input to the linear models was the top 250 genes in the training data, reduced to 30 dimensions via PCA) to predict the mean OR DV score of the clonally-related ORs expressed in the sister cells. The regression was performed via cross-validation, leaving out all the cells from a clone; additionally, the top genes and PCA transformation was identified using only the cells of the training data to avoid any data leakage.

Retinoic acid manipulations

Data from adult mice given either the ALDH1A2 inhibitor WIN 18,446 (RA inh.) or all trans-Retinoic Acid (atRA) during methimazole-induced regeneration were analyzed relative to their respective vehicle controls. Cell types were identified via clustering the latent space of scVI models and cNMF GEP loadings were applied to each cell, as described above for the integrated dataset. Analyses were either performed at the cell level, using the observed DV score for each cell, or at the OR level, using the DV score associated with each OR as measured using all cells in the integrated datasets. To assess distributional shifts, the median DV score across all cells or the number of cells expressing ORs with given associated DV scores were compared between the data from each condition and its respective control; given that fewer cells express the most ventral ORs to begin with, the changes in OSN frequency were also expressed as log₂ fold-changes. Because RA manipulations changed the overall distributions of cells without altering the relationship between cell identity and OR choice, distributional shifts were observed when evaluating all OSN subtypes at the cell level, but minimal changes were observed across conditions for cells of a given OSN subtype. Changes in gene expression were evaluated by z-scoring the expression of each DV gene across all cells and the mean change in z-scored expression was evaluated for cells from drug versus control conditions. Analyses were performed on mature OSNs expressing single ORs, except where, as noted, INPs were used to evaluate changes in DV scores and OR expression in differentiating cells pre-choice.

ChIP-seq data analyses

ChIP-seq data were previously generated and were reanalyzed.^{28,65} The raw fastq files for ChIP-seq data for H3K9me3 and H3K79me3 marks were obtained from the SRA (datasets SRP285789 and SRP096660) and were reprocessed with the Ensembl GRCm39 version 105 mm39 genome, using the default parameters of the nf-core/chipseq Nextflow pipeline (v2.0.0).¹¹² As part of this pipeline, reads were mapped to the genome with BWA, and the normalized read density was summarized at base-pair resolution with UCSC-bedGraphToBigWig. The density of reads across the entire coding and non-coding regions of each OR for each sample were averaged, and the reads for each OR locus (for class II ORs) were correlated with the DV score of OSNs singularly expressing that same OR. The samples evaluated in these analyses were of sorted GBCs, INPs, iOSNs, or mOSNs that were sorted from the entire MOE (and thus contain a mix of cells from the dorsal and ventral regions). The average heterochromatin from these bulk dissections was also compared to those found in Olfr1507+ cells (sorted from Olfr1507-IRES-GFP animals).

Hi-C data analyses

Micro-C data was analyzed following standard approaches. In brief, paired end fastq files were aligned to the Ensembl v105 GRCm39 genome using bwa (bwa mem -5SP -T0). Valid ligation junction events were identified using pairtools parse (–min-mapq=40, –walks-policy=5unique, and –max-inter-align-gap=30), and then sorted (pairtools sort), deduplicated (pairtools dedup), and split (pairtools split) into pairs and pairsam files. Contacts in the pairs files were summarized at a 1kb resolution with cooler and converted into multi-resolution mcool files. To best analyze interchromosomal contacts for individual ORs, data from all biological and technical replicates were combined for downstream analyses, but results were consistent across samples. Using data binned at the 1kb resolution, the ± 5kb region from the start of each class II OR gene was considered as an OR locus, and the number of contacts between that locus and all other OR loci were summarized. To avoid distance-dependent effects, downstream analyses were restricted to

interchromosomal (trans) contacts. Class II ORs were grouped into deciles based on their associated DV scores, and the mean pairwise (10 x 10 deciles) or one to all (10 deciles) contact frequency were evaluated.

Similar results were also observed via Hi-C from OMP-IRES-GFP mice, and from data from individual OR lines. Hi-C experiments were processed with the Nextflow nf-core/hic pipeline (with digestion== “dpnii”), and were aligned to the same reference genome using bowtie2. The resulting multi-resolution mcool files were then evaluated as described above.

HP1 swap experiments

scRNA-seq was performed as described above in the HP1 swap and control animals (i.e. with or without the Foxg1-Cre allele) in young (4–6 week) animals of either sex. Cell type identification, DV scores, and changes to the cell and DV scores in the HP1 swap animals were evaluated in a similar manner to those in the RA manipulations. To assess changes in OR choice for cells of a given DV scores, cells were binned based on their cell DV scores and the percent-normalized DV score associated with each OR was computed for cells in each bin. The cell vs OR DV score mapping was evaluated using robust linear regression (HuberRegressor), and equal numbers of cells were sampled for each OSN subtypes, for subtypes detected in both control and HP1 swap animals. To evaluate changes in the cell to OR DV score mapping, a nearest neighbor approach was used, and a nearest neighbor regression model (KNeighborsRegressor) was fit on all cells from the control mice to predict the DV score associated with the ORs expressed in cells of a given DV score. This model was then applied to the cell DV scores of the HP1 mice, and the residuals from this model, which capture the difference in the associated percentile-normalized DV score of the OR detected in each HP1 swap cell with that predicted based on each cell’s DV score, were evaluated and summarized for HP1 swap cells of a given DV score.

Allelic expression in C57/CAST F₁ hybrids

scRNA-seq was performed in F₁ hybrid animals generated by crossing wild-derived CAST/EiJ female mice with C57BL/6J males. CAST/EiJ mice have on average single nucleotide polymorphisms (SNPs) every 150 bp. Because OR expression is monogenic and monoallelic, the reads that mapped to the chosen OR were used in each mature OSN to infer the strain of the chosen OR. The GRCm39 coordinates of SNPs that distinguished CAST/EiJ and C57BL/6J were obtained from the Mouse Genome Project,^{76,135} and SNPs in ORs of interest were also confirmed by evaluating scRNA-seq data from homozygous CAST/EiJ animals. For each cell, any reads (with MAPQ > 30) that overlapped with homozygous SNPs in CAST/EiJ mice for its chosen OR were evaluated to assess what fraction of reads expressed the CAST/EiJ or C57BL/6J variant. Cells that had reads that overlapped at least one SNP and in which at least 80% of such reads came from a single allele (mean 99.7% of reads) were used for downstream analyses. Similar results were also obtained using Demuxalot to demultiplex the strain of each cell via the set of SNPs across all OR genes.¹³⁶ The strain of the chosen OR was able to be inferred in 81% of cells. The remaining 19% of cells had too few reads that overlapped SNPs, likely because they expressed ORs with minimal variation between strains in their 3’ UTRs. SNPs that led to missense mutations were annotated via SnpEff.¹³⁷

OSN subtypes with significant changes in GEP usages were evaluated empirically using permutation testing, and subtypes with a mean change between strains larger than 1% of shuffles for that GEP were considered as significant. Similar changes were observed across F₀ and F₁ animals, as well as across individual samples; however, some OSN subtypes had few cells for either CAST/C57 (as OR frequency varies across strains) and were likely underpowered.

CUT&RUN in F₁ hybrid animals

Demultiplexed fastq files from CUT&RUN samples were mapped to the Ensembl GRCm39 v105 genome via Bowtie2 (–I 10 –X 700 –no-mixed –end-to-end –no-discordant –very-sensitive).¹¹³ Spike-in reads from *E. coli* DNA were mapped to the *E. coli* genome. However, because F₁ animals have within-animal controls (the reads mapped to each allele), the number of reads that mapped to the *E. coli* genome were not used for normalization. Duplicates were not removed, and data were combined across replicates. As in the F₁ scRNA-seq data, the allele of each read was inferred for reads overlapping SNPs that distinguished CAST/EiJ and C57BL/6J. The number of reads that mapped to each allele were then summarized for given ORs across the entire genomic coordinates for that gene (i.e. from TSS to TES).

OR promoter analyses

Promoter motif enrichment was performed using HOMER,¹³⁸ using the 1kb region upstream of the TSS (i.e. the promoter region) for each OR (findMotifsGenome.pl with –size=given). This approach confirmed the enrichment of known OR promoter motifs (i.e. EBF, the homeobox TF LHX, and ATF); the de novo motif results also identified additional variants of EBF and LHX motifs that were enriched in the OR promoters. EBF motifs were especially enriched in the 0–250 bp upstream of the TSS, as expected for promoter motifs. Next, differential motif enrichment between the top 200 dorsal and ventral class II OR promoters was evaluated using HOMER and MEME, but this approach merely returned the presence of EBF and LHX motifs, suggesting that it may have been data limited. Finally, supervised promoter annotation was performed using HOMER (annotatePeaks.pl) using the default motif files and threshold provided by HOMER for LHX2, EBF, ATF (given their general promoter enrichment identified above) and for NF1, and RARa (given their presence in the DV GEPs). This approach revealed dorsal OR promoters were enriched for EBF binding motifs, and NF1/RAR motifs were more likely to be found in ventral ORs. Comprehensive scanning for motifs for additional TFs and factors (including Mef2a, Emx2, Sox2, CTCF, AP-1, and other common promoter motifs, all via their default motif files provided by HOMER) failed to identify other TFs with dorsoventral biases, suggesting that the factors above may be sufficient, or such differences

may not be captured by the differential presence of putative binding motifs. The odds ratio for the co-occurrence of EBF, NFI, and RAR motifs was evaluated empirically via bootstrapping.

