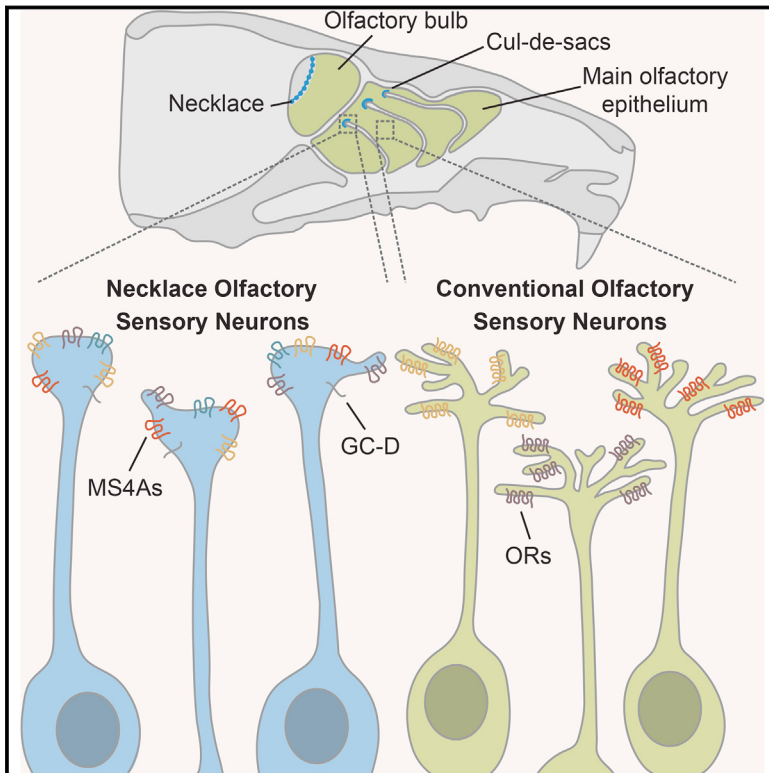


A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction

Graphical Abstract



Authors

Paul L. Greer, Daniel M. Bear, Jean-Marc Lassance, ..., Rory Kirchner, Hopi E. Hoekstra, Sandeep Robert Datta

Correspondence

srdatta@hms.harvard.edu

In Brief

Chemosensory receptors in a subset of mammalian olfactory sensing neurons are structurally distinct from GPCRs, and multiple subtypes are expressed per neuron, suggesting an unexpected mechanism for olfactory detection and encoding.

Highlights

- The MS4A proteins are mammalian chemosensors expressed in the “necklace” olfactory system
- Ligands for these non-GPCR receptors are enriched for food odors and pheromones
- Unlike conventional odor receptors, many MS4As are expressed in each sensory neuron



A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction

Paul L. Greer,^{1,4} Daniel M. Bear,^{1,4} Jean-Marc Lassance,² Maria Lissitsyna Bloom,¹ Tatsuya Tsukahara,¹ Stan L. Pashkovski,¹ Francis Kei Masuda,¹ Alexandra C. Nowlan,¹ Rory Kirchner,³ Hopi E. Hoekstra,² and Sandeep Robert Datta^{1,*}

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

²Departments of Molecular and Cellular Biology and Organismic and Evolutionary Biology, Center for Brain Science, Harvard University, Howard Hughes Medical Institute, Cambridge, MA 02138, USA

³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

⁴Co-first author

*Correspondence: srdatta@hms.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2016.05.001>

SUMMARY

Odor perception in mammals is mediated by parallel sensory pathways that convey distinct information about the olfactory world. Multiple olfactory subsystems express characteristic seven-transmembrane G-protein-coupled receptors (GPCRs) in a one-receptor-per-neuron pattern that facilitates odor discrimination. Sensory neurons of the “necklace” subsystem are nestled within the recesses of the olfactory epithelium and detect diverse odorants; however, they do not express known GPCR odor receptors. Here, we report that members of the four-pass transmembrane MS4A protein family are chemosensors expressed within necklace sensory neurons. These receptors localize to sensory endings and confer responses to ethologically relevant ligands, including pheromones and fatty acids, *in vitro* and *in vivo*. Individual necklace neurons co-express many MS4A proteins and are activated by multiple MS4A ligands; this pooling of information suggests that the necklace is organized more like subsystems for taste than for smell. The MS4As therefore define a distinct mechanism and functional logic for mammalian olfaction.

INTRODUCTION

As animals navigate the natural world, they encounter an unending variety of small molecules, which are rich sources of information that signify the presence of organisms and salient objects in the environment. The olfactory system detects many of these molecules through odorant receptor proteins expressed by peripheral olfactory sensory neurons (OSNs), which are coupled to higher brain circuits mediating odor perception (Axel, 1995; Ihara et al., 2013). In mammals, the olfactory system is divided into multiple, parallel processing streams made up of anatomically and molecularly distinct sensory neuron populations. The largest subdivision, the main olfactory system, is capable of detecting nearly all volatile odorants and plays key roles in

odor discrimination and learning. Smaller subsystems (such as the vomeronasal system) are thought to play a more specialized role in odor perception, discriminating odors of innate significance and releasing specific patterns of reproductive, agonistic, or defensive behavior (Munger et al., 2009).

The main and vomeronasal olfactory systems each express characteristic odorant receptor families that belong to the G-protein-coupled receptor (GPCR) superfamily; these receptor families define the specific receptive fields and therefore the function of each subsystem (Buck and Axel, 1991; Dulac and Axel, 1995; Herrada and Dulac, 1997; Liberles and Buck, 2006; Liberles et al., 2009; Matsunami and Buck, 1997; Rivière et al., 2009; Ryba and Tirindelli, 1997). The identification of these receptor genes (including the odorant receptors [ORs], vomeronasal type 1 receptors, vomeronasal type 2 receptors [V2Rs], formyl peptide receptors, and the trace amine-associated receptors) has revealed a key organizational principle: each mature olfactory sensory neuron (with the exception of those within the basal subdivision of the vomeronasal system) expresses just a single receptor gene of the hundreds encoded in the genome (Dalton and Lomvardas, 2015). This pattern of expression defines specific information channels in the olfactory system, as the axons of those sensory neurons that express the same odorant receptor converge on a small number of insular structures within the olfactory bulb called glomeruli; these glomeruli are differentially recruited as animals sense distinct smells, enabling the brain to discriminate odors detected by the nose (Mori and Sakano, 2011). Individual basal vomeronasal sensory neurons also target specific bulb glomeruli but express two V2Rs instead of a single receptor (Martini et al., 2001).

While the identification of odorant receptor genes has led to deep insight into the sensory tuning and functional architecture of the main and vomeronasal subsystems, there are additional mammalian subsystems whose modes of odor detection—and therefore function—are less clear. Particularly mysterious is the “necklace” subsystem, which is distinguished by its unusual anatomy: OSNs within this subsystem are concentrated in the recesses of the olfactory epithelium (the “cul-de-sac” regions), and project axons to a ring of 12–40 apparently interconnected glomeruli that encircle the caudal olfactory bulb like beads on a necklace (Juifls et al., 1997; Shinoda et al., 1989). Necklace

sensory neurons and glomeruli respond to a diverse range of chemical stimuli, including gases (such as carbon disulfide and carbon dioxide), pheromones (such as 2,5-dimethylpyrazine (2,5-DMP), 2-heptanone, and E-farnesene), plant-derived odorants, and urinary peptides (Fülle et al., 1995; Hu et al., 2007; Juilfs et al., 1997; Leinders-Zufall et al., 2007; Meyer et al., 2000; Munger et al., 2010; Sun et al., 2009). While many of these ligands have innate significance for the mouse, the specific role of the necklace in olfactory perception remains unclear.

Intriguingly, necklace OSNs do not express the signaling proteins known to mediate GPCR-based chemotransduction in the rest of the main olfactory epithelium (Juilfs et al., 1997; Meyer et al., 2000). While the ability of the necklace system to detect and behaviorally respond to gases and peptides requires the single-pass transmembrane protein guanylate cyclase-D (GC-D), which is specifically expressed in all necklace neurons, the remainder of the diverse sensory responses observed in this system are unexplained (Guo et al., 2009; Leinders-Zufall et al., 2007; Sun et al., 2009). These observations suggest that necklace OSNs harbor an as-yet unrecognized receptor type, whose identification could reveal key features of the functional organization and neural logic that governs the necklace system.

Here, we show that necklace OSNs express a previously unidentified class of chemoreceptor encoded by the *Ms4a* gene family. Each MS4A protein detects a specific set of odors—including ethologically relevant odorants like fatty acids and the putative mouse pheromone 2,5-DMP—that stimulate necklace sensory neurons in vivo. Ectopic expression of MS4A proteins is sufficient to confer responses to MS4A ligands upon conventional olfactory neurons. However, unlike all known mammalian olfactory receptors, the *Ms4a* genes do not belong to the GPCR superfamily and are not expressed in the conventional one-receptor-one-neuron pattern; instead, each *Ms4a* gene encodes a four-pass transmembrane protein, and many *Ms4a* family members are expressed in every necklace sensory neuron. Taken together, this work defines a new mechanism for mammalian olfaction and identifies a population of atypical olfactory sensory neurons that each express many members of a receptor gene family, suggesting a distinct perceptual role for odor information coursing through the necklace subsystem. Because MS4A proteins are also expressed in chemosensory cells that reside outside of the nasal epithelium, these findings further suggest a broader role for the MS4A proteins in the detection of chemical cues.

RESULTS

To identify candidate receptor gene families specific to the necklace olfactory system, RNA sequencing (RNA-seq) was used to compare transcripts expressed by FACS-isolated GC-D-expressing and conventional OSNs (Figures 1A, 1B, S1A, and S1B). This analysis failed to reveal expression of known odorant receptor families or enrichment of other GPCR subfamilies within necklace sensory neurons. Consistent with this and prior reports, *Golf*, *Adcy3*, *Cnga2*, *Cnga4*, *Trpc2*, and *Trpm5*—gene products required for odor-related signal transduction in con-

ventional OSNs—also were not expressed in GC-D cells (Munger et al., 2009). To screen for potential non-GPCR odorant receptors, the RNA-seq data were filtered to identify transmembrane protein families that exhibit sufficient molecular diversity to potentially interact with a wide range of ligands. As detailed below, this screen identified the *membrane-spanning, 4-pass A* (*Ms4a*) genes, which encode a class of four-transmembrane (4TM)-spanning proteins that are structurally distinct from GPCRs (Eon Kuek et al., 2015).