Olfactory bulb glomerular positions

The olfactory bulb glomerular map was generated via spatial transcriptomics using the 10x Visium platform, as recently described by Klimpert and colleagues.⁸⁰ In brief, putative glomeruli were identified using a Bayesian approach that used the expression of individual OR genes across the entire 3D volume to infer the location and number of glomeruli for each OR. The outputs of this model were manually curated to keep high-confidence identified glomeruli, and an axis of symmetry was identified that separated glomeruli from the medial and lateral domains. Importantly, the two hemibulbs are also symmetric, and once aligned to each other sister glomeruli expressing the same OR were found on average 230 μm away from each other ($\sim 1\text{--}2$ glomerular lengths), reflecting a combination of both biological variability and spatial errors in the glomerular detection and alignment procedures; such deviations are also similar to those from previous estimates.³⁴

Glomeruli were identified based on OR gene expression and could thus be linked to the scRNA-seq or MERFISH data for the associated OSN subtype expressing that OR. First, OB positions were used to predict the scRNA-seq measured GEP scores for each OSN subtypes, via a cross-validated support vector machine regression model (SVR with $C=150$ and “rbf” kernel). Predicting GEP usages from glomerular positions is easier than the reverse, due to the higher dimensionality of the 3D positions and due to the fact that the DV score mapped onto an axis in the OB that was not fully aligned with the cartesian D-V axis but rather correlated with both the D-V and A-P position of glomeruli. Similar results were also obtained through canonical correlation analysis, which identified optimal rotations to align the GEP and glomerular spaces, further demonstrating that the DV score was well-aligned ($\rho > 0.9$) with the top canonical correlation axis. Second, predictions of the 3D glomerular position for each OSN subtype was performed separately for glomeruli in each domain. Elastic-net-regularized linear regression models ($\alpha=1$, $l1_ratio=0.9$) were trained on either GEP usage (e.g. $\text{GEP}_{\text{Dorsal}}$, $\text{GEP}_{\text{Ventral}}$, and the DV score) or all $\sim 1,300$ variable genes to predict the glomerular position along each axis and the 3D error of these predictions was evaluated for each glomerulus. For both models, five-fold cross validation was performed; this procedure was repeated 100 times, and the distribution of the accuracies and or errors across restart was reported.

QUANTIFICATION AND STATISTICAL ANALYSIS

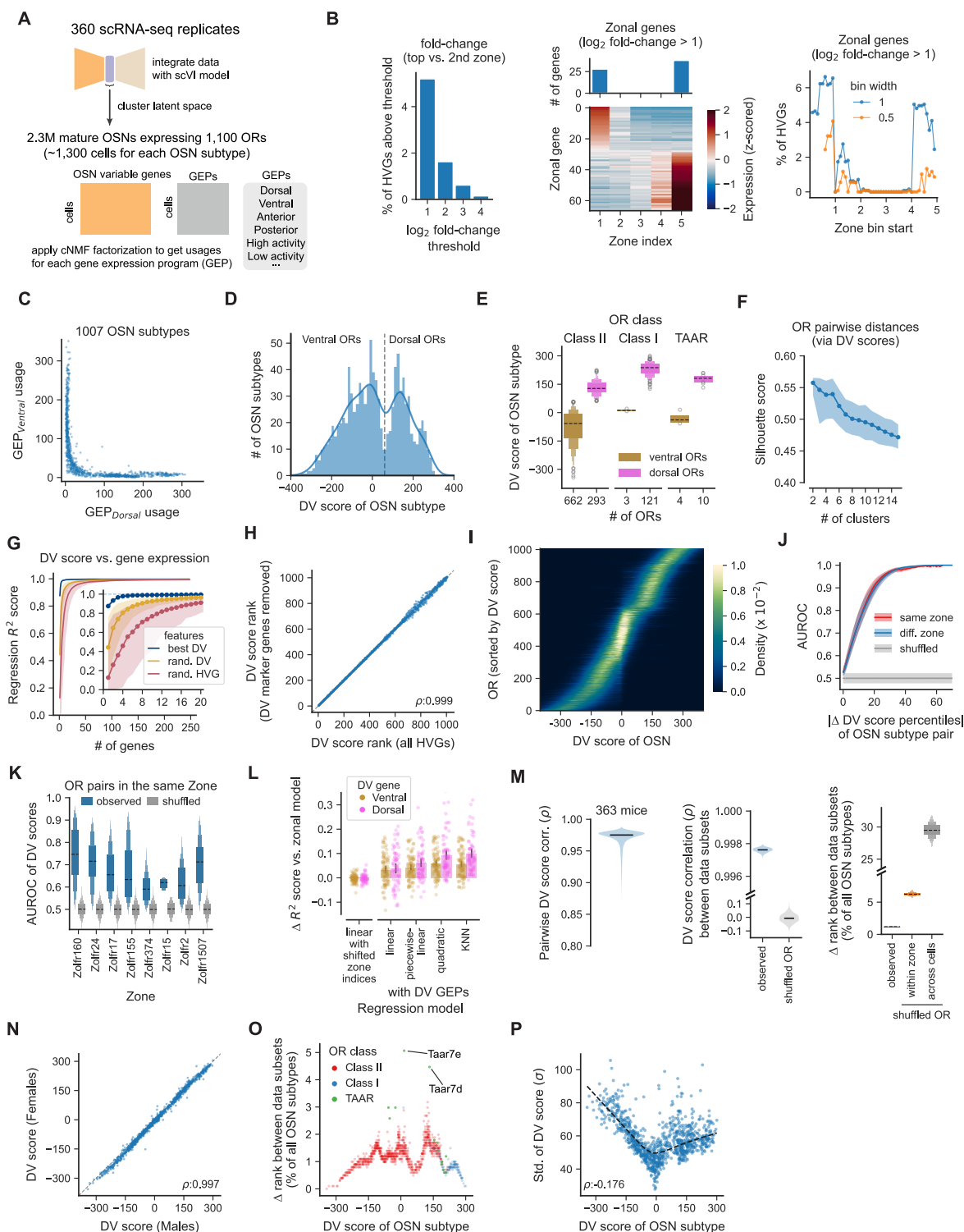
Software used for data analysis

Single-cell analyses were performed in python (versions 3.6–3.12) using the Scanpy and scvi-tools packages,^{108,109} as well as custom-written scripts using the open-source python scientific stack (pysam, SciPy, NumPy, scikit-learn, umap-learn, pandas, statsmodels, matplotlib, seaborn, and numba). cNMF was performed using modified versions of the code from <https://github.com/dylkot/cNMF>.

Statistical testing

Hypothesis testing was performed using parametric or non-parametric statistical tests, as indicated, or p-values were calculated empirically using resampling-based permutation tests. The precision of sample statistics and regression trend lines were evaluated using bootstrapping, and, except where noted, plots and error bars depict the mean and the 95% confidence intervals of the mean across 1,000–10,000 bootstraps. Throughout the paper, a non-parametric version of the box plot (also known as a letter-value plot) was used to represent multiple quantiles and tails of large distributions of data (e.g. to summarize across OSN subtypes or sets of ORs or OR pairs) in an agnostic way that does not require setting bandwidth parameters as in violin plots or kernel density estimates. Like a conventional box plot, the largest box represents the interquartile range (25–75 percentile) and the median (dotted-line). Subsequent boxes recursively represent exponentially-smaller quantiles (the 12.5–25 and 75–87.5 percentiles, then 6.25–12.5 and 87.5–93.75 percentiles, then 3.125–6.25 and 93.75–96.875 percentiles, and so forth. The results of additional statistical tests are listed in [Table S5](#).