RNA-seq revealed transcripts for several *Ms4a* family members in GC-D-expressing cells (Figure 1B). Because *Ms4a* transcripts had low RNA-seq read counts, the Nanostring single-molecule detection technique was used to unambiguously determine the presence of every member of the *Ms4a* gene family in GC-D cells (Khan et al., 2011). This analysis revealed the reproducible expression of 12 *Ms4a* family members (of the 17 members annotated in the mouse genome), demonstrating that GC-D cells express low levels of a specific subset of *Ms4a* genes (Figure 1C). None of these 12 genes was detected in conventional OSNs above background (data not shown). Notably absent from GC-D cells are the two best-studied *Ms4a* genes, *Ms4a1* and *Ms4a2*, which have been implicated in calcium signaling downstream of the B cell receptor and high-affinity Fc-Epsilon receptor, respectively but whose precise function remains unclear (Bubien et al., 1993; Dombrowski et al., 1998; Koslowski et al., 2008; Lin et al., 1996; Polyak et al., 2008).

To assess the molecular diversity of the MS4A family, *Ms4a* genes were identified from representative species of all major mammalian lineages. We then asked how different these genes were within a given species, as amino acid differences are a prerequisite for individual MS4As to interact with distinct odors. Multiple sequence alignments revealed substantial intraspecies diversity among the MS4As, particularly within the extracellular domains of the protein, whose length is variable (Figures S2A and S2B). This diversity is comparable to that observed in the third through seventh transmembrane domains in conventional ORs, the regions thought to form ligand-binding pockets that give rise to odorant specificity (Buck and Axel, 1991; Man et al., 2004; Singer, 2000). These findings raise the possibility that each MS4A within a given organism may interact with a distinct set of extracellular cues.

Ms4a genes are found in a single genomic cluster in all queried mammals; this organization, suggestive of tandem duplication, is also found in known chemoreceptor gene families and is thought to facilitate diversification of family members (Figure 2A) (Nei et al., 2008). We therefore also assessed differences between *Ms4a* genes across species to identify those regions of the MS4A protein subject to diversifying or purifying selection. MS4A proteins were significantly divergent across evolution (Figure S2C) (Nei et al., 2008); the most rapidly evolving amino acid residues in the MS4A proteins are highly enriched in the predicted extracellular loops, where contact with environmental chemical stimuli could occur (Figure 2B). Bitter taste receptors, which accommodate the specific diet of their host organism, exhibit a similar pattern of diversifying selection (Figure 2B) (Hayakawa et al., 2014; Wooding, 2011). In contrast, members of the *Orai* family, which

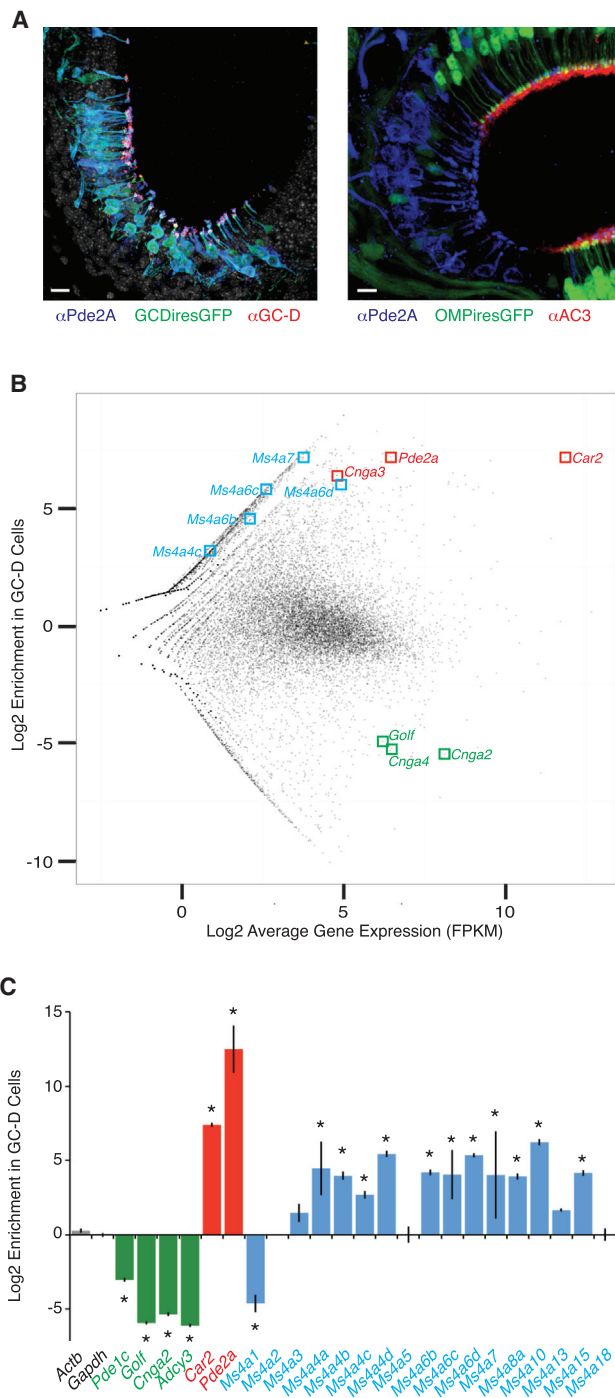


Figure 1. Expression of *Ms4a* Genes in Necklace Sensory Neurons
 (A) (Left) The *Gucy2d-IRES-TauGFP* allele marks necklace sensory neurons expressing PDE2A (blue) and GC-D (red). Grayscale is nuclear counterstain (DAPI). (Right) Pde2a+ necklace sensory neurons reside in epithelial cul-de-sacs and do not express the *Omp-IRES-GFP* allele or the conventional OR signal transduction protein adenylyl cyclase3 (red). Scale bars, 10 μ M.
 (B) Enrichment versus expression plot for every sequenced mRNA in GC-D+ and OMP+ sensory neurons. Each point is a transcript with detectable RNA-seq reads, with marker genes associated with GC-D cells (*Car2*, *Pde2a*, and *Cnga3*) and OMP cells (*Golf*, *Cnga2*, and *Cnga4*) labeled in red

encode 4TM proteins and are not thought to detect environmental chemical stimuli, show no signs of diversifying selection (Figure 2B) (Amcheslavsky et al., 2015). The MS4A family therefore exhibits a pattern of expansion and divergence similar to other chemosensors, although the observation that both within- and between-species MS4A sequence variability is enriched within extracellular loops—rather than traditional hydrophobic binding pockets—suggests that, if the MS4As detect odors, they do so through a distinct domain from conventional GPCR ORs.

MS4A Proteins Confer Odor Responses In Vitro

Although no endogenous or natural ligands have been identified for any member of the MS4A family, the specific expression of a molecularly diverse complement of MS4As within olfactory sensory neurons—taken with prior evidence suggesting involvement in calcium signaling—raised the possibility that *Ms4a* genes encode a novel class of chemoreceptor (Bubien et al., 1993; Dombrowicz et al., 1998; Koslowski et al., 2008; Lin et al., 1996; Polyak et al., 2008). To test whether specific interactions between odors and MS4A proteins induce calcium influx in cells, we heterologously expressed individual MS4A proteins together with the genetically encoded fluorescent calcium indicator GCaMP6s in HEK293 cells (Figures S3A and S3B); expression of MS4A proteins did not increase the baseline rate of calcium transients (Figures S3C and S3D). Six MS4A proteins (selected for their structural diversity) were exposed to 11 compound mixtures, each of whose constituents shared similar chemical structure. These mixes were designed to cover a broad swath of odor space and included known ligands for conventional and necklace glomeruli (Gao et al., 2010; Saito et al., 2009). Increases in intracellular calcium were observed during odor exposure for specific MS4A protein/odor mixture pairs, demonstrating that MS4A proteins transduce signals reflecting the presence of extracellular small molecule ligands (Figures 3A and 3B).

MS4A sensory responses were specifically tuned to particular odor categories, with responses enriched for long-chain fatty acids, steroids, and heterocyclic compounds. To identify individual MS4A ligands, the odorant mixture that evoked the largest response for each MS4A was broken down into its monomolecular constituents (Figure 3C). Individual MS4As conferred responses to a specific subset of odor molecules within a functional class. For example, cells expressing MS4A6C responded to a known ligand for necklace glomeruli, 2,5-dimethylpyrazine (2,5-DMP), as well as to a molecule not previously known to activate the necklace, 2,3-dimethylpyrazine (2,3-DMP)—but only weakly to 2,6-dimethylpyrazine (2,6-DMP) and not at all to other structurally similar molecules in the parent mixture (Figure 3C).

and green, respectively; reliably detected *Ms4a* family members are highlighted in blue.

(C) NanoString-based mRNA quantification of GC-D cells relative to OMP cells. Marker genes for OMP and GC-D cells are shown in green and red, respectively. Twelve of the 17 annotated *Ms4a* genes are significantly enriched in GC-D cells relative to OMP cells. Data are presented as mean $\pm$ SEM from three independent biological samples. * $p < 0.05$, unpaired t test. See also Figure S1.

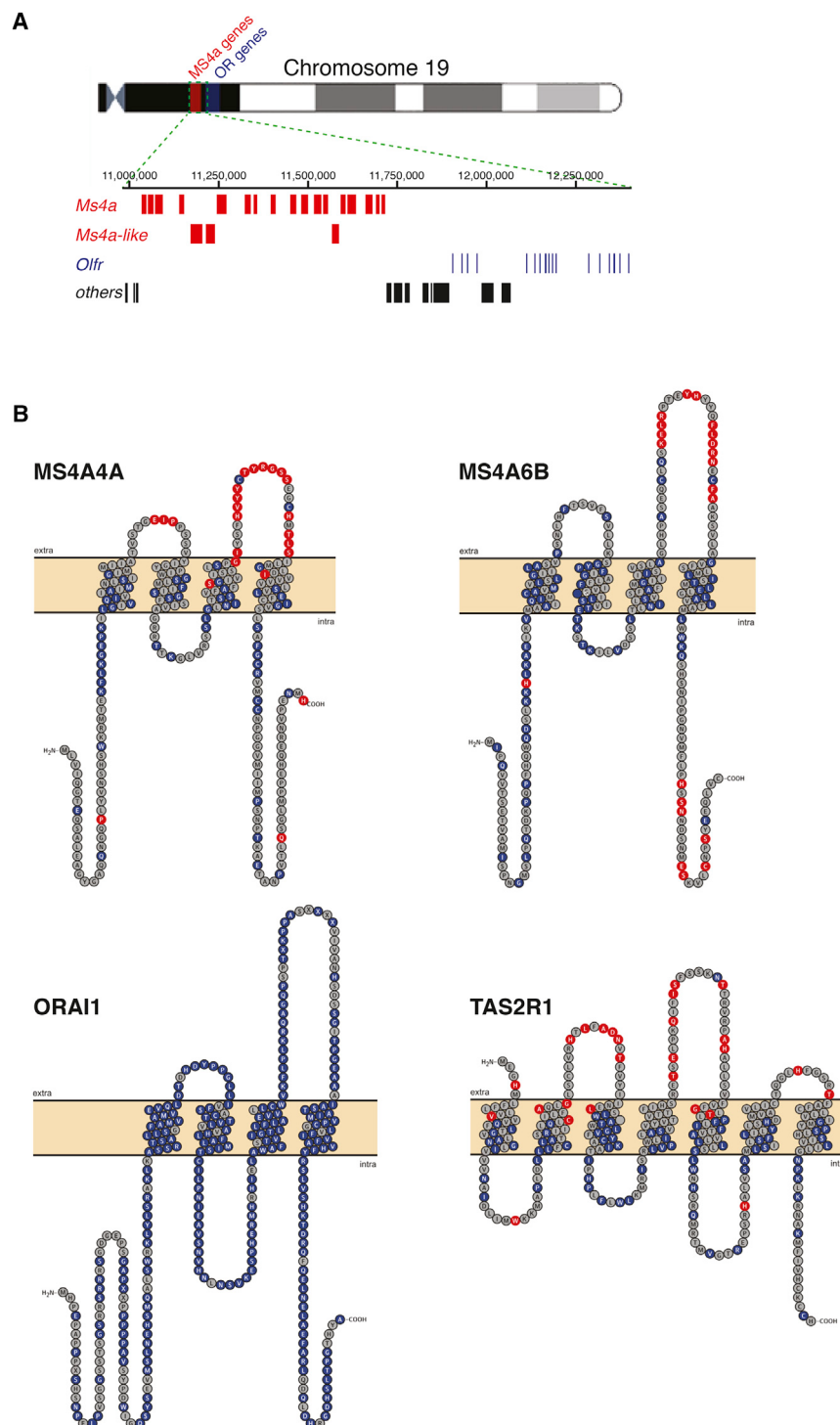


Figure 2. Genomic Clustering and Positive Selection of the *Ms4a* Gene Family

(A) Graphical representation of chromosome 19 of *Mus musculus*, including the entire *Ms4a* gene family (red); immediately telomeric to the *Ms4a* gene cluster resides a large group of conventional mammalian odorant receptor genes (blue).

(B) Topographical representations of the primary sequences of *Mus musculus* MS4A4A (upper-left), MS4A6B (upper-right), ORAI1 (lower-left), and TAS2R1 (lower-right). Amino acid residues under strong purifying selection are shown in blue, whereas those under positive selection are shown in red (posterior probability > 0.90, see [Experimental Procedures](#)); residues under positive selection are enriched within extracellular loops of MS4A proteins and bitter taste receptors ($p = 5.06 \times 10^{-16}$ and $p = 6.76 \times 10^{-7}$, respectively, hypergeometric test).

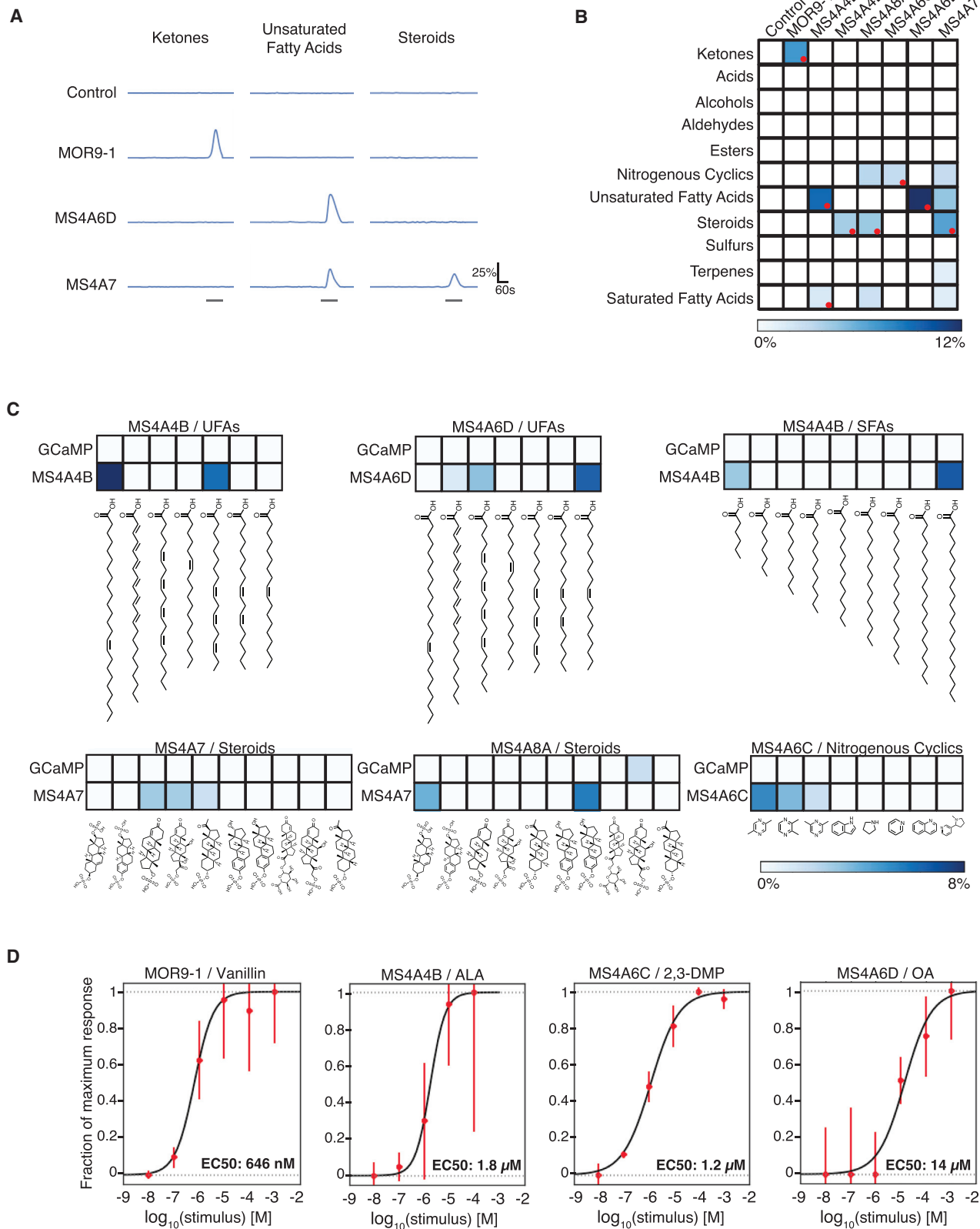
See also [Figure S2](#).

We also asked whether the MS4As and conventional ORs responded to odors with similar kinetics in vitro. No statistically significant differences were observed in time to response onset, time to half-maximal response, or time to peak response either between four queried MS4As or between these MS4As and the conventional odorant receptor MOR9-1 ([Figure S4C](#)). However, because the bulk calcium imaging assay used to screen MS4A ligands is not optimized for assessing response timing, we verified these results using GCaMP6f (a more rapidly responding calcium indicator than GCaMP6s), a faster imaging rate, and a stimulus pencil to focally deliver odorants directly above the imaged cells. These experiments revealed that MS4A responses occur seconds after odor presentation, with similar or slightly faster response dynamics to those observed with MOR9-1 ([Figures S4D and S4E](#); see [Experimental Procedures](#)). Dose-response curves revealed low micromolar EC_{50} s for three specific MS4A/odor pairs, similar to the EC_{50} observed for MOR9-1/vanillin ([Figure 3D](#)) and well within the range of EC_{50} s typically observed for conventional odorant receptor/ligand pairs in vitro ([Saito et al., 2009; Mainland et al., 2015](#)).

Depleting extracellular calcium abolished

Several additional ligand-receptor relationships were identified, including MS4A4B/alpha-linolenic acid (ALA), MS4A6D/oleic acid (OA), MS4A6D/arachidonic acid (AA), MS4A4D/4-pregnan-11 β ,21-diol-3,20-dione 21-sulfate, MS4A7/5-pregnan-3 α -ol-20-one sulfate, and MS4A8A/4-pregnan-11 β ,21-diol-3,20-dione,21-sulfate ([Figures 3C, S4A, and S4B](#)).

MS4A-ligand-dependent calcium transients ([Figure S4F](#)). Taken together, these results demonstrate that individual MS4A proteins enable calcium influx in response to specific monomolecular odorants in heterologous cells, with different MS4A proteins conferring responses to different ligands. The simplest explanation for this result is that the MS4A proteins are odorant receptors.



(legend on next page)

Many *Ms4a* Genes and Proteins Are Expressed in Each Necklace Sensory Neuron

The functional organization of the mammalian olfactory system depends on each mature OSN expressing just one (or two) of the hundreds of possible olfactory receptor genes (Dalton and Lomvardas, 2015). We therefore asked whether the MS4As are expressed in a one-receptor-per-neuron pattern within GC-D cells, which would suggest that the necklace system follows a similar functional logic to the main and accessory olfactory systems. Target *Ms4a* mRNA molecules within dissociated GC-D cells were labeled using a single-molecule detection approach (RNAScope), in which messages are detected as diffraction-limited fluorescent puncta whose abundance reflects transcript levels (Wang et al., 2012). RNAScope probes generated multiple puncta in individual GC-D cells for each of the 12 *Ms4a* genes found by RNA profiling, consistent with the presence of these *Ms4a* messages in GC-D cells (Figure 4A). Conversely, RNAScope failed to identify *Ms4a1*, *Ms4a2*, and *Ms4a5* puncta in GC-D cells above the background detection rate, consistent with the absence of these specific *Ms4a* genes in our earlier RNA-profiling experiments (Figures 4B and S5A).

Using a stringent criterion (in which a cell is counted positive only if it harbors two or more puncta), transcripts for each of the 12 *Ms4a* genes expressed in GC-D cells were found in 5%–30% of GC-D cells (Figure 4B). Moreover, employing a less stringent criterion for *Ms4a* positivity—in which cells with any *Ms4a* puncta are considered positive—reveals that individual *Ms4a* family members may be expressed in >50% of GC-D cells (Figure S5B). Under both of these analyses, the proportions of cells expressing each *Ms4a* message sum to significantly greater than 100%, raising the surprising possibility that each GC-D cell expresses more than one *Ms4a* gene. To directly test whether *Ms4a* genes are co-expressed, we simultaneously labeled cells in two different colors with RNAScope probes recognizing distinct *Ms4a* genes. Every tested pair revealed a significant rate of cells that were positive for more than one *Ms4a* gene, demonstrating that, unlike conventional ORs, multiple *Ms4a* genes are co-expressed in individual necklace sensory neurons (Figures 4C, S5C, and data not shown). These data are consistent with a model in which each GC-D cell expresses multiple *Ms4a* genes.

Because different probes could have distinct false-negative rates (and because *Ms4a* transcript levels are low), RNAScope could not be used to definitively determine the number of unique *Ms4a* genes expressed per necklace OSN (Kim et al., 2015). We therefore asked whether the *Ms4a* expression pattern was more

apparent at the protein level. Peptide antibodies were raised against several MS4As and were used to stain the olfactory epithelium. Consistent with every necklace cell expressing all members of the MS4A family (of the subset expressed in GC-D cells), each of five different anti-MS4A antibodies labeled >95% of GC-D neurons (Figures 5A and 5B). As expected, these antibodies did not label conventional OMP+ OSNs (Figure 5A), and control antibodies against MS4As not detected in necklace neurons by RNA profiling—MS4A1, MS4A2, and MS4A5—did not label GC-D cells (Figure 5B and data not shown). Purified anti-MS4A antibodies were specific in vivo as assessed by peptide competition and in vitro as assessed by staining HEK293T cells overexpressing individual MS4A proteins, although minor cross-reactivity was observed between pairs of highly homologous MS4A proteins (e.g., MS4A6B/MS4A6C and MS4A4B/MS4A4C) (Figures S6A and S6B).