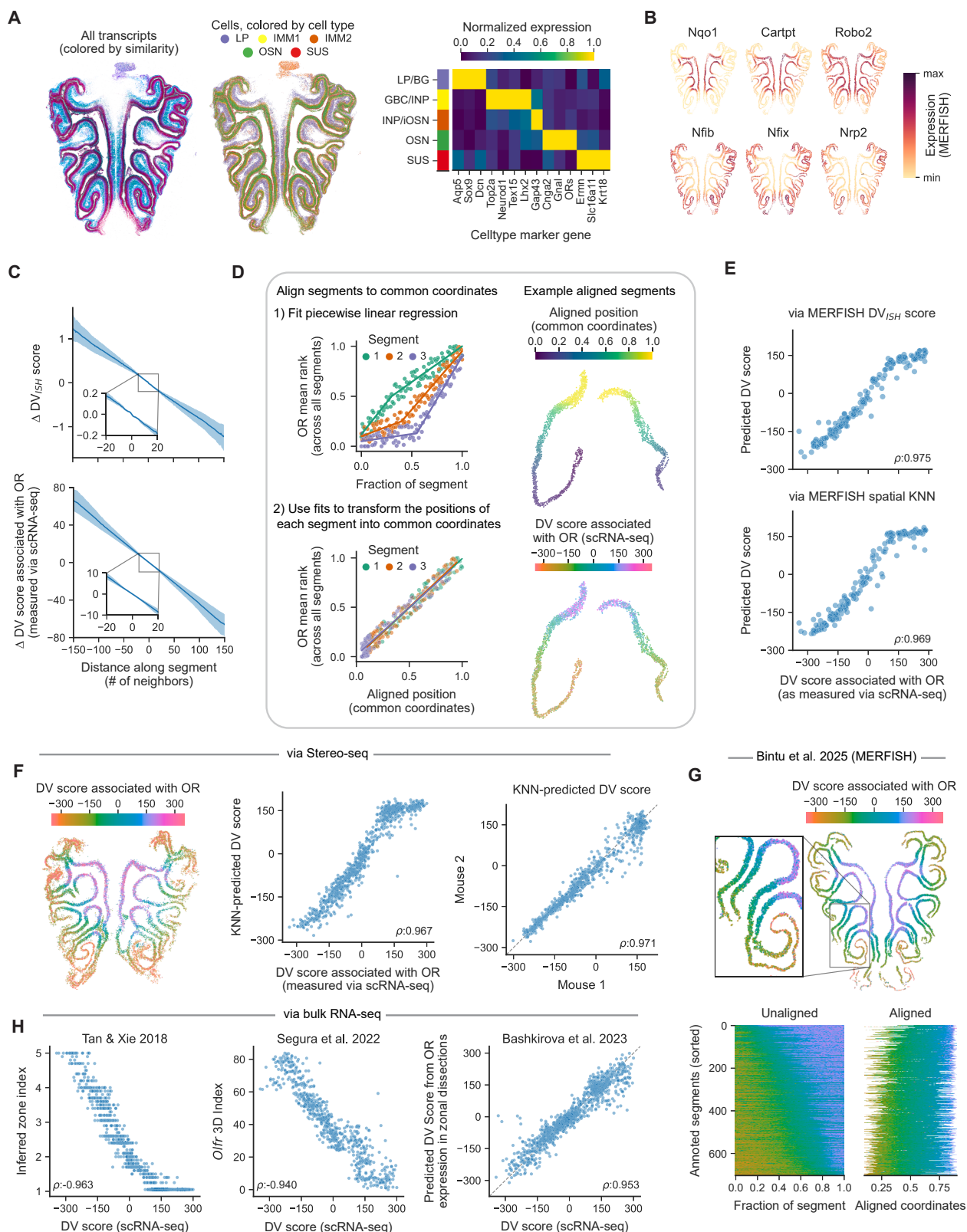
Supplemental figures



(legend on next page)

Figure S1. DV scores vary across OSN subtypes in a stereotyped manner, related to Figure 1

- (A) Schematic depicting the approach for integrating single-cell RNA sequencing (scRNA-seq) data from hundreds of replicates. Mature OSNs were identified by clustering the latent space of a scVI variational autoencoder model, and gene expression programs (GEPs) were identified using consensus non-negative matrix (cNMF), as previously described.⁵¹
- (B) (Left) The fraction of the ~1,300 OSN highly variable genes (HVGs) that showed enriched expression ($\log_2$ fold-changes in the top zone over the second-highest zone above the thresholds indicated on the x axis) for OSN subtypes in a single zone, using discretized zonal indices, as previously inferred for each OR.¹⁶ (Middle) The z-scored expression across OSN subtypes of zonally restricted HVGs (using a threshold of $\log_2$ fold-change > 1). (Right) The number of HVGs with $\log_2$ fold-change > 1 for the OSNs found in zonal bins of the given widths and start positions.
- (C) The mean usage of GEP_{Dorsal} and GEP_{Ventral} by each OSN subtype (defined as the set of OSNs singularly expressing a given OR), for the 1007 OSN subtypes with at least 150 cells in our dataset.
- (D) Histogram depicting the distribution of mean DV scores across OSN subtypes (after DV scores were binned into 61 equal-width bins). OSN subtypes were categorized as dorsal if their mean DV score was greater than 60.
- (E) The distribution of DV scores for each OSN subtype, across different OR classes, and separated for OSN subtypes that were categorized as dorsal or ventral (with the number of ORs in each category indicated); TAARs, trace amine-associated receptors.
- (F) Silhouette scores (quantifying clustering quality) from hierarchical clustering of the pairwise distance matrix of DV scores between OSN subtypes (number of clusters indicated on the x axis, dark line and shading indicate the mean and 95% CIs across restarts in which equal numbers of OSNs were sampled per subtype).
- (G) Accuracy of linear regression models at predicting the DV scores for each OSN subtype, using increasing numbers of random DV or HVGs, or the best set of a given number of DV genes.
- (H) DV scores for each OSN subtype, as measured using either all HVGs or after excluding all known DV marker genes.
- (I) Heatmap depicting, for each OR, kernel density estimates of the distribution of DV scores across the set of OSNs expressing that OR.
- (J) As measured by the area under the receiver operating characteristic curves (AUROC), the distinguishability of OSN subtypes within the same zone, using the zonal identities of Tan and Xie,¹⁶ plotted as a function of the difference in DV scores (as a percentile of all OSN subtypes) of the pair (with lines and shading reflecting the mean and interquartile range for the pairs at each percentile). Shuffles depict the AUROCs when OR labels are shuffled across cells.
- (K) The distribution of AUROC values for ORs found within the same “Zolf” zone, as defined by Zapiec and Mombaerts.¹²
- (L) Accuracy of regression models at predicting the expression of each dorsal or ventral DV gene in each OSN subtype using either its zonal identity or its DV GEP usage.
- (M) (Left) The distribution of pairwise correlations in DV scores between the $_{363}C_2$ pairs of mice in the dataset. (Middle) The correlation of OSN subtype DV scores across different subsets of cells (generated by randomly sampling 150 cells per subtype and splitting the sample into two subsets), compared with the correlation when subtype labels were shuffled across cells. (Right) The mean absolute difference in ranks of the DV score for the data on the left, plotted as a percentage of the total number of OSN subtypes, and compared with a model in which ORs were shuffled across cells of a given zone.
- (N) DV scores for each OSN subtype, as measured in male and female mice.
- (O) The mean change in rank of the DV scores for each OSN subtype across data partitions, as a function of its DV score and colored by the OR class of its associated OR.
- (P) The standard deviation of the DV score for each OSN subtype, as a function of each subtype’s DV score.



(legend on next page)

Figure S2. DV scores predict spatial positions across datasets, related to Figure 2

(A) (Left) All transcripts from the MERFISH experiment, aligned to a common template via STalign and colored by their transcriptional similarity via Baysor. (Middle) Cells, segmented via Proseg, clustered based on their gene expression, and colored by their corresponding cell type labels. LP, lamina propria. IMM, immature cells (GBCs, INPs, and iOSNs). (Right) Heatmap of max-normalized marker gene expression, as measured via MERFISH, for each cell type.

(B) Expression of the indicated example DV genes, as measured directly via MERFISH.

(C) (Top) The observed change in DV_{ISH} scores (i.e., the estimate of the DV score obtained from a subset of the DV genes that were queried via MERFISH) as a function of changes in spatial position within an epithelial segment, with inset demonstrating the continuity of changes in the DV_{ISH} scores as a function of physical space. Shading is mean $\pm$ 95% CIs across epithelial segments. (Bottom) The change in DV scores of the OSN subtypes (as measured via scRNA-seq) at increasing distances along each epithelial segment.

(D) Schematic depicting how individual epithelial segments were aligned to a common coordinate system using the positions of the identified ORs in a manner that was agnostic to the DV score of their respective OSN subtypes, and two example individual segments colored by either their common coordinate-aligned position or by the DV scores associated with ORs detected in each OSN.

(E) (Top) Predictions of the DV score associated with each OSN subtype using regression models trained on the DV_{ISH} scores for each OSN. (Bottom) Predictions of the DV score for each MERFISH-identified OSN subtype based upon the DV scores (as measured via scRNA-seq) associated with the ORs of spatially neighboring OSNs.

(F) (Left) The DV scores (as measured via scRNA-seq) associated with the ORs detected in each region of the MOE via Stereo-seq, a spatial transcriptomic method. (Middle) Predictions of the DV score for each OSN subtype based upon the DV scores associated with the ORs of spatially neighboring OSNs. (Right) Consistency of predicted DV scores across animals.

(G) OR positions in a separate MERFISH dataset from Bintu et al.⁵⁴ (Left) The DV scores associated with the ORs detected in an example section. (Right) DV scores associated with each OR in the epithelial segment from all replicates, before and after the alignment procedure depicted in (D).