If MS4As function as necklace chemoreceptors, they must be present at sensory endings where transduction of odorant binding occurs (Barnea et al., 2004). Consistent with this possibility, high-resolution imaging demonstrated that each MS4A antibody strongly labeled the dendritic knobs of GC-D cells, with some staining apparent in the cilia as well (Figure 5C). Sensory neurons in the main olfactory epithelium also traffic receptors to their axonal endings, which terminate in glomeruli in the olfactory bulb. To test whether MS4A proteins are similarly localized to axonal endings, olfactory bulb tissue sections were probed with anti-MS4A antibodies, which revealed that each MS4A antibody labeled every necklace glomerulus—but failed to label any conventional glomeruli—within the olfactory bulb (Figure 5D). Taken together, these data demonstrate that MS4A proteins are appropriately positioned within sensory endings to detect chemical cues in the environment and that every glomerulus in the necklace system receives input from sensory afferents potentially representing information pooled from all of the MS4As expressed in the necklace system; this pattern of organization differs sharply from that apparent in the rest of the olfactory system, where individual receptors (or pairs of receptors) define specific glomerular information channels.

Necklace Olfactory Neurons Respond to MS4A Ligands In Vivo

The expression of multiple MS4A family members in GC-D cells predicts that the chemoreceptive fields of individual necklace olfactory neurons in vivo should include the ligands identified for different MS4A proteins in vitro. To address this possibility, the intact olfactory epithelium was explanted, and functional

Figure 3. MS4A Proteins Confer Responses to Odorants

(A) GCaMP6 fluorescence in response to indicated chemical mixtures in representative HEK293 cells expressing either the indicated MS4A protein or GPCR mOR + G protein (odor delivery indicated by gray bars after accounting for line and mixing delays, see Experimental Procedures). (B) Responses of expressed MS4A protein/odor mixture pairs performed as in (A) (10 μ M per odor, see Table S3 for mixture definitions, 97 total compounds). Color code indicates percentage of cells responding ($n = 3$, total cells in experiment > 50,000) after thresholding statistically significant responses (see Experimental Procedures). Deconvolved mixture-MS4A pairs are indicated with red circles. (C) Deconvolution identifies monomolecular odors that activate each MS4A receptor. Individual odors delivered at 50 μ M in liquid phase ($n = 3$, total cells in experiment > 68,000) to cells co-expressing GCaMP6s and the indicated MS4A receptor (bottom) or GCaMP6s alone (top). The aggregate percent of cells that responded to each chemical across three independent experiments is color mapped as in (B). SFA, saturated fatty acids; UFA, unsaturated fatty acids. (D) Dose-response curves reveal low micromolar EC_{50} s for MS4A4B/ALA, MS4A6C/2,3-DMP, and MS4A6D/OA. Each data point represents the mean $\pm$ SEM from at least four independent coverslips.

See also Figures S3 and S4.

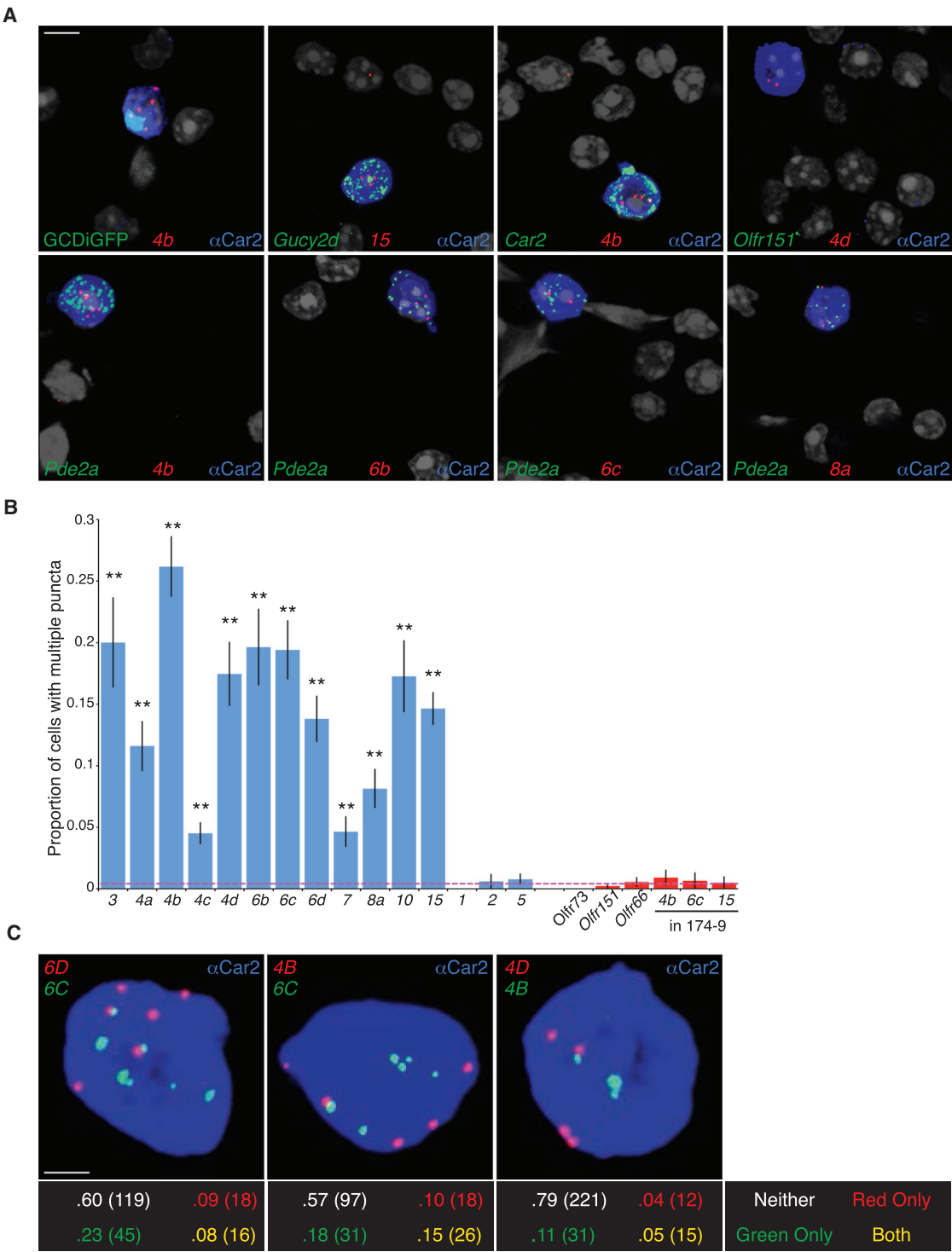


Figure 4. Multiple *Ms4a* Genes Are Expressed in Each Necklace Sensory Neuron

(A) RNAScope single-molecule fluorescent in situ hybridization of dissociated olfactory epithelial cells detects *Ms4a* family members (red). Necklace cells identified via co-labeling with an antibody against *Car2* (blue), GFP from the *Gucy2d-IRES-TauGFP* allele (green, upper-left), or an RNAScope probe against a necklace marker gene (green, all panels except the upper-left). Necklace cells are not marked by a probe against the conventional OR gene *Olfr151* (upper-right). Nuclei marked by DAPI (grayscale); cytoplasmic anti-CAR2 signal was saturated to demarcate the entire volume of each GC-D cell. Scale bar, 5 μ M.

(B) Proportion of *Car2*+ necklace OSNs with two or more detected puncta for each *Ms4a* (blue bars) and *Olfr* probe (red bars, including *Ms4a* puncta in OR174-9-IRES-GFP-expressing cells; n = 3 experiments, between 150 and 750 cells/probe, error bars are standard error of the proportion). Dashed red line represents the

(legend continued on next page)

responses were assessed by multiphoton microscopy as odorants were delivered in liquid phase. Necklace neurons were labeled with the fluorescent calcium reporter molecule GCaMP3 using the *Emx1-Cre* driver line; the *Emx1* gene was enriched in RNA-seq and NanoString analyses of GC-D neurons and distinguishes necklace cells from conventional OSNs, which express the related protein EMX2 (Figure S7A and data not shown) (Hirota and Mombaerts, 2004). Because *Emx1-Cre* labels a number of non-GC-D cells in the nasal epithelium (whose identity is unclear), we specifically imaged cul-de-sac regions, where nearly all GCaMP3-positive cells belong to the necklace (Figure S7B), and heuristically identified necklace neurons as those that responded to carbon disulfide, a known necklace ligand whose detection requires the GC-D protein (Munger et al., 2010).

Necklace neurons were activated by a mixture of the unsaturated fatty acids oleic acid and α -linolenic acid (UFAs) and by a mixture of 2,3- and 2,5-dimethylpyrazine (DMPs)—two chemical classes that stimulated MS4A proteins in vitro—but rarely by mixtures of ketones, esters, or alcohols, which were not identified as MS4A ligands (Figures 6A, 6B, and 3B). Importantly, UFAs and DMPs do not activate adjacent conventional OSNs, (i.e., those that responded to the control odor mixtures, but not to CS₂ [data not shown]); necklace cells are therefore specifically tuned to MS4A-activating compounds. These data also demonstrate that single necklace cells respond to multiple compounds that individually activate different MS4A proteins in vitro, as oleic acid and α -linolenic acid are ligands for MS4A6D and MS4A4B, respectively, and the dimethylpyrazines are ligands for MS4A6C; indeed, in separate experiments, individual necklace cells responded to multiple monomolecular MS4A ligands (Figures 6A, 6B, and 3C and data not shown). These results strongly suggest that co-expression of *Ms4a* genes confers upon each GC-D cell a tuning profile that is broader than that associated with any single MS4A protein.

Given that MS4A ligands possess a range of volatilities, we wished to directly demonstrate that these odorants can activate necklace sensory neurons in intact mice. Freely behaving mice were exposed to MS4A ligands in gas phase, and then activation of GC-D cells was measured using an antibody against phosphorylated ribosomal S6 protein, an established marker of prior OSN activity (Jiang et al., 2015). These experiments revealed that the MS4A6C ligands 2,5-DMP and 2,3-DMP, the MS4A4B ligand α -linolenic acid, and the MS4A6D ligands oleic acid and arachidonic acid reliably activated GC-D cells in vivo to a similar extent as the positive control carbon disulfide (Figure 7A). Conversely, general odorants (including acetophenone and eugenol) and compounds that only weakly activated MS4A-expressing HEK293 cells (such as 2,6-DMP) did not activate necklace cells above the background rate of plain air (Figure 7A). While prior work had implicated 2,5-DMP as a necklace ligand, 2,3-DMP,

arachidonic acid, and the fatty acids had not been previously shown to activate this olfactory subsystem. These experiments demonstrate that, in the context of active exploration, necklace sensory neurons respond to ligands for MS4A receptors.

It is not clear how soluble ligands such as urinary peptides gain access to necklace sensory neurons within the olfactory epithelium; nevertheless, one soluble class of MS4A ligand—the sulfated steroids—also activated the necklace, albeit more weakly than observed for other odorants (control DMSO = 8.7 ± 0.6 versus sulfated steroids 1,3,5(10)-estratrien-3,17 β -diol disulfate/1,3,5(10)-estratrien-3,17 α -diol 3-sulfate $12.9\% \pm 1.5\%$ cells positive, $n = 4$, $p < 0.05$ unpaired t test). It is notable that, both in explants and in vivo, the MS4A ligands and the control ligand carbon disulfide activated a fraction of the necklace sensory neurons. While these partial responses (also previously observed for urinary peptides) (Leinders-Zufall et al., 2007) may reflect specific technical features of these experiments, the consistency of this observation across ligands and preparations raises the possibility that cellular responses in the intact necklace system may be context or state dependent.

Altogether, these results support a model in which MS4A receptors bind inhaled odorants and induce calcium influx into necklace OSNs. This model raises the possibility that ectopically expressed MS4A protein will confer its in vitro chemosensitivity onto the receptive fields of conventional olfactory neurons. To test this hypothesis directly, mouse nares were irrigated with adenovirus encoding bicistronic *Ms4a6c-IRES-GFP* transcript, yielding sparse populations of GFP+ OSNs. This approach generated neurons that expressed MS4A6C protein as well as a subset of cells that were infected but did not express MS4A6C, which served as an internal control (Figure 7B); the failure of MS4A6C protein expression in some GFP-positive cells may reflect stochastic (and potentially cell-type-specific) effects related to ectopic chemosensor expression within mature OSNs (Tsai and Barnea, 2014). Ectopic MS4A6C protein was localized to dendritic endings in infected neurons, suggesting that MS4A proteins associate intrinsically with sensory structures even outside of the molecular milieu of the necklace (Figure 7B and data not shown). Quantitative analysis of neural activity (assessed by phosphorylated S6 protein levels, Figure 7C, blue channel) after exposing awake, behaving animals to odors in gas phase revealed that ectopic MS4A6C confers responses to its in vitro ligands 2,3-DMP and 2,5-DMP, but not to the control odorants eugenol and acetophenone (Figure 7C). Similarly, adenovirally mediated ectopic expression resulted in effective targeting of MS4A6D to sensory endings in conventional olfactory sensory neurons and conferred specific responses to the MS4A6D ligand oleic acid, but not to the control odorant eugenol (Figure 7C). These results demonstrate that an MS4A protein can directly impart its chemoreceptive properties to generic olfactory neurons in vivo and, furthermore, that signaling downstream of

average value of negative controls. All *Ms4a* probes (other than *Ms4a1*, *Ms4a2*, *Ms4a5*) give a significantly higher proportion of positive cells than negative controls (** $p < 0.01$, one-tailed Z test on population proportions).

(C) Representative images of Car2+ (blue) cells labeled with probes against the indicated *Ms4a* family members (top panels). The proportion and total number of cells with multiple puncta for neither, one, or both colors were quantified (bottom panels, total cell number in parentheses). Each pair shows significantly more double-positive cells (yellow) than expected ($p < 0.05$, Fisher's exact test on 2×2 table). Scale bar, 2 μ m.

See also Figure S5.

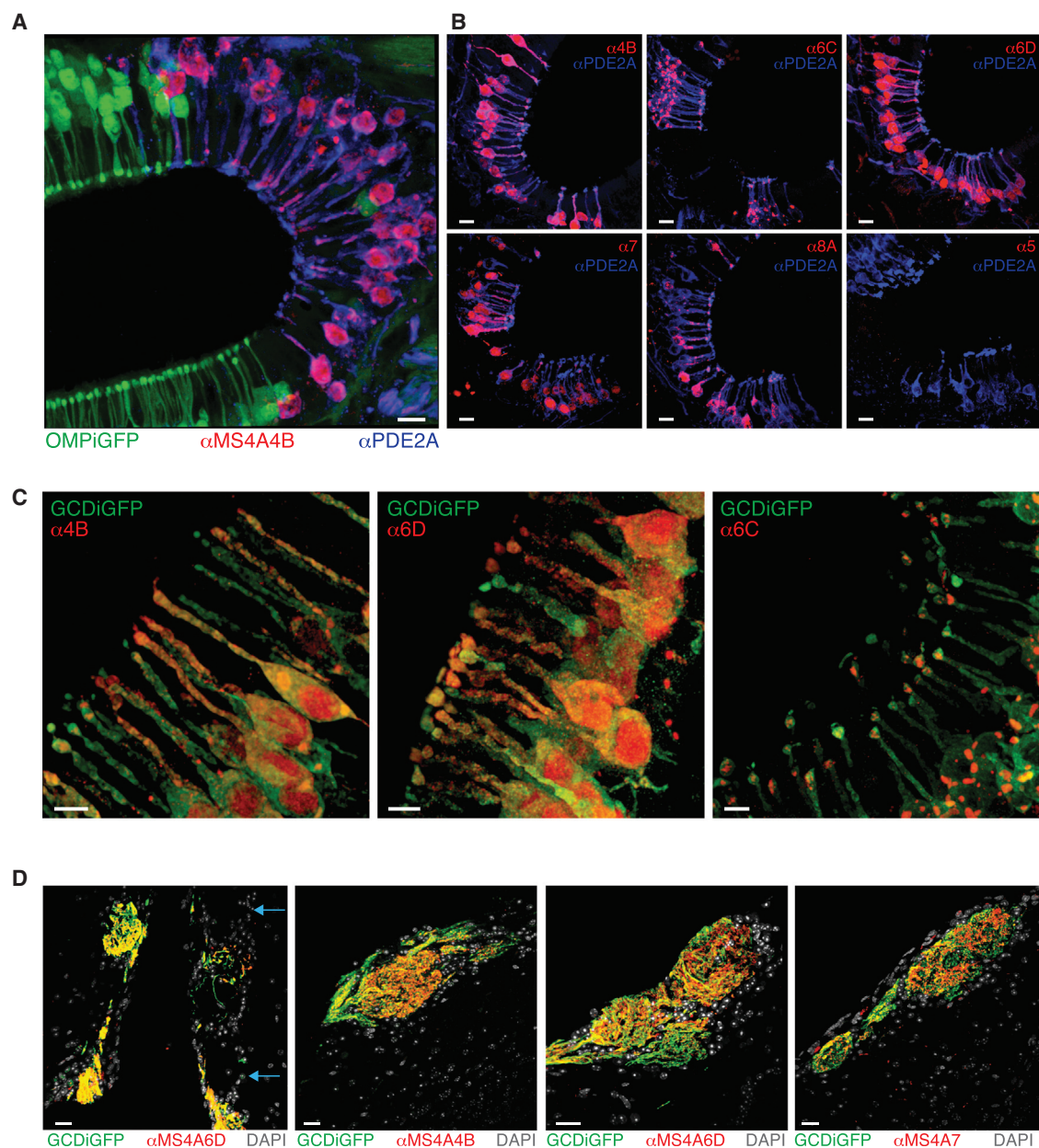


Figure 5. Multiple MS4A Proteins Are Expressed within Necklace Sensory Endings and Glomeruli

(A) Anti-MS4A4B antibody stains every anti-PDE2A+ cell, but no OMP-IRES-GFP+ cells, in sections of the olfactory epithelium. Scale bar, 10 μ M.

(B) Representative images of immunostaining with antibodies against five different MS4A family members, each of which stains >95% of anti-PDE2A+ necklace cells in epithelial cul-de-sacs; control antibody against MS4A5, which is not detected at the mRNA level in GC-D cells, does not label necklace cells (lower-right). Scale bars, 10 μ M.

(C) Anti-MS4A antibodies label dendritic knobs. Many MS4A antibodies also appear to stain perinuclear and nuclear regions; although the origin of this staining, which is eliminated by peptide competition (Figure S6B), is unknown, it may represent MS4A protein trapped within the endoplasmic reticulum or MS4A protein fragments (Cruse et al., 2013). Scale bars, 5 μ M.

(D) Anti-MS4A6D staining overlaps with all GCD-IRES-TauGFP+ necklace glomeruli in sections of the olfactory bulb (left). Blue arrows mark non-necklace glomeruli, which are not stained by anti-MS4A6D antibody. Similarly, anti-MS4A4B and anti-MS4A7 antibodies label each necklace glomerulus (right panels). Scale bars, 20 μ M.

See also Figure S6.

MS4A proteins does not require necklace-specific molecular components such as GC-D, PDE2A, or CNGA3. The concordance of necklace cell responses with the chemoreceptive

fields of MS4A proteins expressed ectopically—both in vitro and in vivo—implies that the necklace subsystem uses the *Ms4a* family to sense odors.