(H) (Left and middle) Comparison of the DV scores for each OSN subtype (as identified via scRNA-seq GEP usage) with the OR spatial indices described in past reports that measured the expression of OR genes via bulk RNA-seq in microdissected regions of the MOE.^{16,17} (Right) Results of support vector regression models trained to predict the DV score associated with each OR using the levels of OR expression measured via bulk RNA-seq in microdissected samples from Bashkirova et al.²⁸

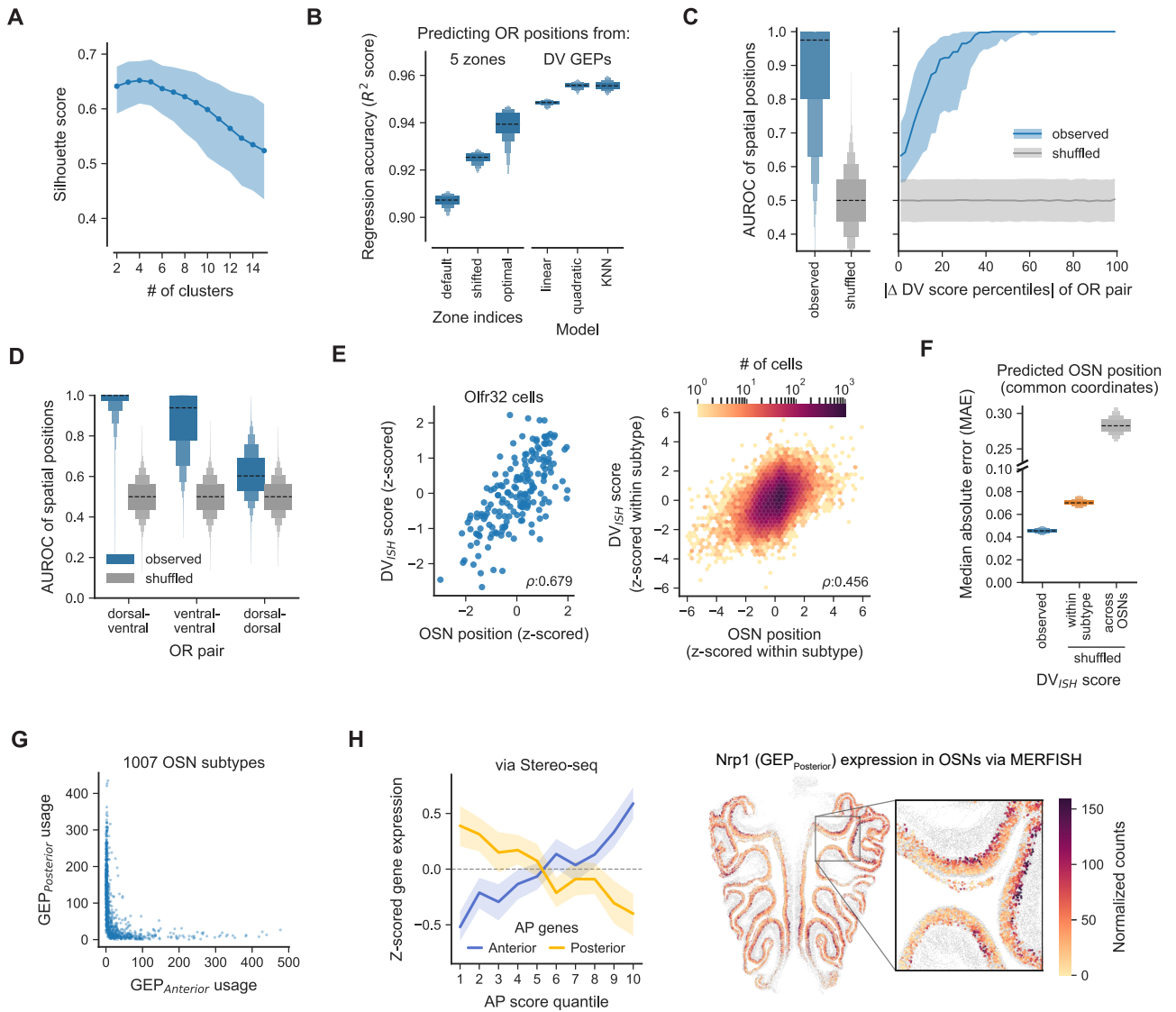


Figure S3. OR positions are not clustered into zones, and AP genes vary along orthogonal axes, related to Figure 2

(A) Silhouette scores from hierarchical clustering of the pairwise distance matrix (in common coordinate space) of the spatial positions of OSN subtypes, with the indicated number of clusters, with the mean and 95% CIs across restarts.

(B) Accuracy of regression models at predicting the spatial position of each OSN subtype using either its zonal identity or its DV GEP usages. Regression features were transformed as indicated.

(C) The discriminability, as measured via area under the receiver-operator curve (AUROC), of pairs of ORs based on their spatial positions, plotted as a function of the difference in DV scores (as a percentile of all OSN subtypes).

(D) The discriminability (AUROC) of the spatial positions of indicated types of OR pairs.

(E) (Left) For OSNs expressing *Olf32*, the z-scored DV_{ISH} score as a function of the z-scored spatial position of each cell. (Right) The DV_{ISH} score for all OSNs as a function of its position, for which both positions and DV_{ISH} scores were z-scored within each OSN subtype.

(F) The median absolute error (MAE) of the predicted spatial position of each OSN in the aligned common coordinates, using regression models trained on each cell's DV_{ISH} score. Predictions were compared with data in which the DV_{ISH} score was shuffled within each OSN subtype or shuffled across all OSNs.

(G) The mean usage of GEP_{Anterior} and GEP_{Posterior} by each OSN subtype.

(H) (Left) The expression of anterior and posterior genes, as measured spatially via Stereo-seq, across OSN subtypes with the given AP scores. (Right) The expression of the GEP_{Posterior} axon guidance gene *Nrp1*, as measured directly via MERFISH. Gene expression was normalized based on the total number of transcripts detected in each OSN.

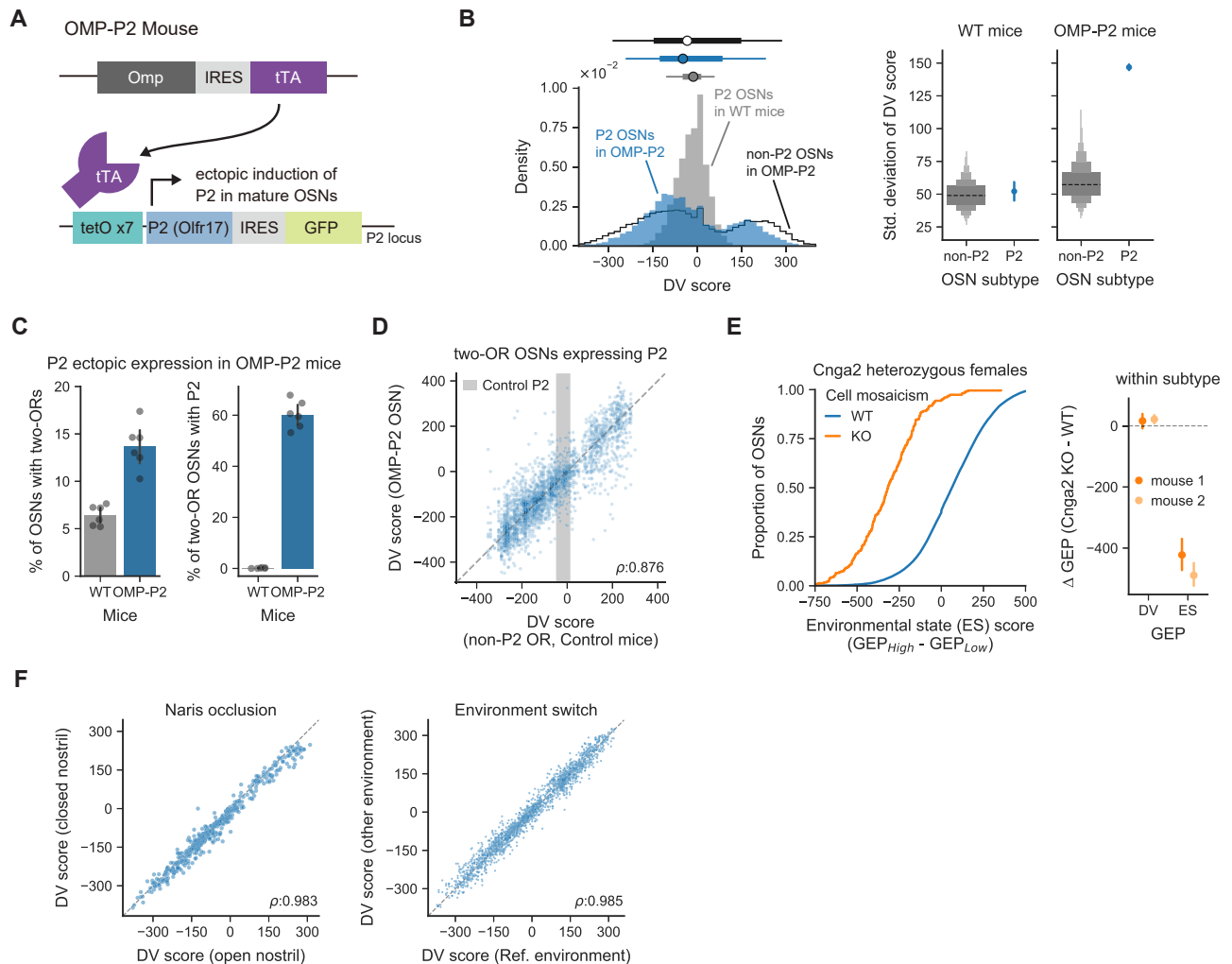


Figure S4. OR expression or activity has little effect on DV scores, related to Figure 3

(A) Schematic depicting induction of the ectopic expression of the P2 OR in mature OSNs in the OMP-P2 mouse.