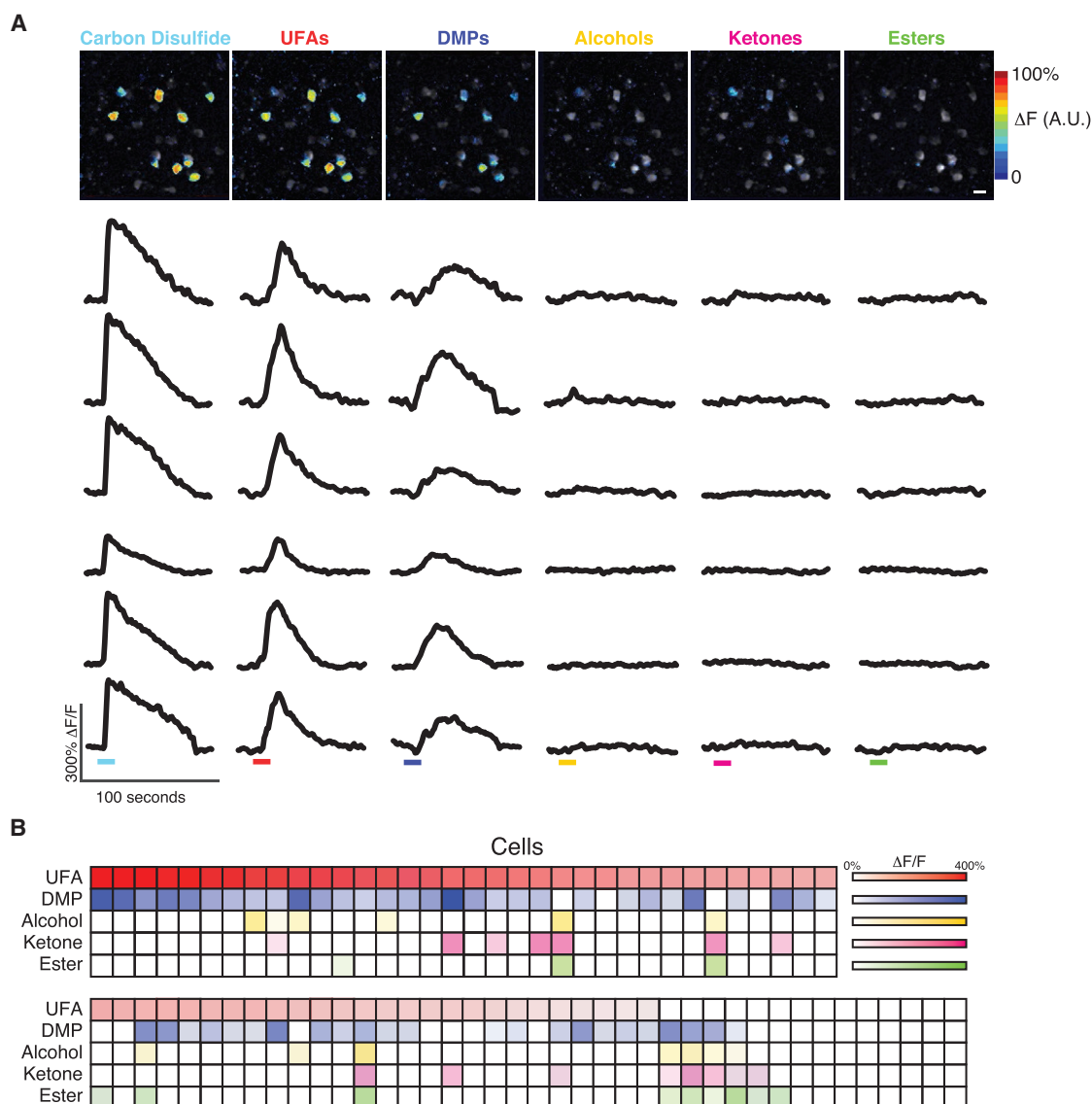


Figure 6. Multiple In Vitro MS4A Ligands Activate Single Necklace Cells In Vivo

(A) Functional responses of individual necklace cells (labeled with *Emx1-cre/GCaMP3*) to the indicated odorants were heatmapped (top row), and dF/F traces for six cells (drawn from the heatmapped image) were plotted (bottom rows). Colored bars represent the 10 s odor delivery period (after accounting for line and mixing delays, see [Experimental Procedures](#)). All mixtures were at 100 μ M total odorant concentration (DMP: 2,3-DMP and 2,5-DMP; UFA: oleic acid and α -linolenic acid, ketones, esters, and alcohols as in [Table S3](#)). Scale bar, 10 μ M.

(B) The odor-tuning properties of 74 CS₂-responsive cells (columns) across eight experiments were quantified, with color coding as in the top panel. DMPs and UFAs each activated significantly more cells (response >25% dF/F) than each negative control ($p < 0.0001$, Fisher's exact test, corrected for multiple comparisons). Significantly more cells responded to both UFAs and DMPs than were expected by chance ($p < 0.01$, Fisher's Exact Test).

See also [Figure S7](#).

DISCUSSION

Insects and mammals use similar molecular mechanisms to detect light, heat, and several gases, suggesting that solutions to common sensory problems are often conserved ([Caterina, 2007](#); [Dhaka et al., 2006](#); [Terakita, 2005](#)). However, peripheral mechanisms for odor detection differ among phyla; insects like *Drosophila melanogaster* deploy several structurally distinct ion-

otropic odorant receptor classes to interrogate the chemical world, whereas mammals were thought to detect smells exclusively through metabotropic GPCRs ([Silbering and Benton, 2010](#)). Our identification of a non-GPCR family of odorant receptor reveals an unexpected similarity between the mammalian olfactory system and that of insects: both use multiple unrelated receptor types to transduce chemosensory cues into intracellular signals.

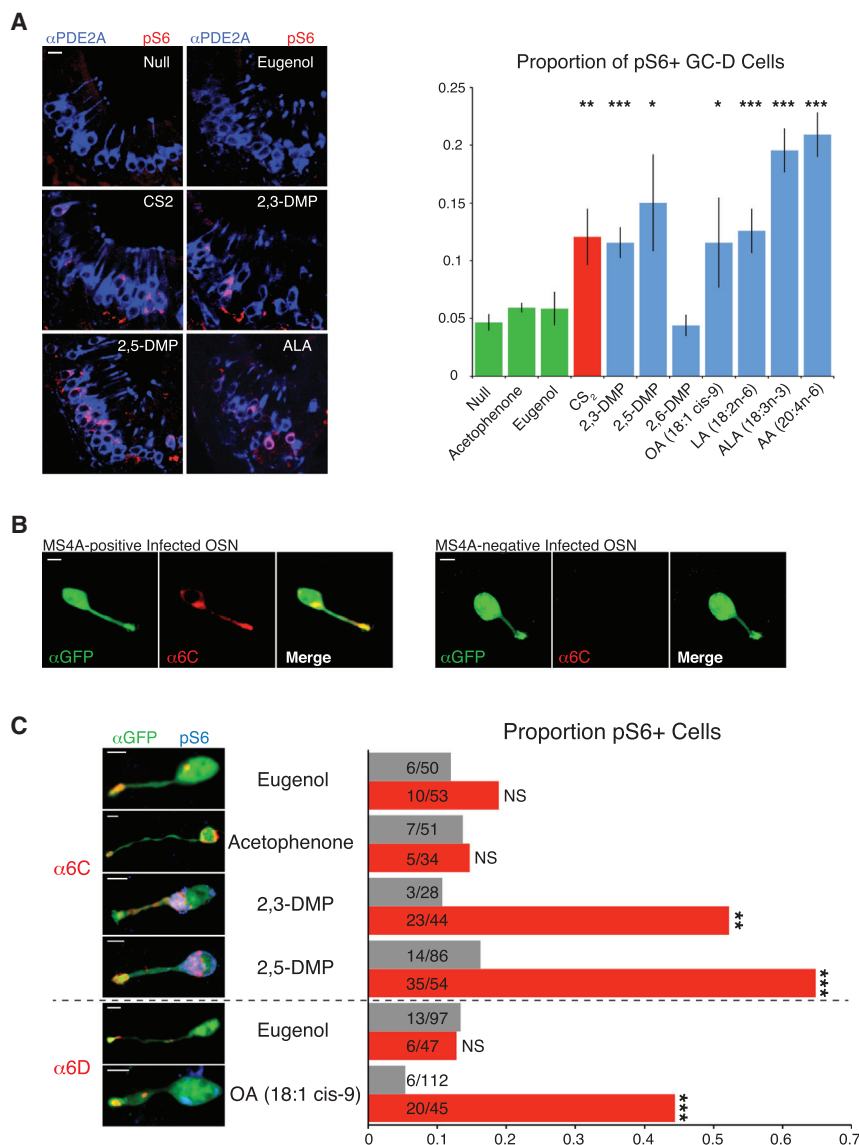


Figure 7. MS4A Ligands Activate Necklace Sensory Neurons, and MS4A Proteins Confer MS4A Ligand Responses to Conventional Olfactory Sensory Neurons in Awake, Behaving Mice

(A) Example images of cul-de-sacs from mice exposed to the indicated odorant, immunostained for the necklace cell marker PDE2A (blue) and the neuronal activity marker phospho-S6 (pSerine240/244) (red) (left panels). Quantification of the proportion of pS6+ necklace cells in odor-exposed mice (right panel, mean $\pm$ SEM, $n \geq$ three independent experiments, * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.0001$, unpaired t test compared to null exposure). Scale bar, 10 μ M.

(B) Olfactory epithelial sections of mice infected with adenovirus carrying an *Ms4a6c-IRES-Gfp* expression cassette reveal a subset of virally infected cells (green) that also express MS4A6C protein (red). Scale bars, 5 μ M.

(C) Representative images (left panels) and quantification (right panel) of phospho-S6-positive OSNs infected with *Ms4a6c-IRES-Gfp* or *Ms4a6d-IRES-Gfp*-expressing adenovirus and exposed to the indicated odorant (DMP, dimethylpyrazine; OA, oleic acid). Gray bars: GFP-positive/MS4A6C- or MS4A6D-negative cells; red bars: GFP-positive/MS4A6C or MS4A6D-positive cells ($n \geq$ three animals per odor, ** $p < 0.001$, *** $p < 0.0001$, Fisher's exact test comparing MS4A-positive to MS4A-negative cells for each odorant). Scale bars, 5 μ M.

An Alternative Logic Organizes the Necklace Olfactory System

We find that MS4A ligands are enriched for molecules with innate significance for the mouse, including the aversive pheromone 2,5-DMP and several appetitive long-chain fatty acids; interestingly, seeds and nuts, which are a major food source for mice in the wild, have high concentrations of those specific fatty acids detected by the MS4As (Sabir et al., 2012). Like the

Multiple lines of evidence indicate that *Ms4a* genes encode a novel family of chemoreceptors. Mammalian MS4A proteins are localized to the dendritic endings of olfactory sensory neurons, the site of odorant chemotransduction, and contain hypervariable regions that can potentially interact with diverse extracellular cues. Expression of MS4As in both HEK293 cells in vitro and conventional OSNs in vivo confers specific responses to odorants, indicating that MS4A proteins are sufficient to transduce the binding of extracellular ligands into intracellular signals. Furthermore, individual necklace sensory neurons, which co-express multiple MS4A proteins, each respond in vivo to multiple MS4A ligands identified in vitro. These data demonstrate that the MS4As define a new mechanism and logic for mammalian chemosensation and are likely responsible for endowing the necklace olfactory subsystem with specific sensory odor tuning properties.

MS4A ligands, many of the molecules previously shown to trigger activity within the necklace olfactory system also have innate meaning for mice; these include carbon dioxide, which mice robustly avoid, and carbon disulfide and the peptides guanylin and uroguanylin, each of which promotes the social transmission of food preferences (STFP) between mice (Arakawa et al., 2013; Hu et al., 2007; Munger et al., 2010).

The tuning properties of the necklace system appears to reflect the co-expression of genes encoding multiple MS4A receptors within individual necklace sensory neurons. This pattern of odor receptor gene expression stands in stark contrast to the canonical one-receptor-per-neuron rule that (to a first approximation) organizes the remainder of the mammalian and the entirety of the *Drosophila* olfactory systems. In those systems, each OR is associated with one or a small number of glomeruli in the brain; in the necklace system, each sensory neuron and glomerulus is, in principle, capable

of conveying odor information detected by all of the MS4A receptors expressed in the nose.