(B) (Left) Histogram depicting the DV scores for OSNs singularly expressing P2, from either the entire integrated dataset (gray) or from OMP-P2 mice (blue), as well as the distributions of DV scores across all non-P2-expressing OSNs in OMP-P2 mice (black line). Box and whisker plots depict the median, interquartile range, and 5th–95th percentiles. (Right) The distribution of standard deviations of the DV score for OSNs expressing P2 and non-P2 ORs from wild-type (WT) mice or OMP-P2 mice. Error bars indicate 95% CIs.

(C) The percentage of OSNs expressing two ORs in WT and OMP-P2, and the percentage of these two-OR OSNs expressing P2 (and another non-P2 OR). Dots indicate individual replicates, and error bars represent the mean and 95% CIs.

(D) The DV score for two-OR OSNs in OMP-P2 mice expressing P2 and a non-P2 OR, plotted as a function of the DV score associated with the non-P2 OR in control mice. Shaded gray area depicts the interquartile range of the DV score for OSNs in control mice that singularly express P2.

(E) (Left) Cumulative distribution of environmental state (ES) scores, which reflect ongoing OR-dependent activity and are computed as the difference between the GEP_{High} and GEP_{Low} activity GEPs,⁵¹ for OSNs derived from Cnga2 heterozygous females expressing either the WT or loss-of-function Cnga2 allele. (Right) Within OSN subtype change (KO – WT) in GEP usage for the DV and ES GEPs; note that ES scores depend upon CNGA2, while DV scores do not.

(F) (Left) DV scores for each OSN subtype for OSNs from open or occluded nostrils after 4 weeks of chronic naris occlusions. (Right) DV scores for each OSN subtype from mice housed in separate olfactory environments, whose odorants differentially activate the OR repertoire.

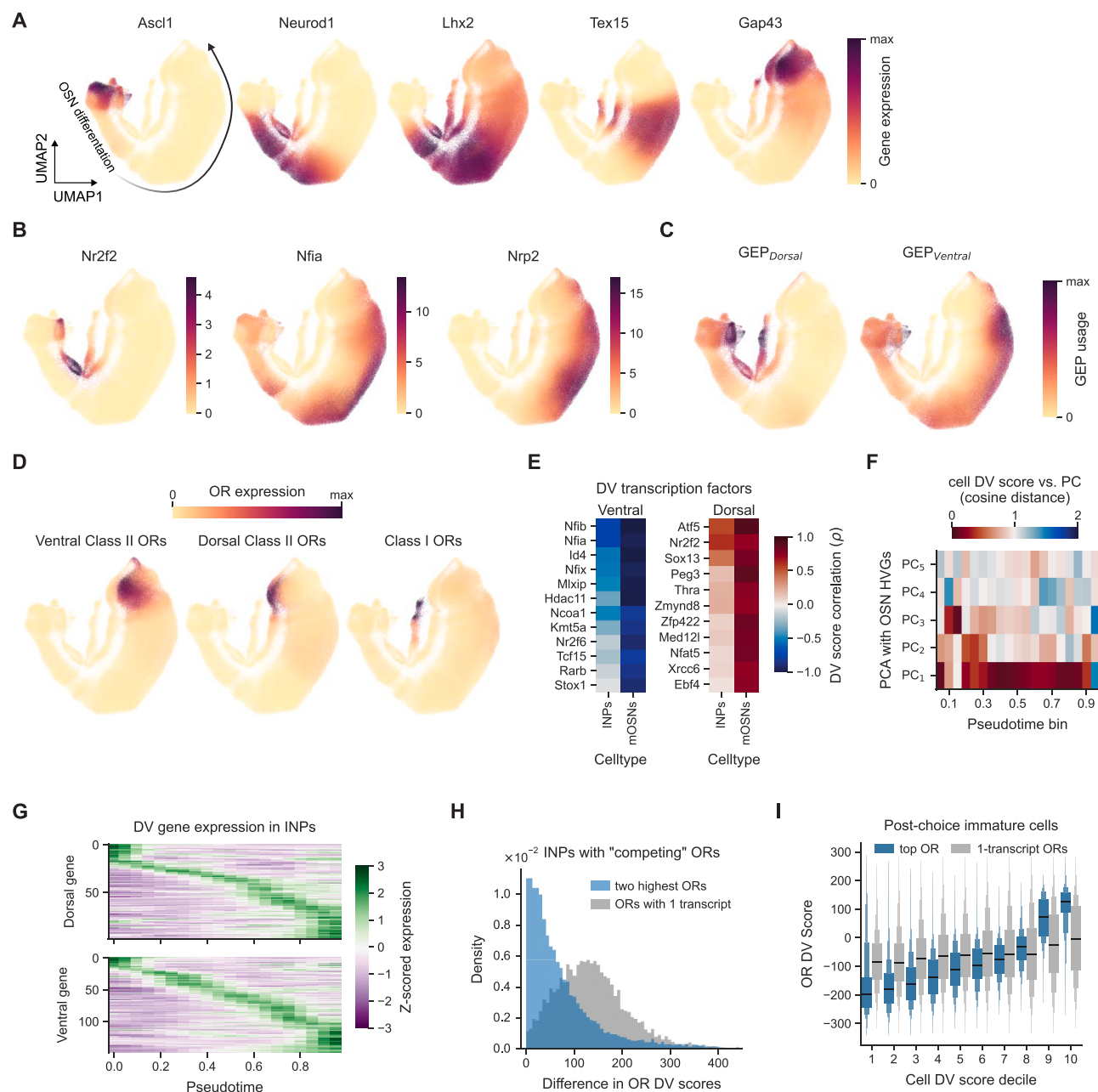


Figure S5. Variation in gene expression during OSN differentiation, related to Figure 3

(A) UMAP plot depicting the indicated marker genes across the differentiating OSN dataset (individual dots are single cells, shading is normalized gene expression); note, the majority of immature OSNs and all mature OSNs are excluded in this dataset. For visualization purposes, gene expression was smoothed based on the nearest neighbor graph.

(B) Similar to (A), but for the expression of example DV genes in differentiating OSNs.

(C) GEP usages for GEP_{Dorsal} and GEP_{Ventral} for each differentiating cell, calculated using the GEP weights for each gene as identified in mature OSNs.

(D) UMAP plots depicting the maximum levels of expression for cells expressing the indicated types of ORs. Note the separate developmental trajectories adopted for differentiating OSNs that eventually choose each type/class of OR.

(E) The correlation of the cellular expression of each of the indicated transcription factors (all of which participate in the DV score and are therefore DV genes) with the DV scores of each cell (with data derived from either the mature OSN or differentiating OSN datasets).

(F) The similarity (cosine distance) of the cell DV score and the top principal components of gene expression, where PCA was performed using OSN variable genes for cells in each pseudotime bin. Note, low-level OR expression does not begin until pseudotime > 0.4.

(G) The expression of each dorsal and ventral gene in INPs, z-scored and sorted by their expression as a function of pseudotime.

(legend continued on next page)

(H) For cells expressing multiple “competing” ORs expressed at moderate levels (three or more transcripts), the difference in DV scores of expressed ORs (as determined in mature OSNs) compared with similar differences among the ORs expressed in the same cells at the lowest expression level (i.e., one transcript per cell).

(I) For post-choice cells that express only a single OR (defined here as having three or more transcripts of that OR), the DV score associated with the chosen OR (determined in mature OSNs) compared with the mean DV score associated with all those ORs expressed at low levels (only one transcript per cell), plotted as a function of each cell’s DV score (split into deciles).

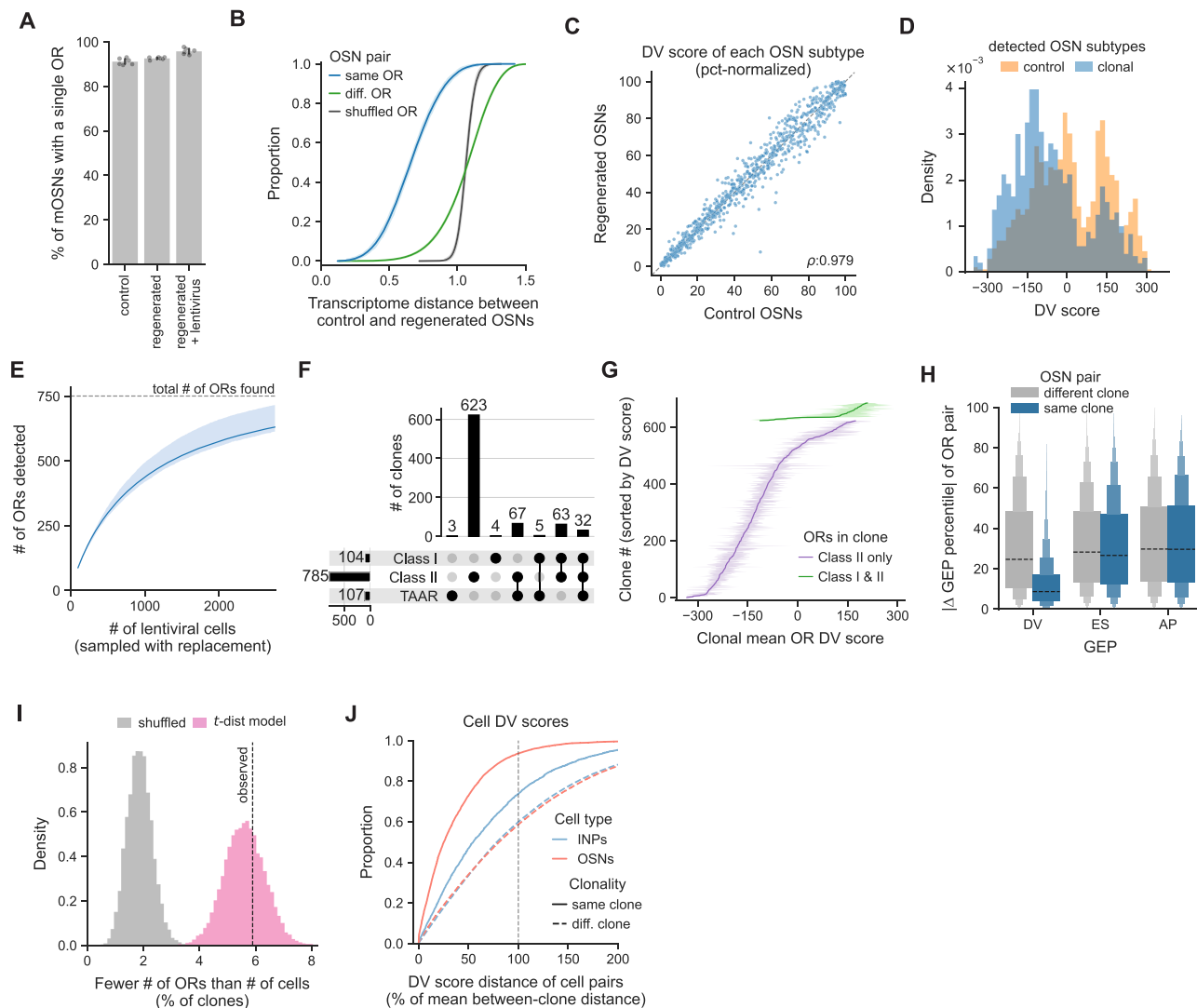


Figure S6. OSN clones have restricted cell fates, related to Figure 3

(A) The percentage of OSNs expressing a singular OR in either WT control mice or regenerated following methimazole injection in the lentiviral lineage tracing experiment. Dots indicate individual replicates and error bars indicate 95% CIs.

(B) The transcriptome distance, defined as the cosine distance of GEP usages, between control and regenerated OSNs that express either the same or different ORs, or when OR labels are shuffled across OSNs. Shaded error indicates the mean and 95% CIs.

(C) Percent-normalized DV scores associated with each OSN subtype in control and regenerated conditions.

(D) Distribution of DV scores for all OSN subtypes in control conditions, and the distribution of DV scores of ORs detected in lentiviral-infected OSNs in the clonal lineage tracing experiment.

(E) Cumulative number of ORs detected as a function of the number of barcoded cells. Line and shaded error indicate the mean and 95% CIs across restarts.

(F) UpSet plot depicting the number of clones that contained OSNs collectively expressing each combination of OR classes. The total number of cells expressing each receptor type is indicated on the bottom y axis. The bar graph indicates the number of clones harboring OSNs expressing receptors of the dotted classes (below each bar). Note that individual clones can harbor OSNs expressing ORs from different classes.

(G) For each clone, the mean and 95% CIs of the DV scores (as measured in control conditions) associated with the OSN subtypes of that clone. Clones are sorted by the mean DV score associated with their expressed ORs.

(H) Distances in DV, AP, and ES scores (as measured in control conditions) of OSN subtypes found in pairs of OSNs, colored by whether each pair was clonally related or not. ES scores reflect ongoing OR-dependent activity, whereas AP scores reflect OR-dependent but odor-independent activity.

(I) The percent of clones with fewer ORs than cells (which thus contain multiple OSNs expressing the same OR) for the observed data, for data with shuffled clonal labels, and for clones generated using the observed t -statistics to sample ORs.

(J) Cumulative distribution of the distance in DV scores of pairs of INPs or OSNs that either came from the same or different clones.

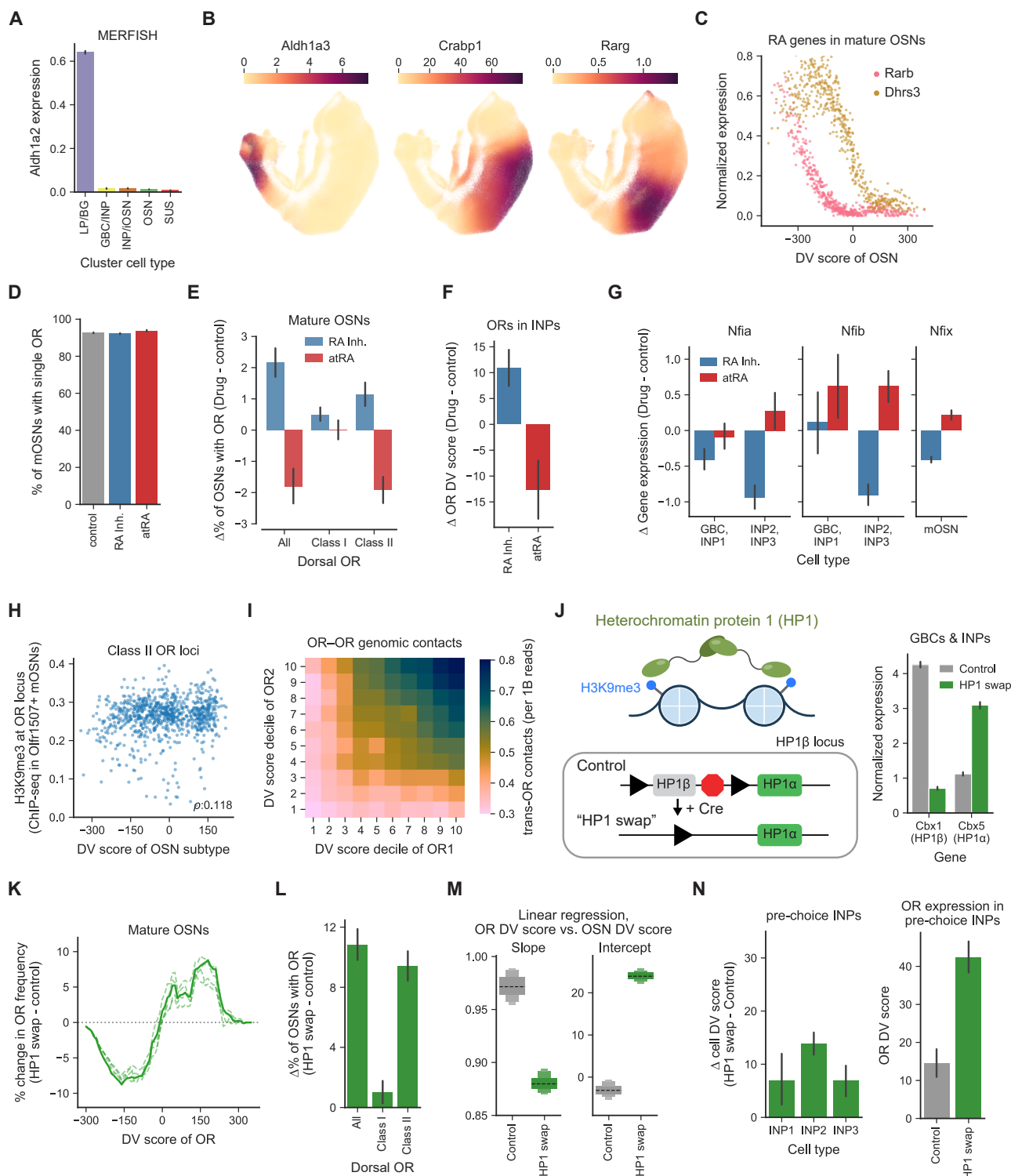


Figure S7. RA signaling components vary spatially to influence OSN identities, and heterochromatin binding influences OR choice, related to Figures 4 and 5

(A) Expression of *Aldh1a2*, as measured via MERFISH, in each cell type (log-normalized number of transcripts per cell). LP/BG, lamina propria and Bowman's gland cells.