The ability of individual necklace neurons to detect multiple innately relevant signals (including gases, peptides, and volatile odors) of widely differing valences suggests two broad models for the perceptual function of the necklace subsystem. First, the necklace could discriminate between odors, albeit through distinct mechanisms from the conventional olfactory system. Odor-evoked differences in activity in individual necklace glomeruli could be generated by subtle differences in MS4A expression within subsets of sensory neurons (but see Figure 5); such differences could also be caused by conventional OSN axons, which may co-mingle with necklace sensory axons in necklace glomeruli (Cockerham et al., 2009; Secundo et al., 2014). Stimulus discrimination could alternatively be achieved through linear or non-linear interactions between intracellular signals downstream of individual MS4A receptors, which could allow the integration or gating of signals arising from distinct odorants and gases detected by single necklace neurons (van Giesen et al., 2016); this process could generate a range of firing rates (depending, for example, on the specific odor components present in a blend) that could then be differentially read-out by the brain.

Second, the main function of the necklace system could be odor detection rather than discrimination. In this model, the necklace system may be organized in a similar manner to the mammalian bitter taste and *C. elegans* olfactory systems, which co-express receptors with widely divergent receptive fields in sensory neurons coupled to circuits mediating stereotypical behavioral outputs, such as attraction or aversion (Adler et al., 2000; Troemel et al., 1995). In the case of the necklace system, this singular behavioral output could be aversion, as many of the MS4A ligands have been shown to be contextually aversive, including (despite their nutritional value) fatty acids at high concentrations (Galindo et al., 2012). Alternatively, the necklace could act as an alert system, signaling the presence of salient cues in the environment (whose identity would be disentangled by other sensory systems); such a notification system could be useful for modulating the internal state of the animal in response to particularly relevant external cues, thereby facilitating both innate responses to odors and odor-related learning processes (potentially including STFP).

In either the discrimination or detection model, the physiological function of the necklace could depend upon interactions between intracellular signals downstream of the MS4A receptors and GC-D. Although the MS4As do not obligately require GC-D to promote calcium influx, GC-D enzymatic activity is sensitive to calcium levels in vitro (Duda and Sharma, 2008). This observation raises the possibility that the MS4A and GC-D pathways intersect in some manner, perhaps facilitating coincidence detection between gases and peptides on the one hand and volatile odors on the other.

Dissecting MS4A Function

Previous work has suggested that the MS4As may facilitate increases in intracellular calcium by acting as co-receptors for associated ligand-binding proteins such as the B-cell receptors (Dombrowicz et al., 1998; Eon Kuek et al., 2015). Here, we show that small molecule ligands cause an MS4A-dependent influx of

extracellular calcium, demonstrating that the MS4A molecules themselves have a receptor function. It is not clear, however, whether the MS4A proteins themselves form a calcium-permeable channel (similar to the *Drosophila* IR odorant receptors) or whether the MS4As act as ligand-binding co-receptors for an ion channel that is expressed in many cell types (Abuin et al., 2011; Benton et al., 2009).

The underlying molecular mechanisms through which the MS4A proteins interact with ligands are not known. The observation that the molecular diversity within the MS4A family is largely found within extracellular domains—rather than transmembrane domains—suggests that the MS4As interact with odor ligands through biophysical mechanisms that are distinct from those used by conventional ORs. These mechanisms may be similar to those used by mammalian bitter taste receptors, whose diversity is also found largely in extracellular domains (Hayakawa et al., 2014; Wooding, 2011). It is also unclear whether the MS4As, despite being co-expressed in single cells, primarily interact with odorants as homomers (like the bitter taste receptors) or whether they heteromerize in a manner that alters their tuning properties (Howie et al., 2009).

The Four-Pass Transmembrane MS4As: Chemical Detectors across Cell Types and Species?

Although a number of cellular roles have been suggested for individual MS4A proteins (largely in the context of the immune system), no clear picture has emerged of the core function of the MS4A family across cell types. The finding that multiple MS4As encode olfactory receptors suggests that they act as chemosensors in a range of tissues; this role may be both exteroceptive, serving to detect small molecules from the outside world, and interoceptive, as several of the ligands for MS4A proteins (e.g., oleic acid and arachidonic acid) are used as signaling molecules physiologically. Consistent with this hypothesis, members of the MS4A6/7 subfamily are expressed in microglia, neuroimmune cells responsible for sensing and responding to a variety of endogenous protein and lipid chemosignals in the mammalian brain; human MS4A8B and MS4A12 are expressed in epithelial cells at the luminal surface of the small and large intestine, tissues that both sense dietary lipids; human MS4A8B is localized specifically in chemosensory cilia in lung cells tasked with probing the sensory environment; and MS4A5 is expressed in mammalian spermatocytes, which chemotax to oocytes (Eon Kuek et al., 2015; Koslowski et al., 2008).

MS4A homologs are present in all mammalian lineages and in many sequenced deuterostomes, including osteichthyes (Figure S2C and data not shown), implying that the evolution of the *Ms4a* genes antedates the advent of the mammalian receptors for taste and for pheromones (Grus and Zhang, 2009). Taken with the finding that the MS4As detect a variety of innately relevant cues, these observations invite the speculation that the MS4A molecules represent an ancient mechanism for sensing ethologically salient small molecules in the environment. Testing this hypothesis will require establishing chemosensory roles for *Ms4a* genes in other species and better definition of those natural ligands that optimally activate the MS4As (given the limits of the synthetic odor panel explored here). It is important to note that all MS4A ligands thus far identified are also detected by

other receptor molecules in the smell and taste systems (Isogai et al., 2011; Mamasuew et al., 2011; Oberland et al., 2015). Nevertheless, the maintenance of the MS4A receptor repertoire for more than 400 million years, especially given the evolutionary success of G-protein-coupled odor receptors, argues that MS4As play an important—and non-redundant—role in sensory physiology (Zuccolo et al., 2010).

Although the *Ms4a* genes are conserved across vertebrates, *Gucy2d* is itself pseudogenized in most primates (Young et al., 2007). It is unclear whether the absence of GC-D reflects a disappearance of the necklace system in primates or merely that GC-D became unnecessary for the tasks required of the primate necklace, given the persistence of the *Ms4a* genes (Young et al., 2007). Similarly, the vomeronasal organ, the sensory epithelium thought to mediate the majority of pheromone responses in rodents, became vestigial ~25 million years ago (Zhang and Webb, 2003). Future investigation of the functional role of *Ms4a* gene families across chordates—and the relevance of interactions between MS4A proteins and ethologically relevant cues like pheromones and fatty acids—will reveal both common and species-specific roles of the MS4As in processing information from the chemical environment.

EXPERIMENTAL PROCEDURES

Graphical representations of the data are presented as the mean $\pm$ SEM unless otherwise noted. All mice were obtained from the Jackson Laboratory, with the exception of *OR174-9-IRES-tauGFP*, which was obtained from the Axel lab. For deep sequencing, olfactory epithelia were dissociated using papain and individual cells were FAC sorted; RNA was then isolated using Trizol (Invitrogen) before SmartSeq2-based amplification and Illumina-based sequencing. NanoString-based RNA quantitation was performed using a modified protocol to optimize for recovery of low-abundance transcripts. MS4A sequences were extracted from the Ensembl and NCBI databases. MS4A proteins were expressed in HEK293 cells using a Tet-regulatable system to control protein expression (Invitrogen); odor delivery was either achieved via bulk exchange or through a focal stimulus pencil placed over specific fields of view, and odor delays were determined using Rhodamine B dye as a surrogate. Odor responses were imaged in both configurations using an Andor Neo sCMOS camera mounted on an Olympus IX83. Single-molecule in situ RNA detection was performed using RNAScope (Advanced Cell Diagnostics); this protocol was adapted to simultaneously immunostain with either anti-CAR2 or anti-GFP antibodies. Custom antibodies against the MS4A proteins were developed in rabbits or guinea pigs (Covance) and were purified using protein A-sepharose and bead-conjugated peptides. Multiphoton-based functional imaging (Prairie Technologies) of explanted olfactory epithelia was performed in superfused carboxygenated modified Ringer's solution, with odor delivery in liquid phase controlled by a solenoid valve system (Warner Instruments); image alignment was achieved using custom feature-tracking and homography algorithms. Human adenoviruses were generated by the UNC Viral Core facility. See [Supplemental Experimental Procedures](#) for details regarding experimental methods, reagent sources, and statistical procedures.

ACCESSION NUMBERS

The RNA-seq data reported in this paper have been deposited in the SRA database at NCBI under the accession number SRP074687.

SUPPLEMENTAL INFORMATION

[Supplemental Information](#) includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.05.001>.

AUTHOR CONTRIBUTIONS

P.L.G., D.M.B., and S.R.D. conceived and designed the experiments. P.L.G., D.M.B., M.L.B., T.T., S.L.P., and A.C.N. carried out all of the wet-lab experiments; evolutionary analysis was performed by J.-M.L., F.K.M., and H.E.H. R.K. performed RNA-seq read quality control and alignment.

ACKNOWLEDGMENTS

We thank Vanessa Ruta, Michael Greenberg, Richard Axel, Steve Liberles, Bernardo Sabatini, Rachel Wilson, Elizabeth Hong, Steve Flavell, Eric Griffith, and members of the Datta Lab for helpful discussions and comments on the manuscript and Allison Petrosino, Neha Bhagat, and Lauren Coritt for laboratory assistance. We thank Andrew Giessel, Alexander Wiltschko, and Jeffrey Markowitz for coding assistance and Hannah Somhegyi for generation of the graphical abstract. P.L.G. was supported by the Nancy Lurie Marks Foundation, the William Randolph Hearst Fund, the Leonard and Isabelle Goldenson Fund, and the Edward and Anne Leffer Foundation. D.M.B. is supported by a NSF predoctoral fellowship and a Sackler Scholarship in Psychobiology. J.-M.L. is supported by the European Molecular Biology Organization (ALTF 379-2011), the Human Frontiers Science Program (LT001086/2012), and the Belgian American Educational Foundation. M.L.B. is supported by a NSF predoctoral fellowship. T.T. is supported by a postdoctoral fellowship from the Alice and Joseph E. Brooks Foundation, the Astellas Foundation for Research on Metabolic Disorders, and the Japan Society for the Promotion of Science postdoctoral fellowship for Research Abroad. S.L.P. was supported by a NIDA F31 (DA036922-01) and is a Stuart H.Q. & Victoria Quan Fellow at Harvard Medical School. H.H. is supported by the Howard Hughes Medical Institute. S.R.D. is supported by fellowships from the Burroughs Wellcome Fund, the Searle Foundation, the Sloan Foundation, the Vallee Foundation, the McKnight Foundation, the Khodadad Program, by grants DP2OD007109 and RO11DC011558 from the National Institutes of Health, and by the Global Brain Initiative from the Simons Foundation.