(B) Expression of genes involved in RA signaling in the differentiating OSN dataset, including the retinaldehyde dehydrogenase 2 (*Aldh1a3*), the cellular retinoic acid binding protein *Crabp1* in immature cells, and the nuclear RAR gamma (*Rarg*).

(legend continued on next page)

(C) Normalized gene expression of the RA-related genes *Rarb* and *Dhrs3* in mature OSNs, as a function of the DV score of each OSN. OSNs were binned into equal-frequency bins, and each dot depicts the average expression in each bin.

(D) The percent of OSNs following methimazole-induced regeneration that expressed single ORs, for animals treated with either control vehicle, the RA inhibitor WIN 18,446 (RA Inh.), or atRA, with error bars indicating 95% CIs.

(E) For mice given each drug, the change in frequency (relative to control mice) of the percent of cells expressing dorsal ORs, plotted for all dorsal ORs and separated by OR class. Error bars depict 95% CIs.

(F) Change in the DV scores associated with each OR in mature OSNs for the ORs expressed in pre-choice INP cells. For cells expressing multiple ORs, the DV scores associated with each OR were weighted by their expression levels.

(G) Expression of the NFI family transcription factors *Nfia*, *Nfib*, and *Nfix* in the indicated cell types relative to their expression in control mice, for mice given either the RA inhibitor or atRA.

(H) As measured via ChIP-seq from OSNs expressing the ventral receptor Olfr1507,²⁸ the levels of the heterochromatin mark H3K9me3 at class II OR loci versus the DV score associated with each OR.

(I) As measured via the high-throughput chromatin conformation method Micro-C, the frequency of interchromosomal OR-OR contacts between pairs of class II OR loci, binned as a function of their respective DV scores.

(J) (Left) Schematic of the interaction between heterochromatin protein 1 (HP1) and histones modified with the heterochromatin-promoting H3K9me3 mark, and depiction of the “HP1 swap” experiment, in which the loss of HP1 β is rescued with HP1 α ; in these experiments, HP1 β was conditionally knocked out of neurons using the Foxg1-Cre mouse line. (Right) The normalized expression of HP1 β and HP1 α in differentiating GBCs and INPs from either control or HP1 swap mice.

(K) Change in frequency of ORs with given DV scores for data from HP1 swap relative to control mice. Dashed lines indicate individual replicates.

(L) Similar to (E) but for the change in the percentage of cells expressing dorsal ORs in HP1 swap mice relative to control mice.

(M) Bootstrap estimates of the slope and intercept for robust linear regression models fit on the cell to OR DV score mapping in each condition.

(N) For pre-choice INPs, (left) the mean change in the cell DV scores (HP1 swap—control) for the cells at the indicated stage or (right) the mean DV scores associated with the ORs (as measured in mature OSNs) for the ORs expressed in INPs from each condition, where for cells expressing multiple ORs the DV scores associated with each OR were weighted by their expression. Error bars depict 95% CIs.

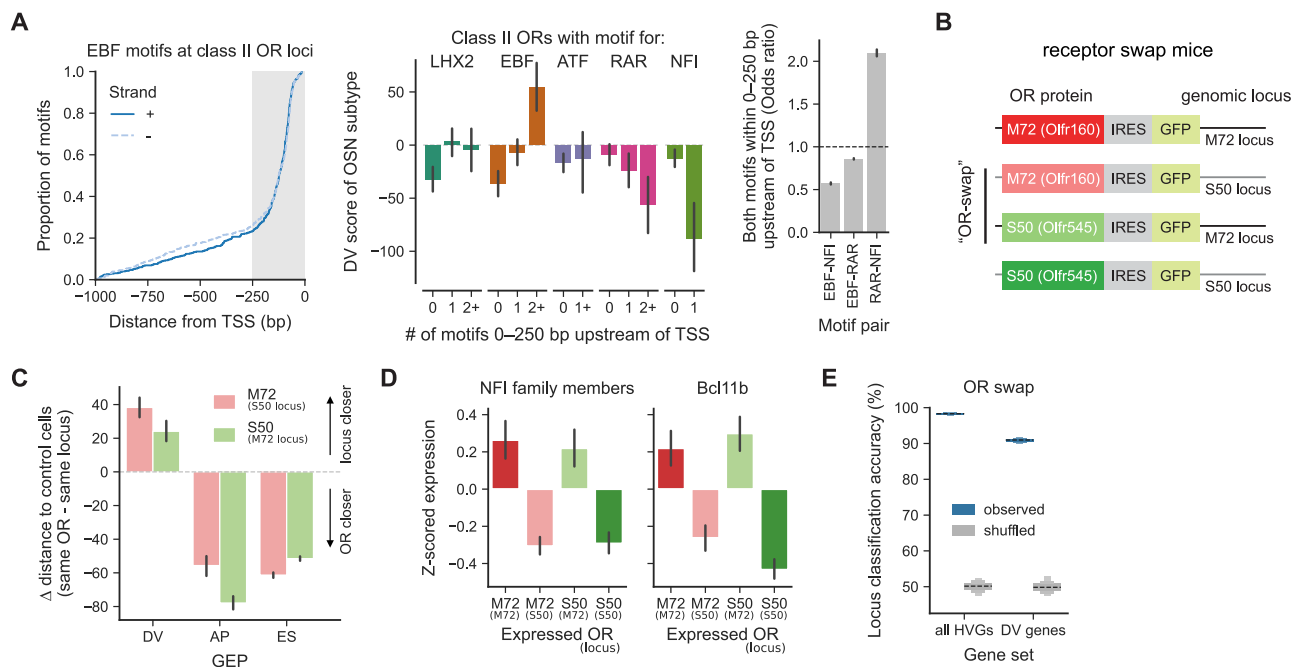


Figure S8. DV scores are a genomic property of each OR, related to Figure 6

(A) (Left) Cumulative proportion of EBF motifs in the 1 kb region upstream of each OR's transcriptional start site (TSS). Note the density of motifs in the 0–250 bp promoter region upstream of each TSS. (Middle) The mean and 95% CIs of the DV score for OSN subtypes whose ORs have the indicated number of motifs for each transcription factor in the upstream 0–250 bp promoter region. (Right) The odds ratio of having both motifs in the promoter regions of OR loci for the indicated motif pairs.

(B) Schematic of the receptor swap experiment, in which OSNs express the M72 or S50 OR from either the M72 or S50 genomic locus. Note that although these mice are referred to as receptor "swaps" because M72 (or S50) is expressed from the S50 (or M72) locus, there remains M72 (or S50) expressed from its original locus in the genome. As a result, in these mice, there are two separate genomic locations associated with the same receptor.

(C) The distances in GEP usages for each cell from the OR swap mice to the mean of the respective control animal (in which ORs were expressed from their endogenous loci). The difference in distances to the same OR versus the same locus in control animals is shown for each GEP.

(D) The average expression of the NFI family and Bcl11b transcription factors (both of which are enriched in ventral OSNs) across the animals examined in the OR swap experiments.

(E) Accuracy of support vector machine classifiers at distinguishing the genomic locus of each cell, using either all HVGs or the DV genes, for the observed data and data in which the genomic locus was shuffled across OSNs.

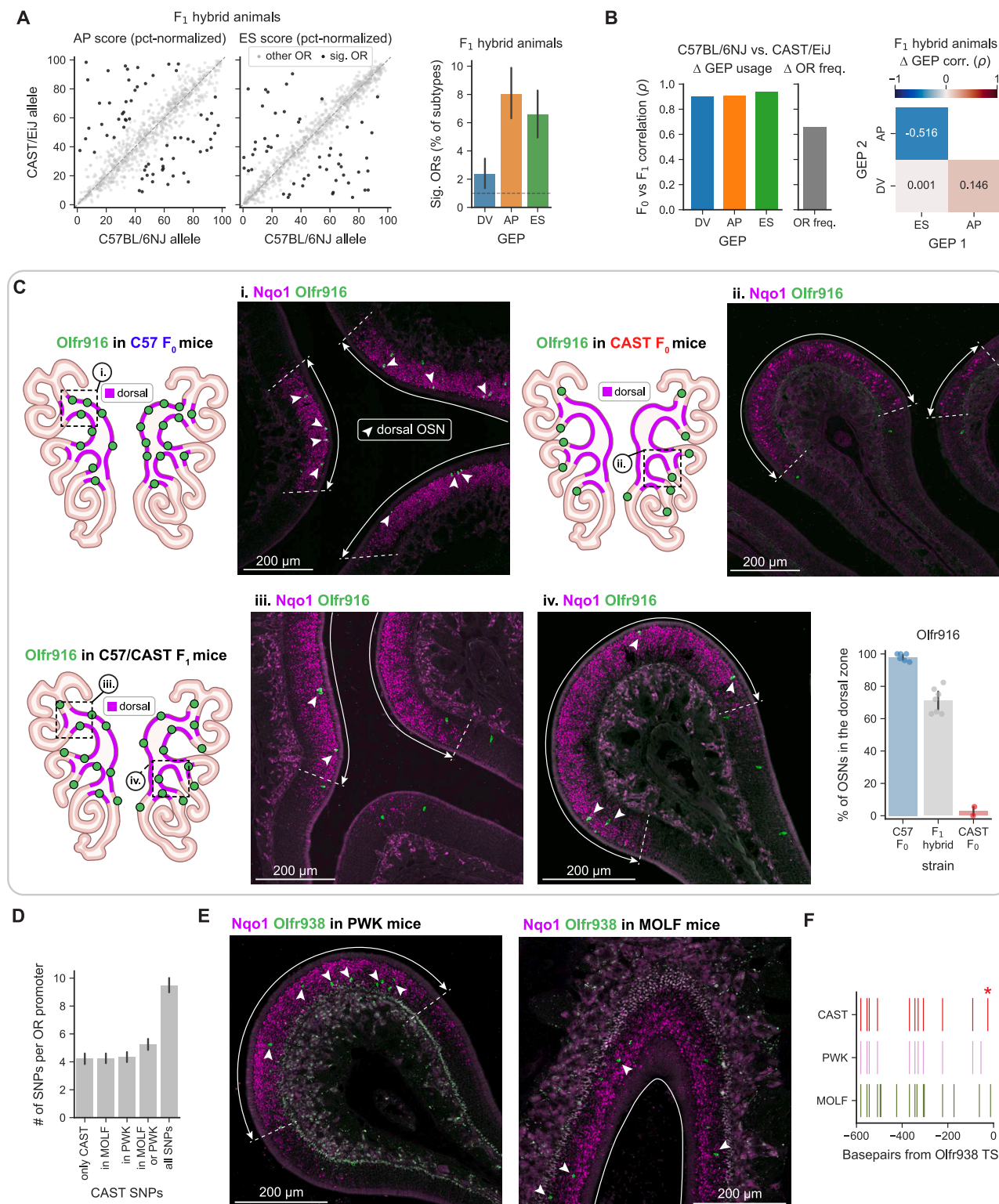


Figure S9. DV scores and OR spatial positions vary across mouse strains, related to Figure 6

(A) AP and ES scores for each OSN subtype, evaluated separately for OSNs from F₁ animals that chose either the C57BL/6J or CAST/EiJ allele for a given OR. OSN subtypes with significant changes in GEP usages across strains are highlighted in black. Note, this strain-specific variation is expected to be the consequence of coding changes in the receptor itself. (Right) The percentage of OSN subtypes with strain-specific changes for each GEP.

(legend continued on next page)

(B) (Left) For OSN subtypes with significant changes for each GEP, the correlation between the change in GEP usages (CAST – C57 allele) in F_1 hybrid animals and the change in GEP usages across F_0 animals (CAST – C57 mice), and the correlation in the change in OR frequency across alleles in F_1 animals with that across strains in F_0 animals is shown. (Right) For each GEP pair, the correlation in the change in GEP usage across strains for the set of OSN subtypes with significant changes in either GEP from each GEP pair is shown.

(C) Schematics and images depicting *in situ* hybridizations for OSNs expressing Olfr916 (in green) in either F_0 or F_1 animals, with the dorsal marker Nqo1 in magenta, the dorsal zone outlined in white, and dorsal ORs labeled with arrowheads. (Lower right) Summary quantification of the percentage of OSNs found in each animal that were located within the dorsal zone, with dots depicting individual replicates and error bars depicting 95% CIs.

(D) The number of SNPs in the 1 kb upstream region of OR loci between CAST and C57 that are either specific to CAST or that are also shared in either PWK/PhJ or MOLF/EiJ mice.

(E) Example *in situ* hybridizations for dorsal Nqo1⁺ OSNs (magenta) expressing Olfr938 (green) in either PWK or MOLF mice.

(F) The presence of SNPs in the upstream region of Olfr938 across CAST, PWK, and MOLF, with an asterisk depicting a SNP present only in CAST.

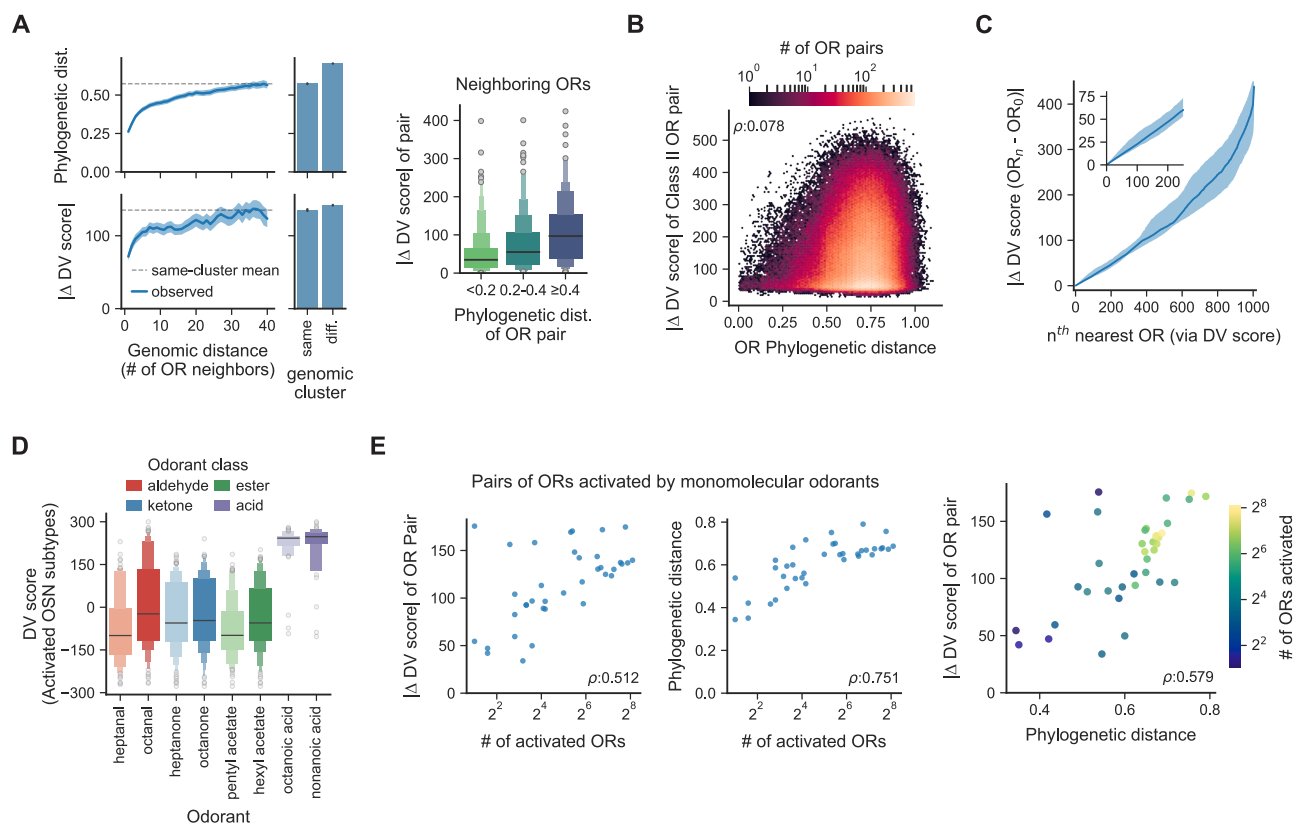


Figure S10. Phylogeny, function, and spatial location are weakly related, related to Figure 6

(A) (Left) For OSN subtypes whose ORs are located in the same genomic cluster on a given chromosome, the phylogenetic distance of their amino acid sequences or the average distance in their associated DV scores, plotted as a function of genomic distance (defined as the number of functional OR genes separating the ORs of each OSN subtype). (Right) The distribution of DV score distances for pairs of OSN subtypes whose ORs are the nearest genomic neighbors, separated based on the phylogenetic distance of the OR pair.

(B) The relationship between phylogenetic distances of pairs of class II ORs as a function of their differences in DV scores.

(C) The median and interquartile range of the DV score distances for the n^{th} nearest OSN subtype, as based on DV scores.

(D) Act-seq-based measurements of acute odor-driven activity (quantified as the Activation Score, which summarizes acute odor-driven increases in ~200 immediate early and other genes⁵¹). The distribution of DV scores for activated OSN subtypes for broadly activating monomolecular odorants. Note, the acids exhibit significant spatial clustering of their responses, likely because they are largely detected by class I ORs located in the dorsal region of the epithelium.

(E) The average DV score distance (left) or mean phylogenetic distance (middle) for pairs of activated ORs, as a function of the number of ORs activated by each odorant, demonstrating that sparse odorants activated more spatially and phylogenetically similar ORs. (Right) The average DV score distance as a function of the mean phylogenetic distance for pairs of ORs activated by each odorant.