Received: December 3, 2015

Revised: March 14, 2016

Accepted: April 28, 2016

Published: May 26, 2016

REFERENCES

- Abuin, L., Bargeton, B., Ulbrich, M.H., Isacoff, E.Y., Kellenberger, S., and Benton, R. (2011). Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* 69, 44–60.
- Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., and Zuker, C.S. (2000). A novel family of mammalian taste receptors. *Cell* 100, 693–702.
- Amcheslavsky, A., Wood, M.L., Yeromin, A.V., Parker, I., Freites, J.A., Tobias, D.J., and Cahalan, M.D. (2015). Molecular biophysics of Orai store-operated Ca²⁺ channels. *Biophys. J.* 108, 237–246.
- Arakawa, H., Kelliher, K.R., Zufall, F., and Munger, S.D. (2013). The receptor guanylyl cyclase type D (GC-D) ligand uroguanylin promotes the acquisition of food preferences in mice. *Chem. Senses* 38, 391–397.
- Axel, R. (1995). The molecular logic of smell. *Sci. Am.* 273, 154–159.
- Barnea, G., O'Donnell, S., Mancina, F., Sun, X., Nemes, A., Mendelsohn, M., and Axel, R. (2004). Odorant receptors on axon termini in the brain. *Science* 304, 1468.
- Benton, R., Vannice, K.S., Gomez-Diaz, C., and Vosshall, L.B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136, 149–162.
- Bubien, J.K., Zhou, L.J., Bell, P.D., Frizzell, R.A., and Tedder, T.F. (1993). Transfection of the CD20 cell surface molecule into ectopic cell types generates a Ca²⁺ conductance found constitutively in B lymphocytes. *J. Cell Biol.* 121, 1121–1132.
- Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65, 175–187.

- Caterina, M.J. (2007). Transient receptor potential ion channels as participants in thermosensation and thermoregulation. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 292, R64–R76.
- Cockerham, R.E., Puche, A.C., and Munger, S.D. (2009). Heterogeneous sensory innervation and extensive intrabulbar connections of olfactory necklace glomeruli. *PLoS ONE* 4, e4657.
- Cruse, G., Beaven, M.A., Ashmole, I., Bradding, P., Gilfillan, A.M., and Metcalfe, D.D. (2013). A truncated splice-variant of the FcεRIβ receptor subunit is critical for microtubule formation and degranulation in mast cells. *Immunity* 38, 906–917.
- Dalton, R.P., and Lomvardas, S. (2015). Chemosensory receptor specificity and regulation. *Annu. Rev. Neurosci.* 38, 331–349.
- Dhaka, A., Viswanath, V., and Patapoutian, A. (2006). Trp ion channels and temperature sensation. *Annu. Rev. Neurosci.* 29, 135–161.
- Dombrowicz, D., Lin, S., Flamand, V., Brini, A.T., Koller, B.H., and Kinet, J.P. (1998). Allergy-associated Fcβeta is a molecular amplifier of IgE- and IgG-mediated in vivo responses. *Immunity* 8, 517–529.
- Duda, T., and Sharma, R.K. (2008). ONE-GC membrane guanylate cyclase, a trimodal odorant signal transducer. *Biochem. Biophys. Res. Commun.* 367, 440–445.
- Dulac, C., and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83, 195–206.
- Eon Kuek, L., Leffler, M., Mackay, G.A., and Hulett, M.D. (2015). The MS4A family: counting past 1, 2 and 3. *Immunol. Cell Biol.* 94, 11–23.
- Fülle, H.J., Vassar, R., Foster, D.C., Yang, R.B., Axel, R., and Garbers, D.L. (1995). A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons. *Proc. Natl. Acad. Sci. USA* 92, 3571–3575.
- Galindo, M.M., Voigt, N., Stein, J., van Lengerich, J., Raguse, J.D., Hofmann, T., Meyerhof, W., and Behrens, M. (2012). G protein-coupled receptors in human fat taste perception. *Chem. Senses* 37, 123–139.
- Gao, L., Hu, J., Zhong, C., and Luo, M. (2010). Integration of CO₂ and odorant signals in the mouse olfactory bulb. *Neuroscience* 170, 881–892.
- Grus, W.E., and Zhang, J. (2009). Origin of the genetic components of the vomeronasal system in the common ancestor of all extant vertebrates. *Mol. Biol. Evol.* 26, 407–419.
- Guo, D., Zhang, J.J., and Huang, X.Y. (2009). Stimulation of guanylyl cyclase-D by bicarbonate. *Biochemistry* 48, 4417–4422.
- Hayakawa, T., Suzuki-Hashido, N., Matsui, A., and Go, Y. (2014). Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. *Mol. Biol. Evol.* 31, 2018–2031.
- Herrada, G., and Dulac, C. (1997). A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* 90, 763–773.
- Hirota, J., and Mombaerts, P. (2004). The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proc. Natl. Acad. Sci. USA* 101, 8751–8755.
- Howie, D., Nolan, K.F., Daley, S., Butterfield, E., Adams, E., Garcia-Rueda, H., Thompson, C., Saunders, N.J., Cobbold, S.P., Tone, Y., et al. (2009). MS4A4B is a GPCR-associated membrane adapter, expressed by regulatory T cells, which modulates T cell activation. *J. Immunol.* 183, 4197–4204.
- Hu, J., Zhong, C., Ding, C., Chi, Q., Walz, A., Mombaerts, P., Matsunami, H., and Luo, M. (2007). Detection of near-atmospheric concentrations of CO₂ by an olfactory subsystem in the mouse. *Science* 317, 953–957.
- Ihara, S., Yoshikawa, K., and Touhara, K. (2013). Chemosensory signals and their receptors in the olfactory neural system. *Neuroscience* 254, 45–60.
- Isogai, Y., Si, S., Pont-Lezica, L., Tan, T., Kapoor, V., Murthy, V.N., and Dulac, C. (2011). Molecular organization of vomeronasal chemoreception. *Nature* 478, 241–245.
- Jiang, Y., Gong, N.N., Hu, X.S., Ni, M.J., Pasi, R., and Matsunami, H. (2015). Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat. Neurosci.* 18, 1446–1454.
- Juif, D.M., Fülle, H.J., Zhao, A.Z., Houslay, M.D., Garbers, D.L., and Beavo, J.A. (1997). A subset of olfactory neurons that selectively express cGMP-stimulated phosphodiesterase (PDE2) and guanylyl cyclase-D define a unique olfactory signal transduction pathway. *Proc. Natl. Acad. Sci. USA* 94, 3388–3395.
- Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the probability of mouse odorant receptor gene choice. *Cell* 147, 907–921.
- Kim, S.T., Do, I.G., Lee, J., Sohn, I., Kim, K.M., and Kang, W.K. (2015). The NanoString-based multigene assay as a novel platform to screen EGFR, HER2, and MET in patients with advanced gastric cancer. *Clin. Transl. Oncol.* 17, 462–468.
- Koslowski, M., Sahin, U., Dhaene, K., Huber, C., and Türeci, O. (2008). MS4A12 is a colon-selective store-operated calcium channel promoting malignant cell processes. *Cancer Res.* 68, 3458–3466.
- Leinders-Zufall, T., Cockerham, R.E., Michalakis, S., Biel, M., Garbers, D.L., Reed, R.R., Zufall, F., and Munger, S.D. (2007). Contribution of the receptor guanylyl cyclase GC-D to chemosensory function in the olfactory epithelium. *Proc. Natl. Acad. Sci. USA* 104, 14507–14512.
- Liberles, S.D., and Buck, L.B. (2006). A second class of chemosensory receptors in the olfactory epithelium. *Nature* 442, 645–650.
- Liberles, S.D., Horowitz, L.F., Kuang, D., Contos, J.J., Wilson, K.L., Siltberg-Liberles, J., Liberles, D.A., and Buck, L.B. (2009). Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc. Natl. Acad. Sci. USA* 106, 9842–9847.
- Lin, S., Cicala, C., Scharenberg, A.M., and Kinet, J.P. (1996). The Fc(epsilon)RIβ subunit functions as an amplifier of Fc(epsilon)RIγ-mediated cell activation signals. *Cell* 85, 985–995.
- Mainland, J.D., Li, Y.R., Zhou, T., Liu, W.L., and Matsunami, H. (2015). Human olfactory receptor responses to odorants. *Sci. Data* 2, 150002.
- Mamasuew, K., Hofmann, N., Breer, H., and Fleischer, J. (2011). Grueneberg ganglion neurons are activated by a defined set of odorants. *Chem. Senses* 36, 271–282.
- Man, O., Gilad, Y., and Lancet, D. (2004). Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci.* 13, 240–254.
- Martini, S., Silvotti, L., Shirazi, A., Ryba, N.J., and Tirindelli, R. (2001). Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ. *J. Neurosci.* 21, 843–848.
- Matsunami, H., and Buck, L.B. (1997). A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* 90, 775–784.
- Meyer, M.R., Angele, A., Kremmer, E., Kaupp, U.B., and Müller, F. (2000). A cGMP-signaling pathway in a subset of olfactory sensory neurons. *Proc. Natl. Acad. Sci. USA* 97, 10595–10600.
- Mori, K., and Sakano, H. (2011). How is the olfactory map formed and interpreted in the mammalian brain? *Annu. Rev. Neurosci.* 34, 467–499.
- Munger, S.D., Leinders-Zufall, T., and Zufall, F. (2009). Subsystem organization of the mammalian sense of smell. *Annu. Rev. Physiol.* 71, 115–140.
- Munger, S.D., Leinders-Zufall, T., McDougall, L.M., Cockerham, R.E., Schmid, A., Wandernoth, P., Wennemuth, G., Biel, M., Zufall, F., and Kelliher, K.R. (2010). An olfactory subsystem that detects carbon disulfide and mediates food-related social learning. *Curr. Biol.* 20, 1438–1444.
- Nei, M., Niimura, Y., and Nozawa, M. (2008). The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat. Rev. Genet.* 9, 951–963.
- Oberland, S., Ackels, T., Gaab, S., Pelz, T., Spehr, J., Spehr, M., and Neuhaus, E.M. (2015). CD36 is involved in oleic acid detection by the murine olfactory system. *Front. Cell. Neurosci.* 9, 366.
- Polyak, M.J., Li, H., Shariat, N., and Deans, J.P. (2008). CD20 homo-oligomers physically associate with the B cell antigen receptor. Dissociation upon receptor engagement and recruitment of phosphoproteins and calmodulin-binding proteins. *J. Biol. Chem.* 283, 18545–18552.

- Rivière, S., Challet, L., Flügge, D., Spehr, M., and Rodriguez, I. (2009). Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* 459, 574–577.
- Ryba, N.J., and Tirindelli, R. (1997). A new multigene family of putative pheromone receptors. *Neuron* 19, 371–379.
- Sabir, A., Unver, A., and Kara, Z. (2012). The fatty acid and tocopherol constituents of the seed oil extracted from 21 grape varieties (*Vitis* spp.). *J. Sci. Food Agric.* 92, 1982–1987.
- Saito, H., Chi, Q., Zhuang, H., Matsunami, H., and Mainland, J.D. (2009). Odor coding by a Mammalian receptor repertoire. *Sci. Signal.* 2, ra9.
- Secundo, L., Snitz, K., and Sobel, N. (2014). The perceptual logic of smell. *Curr. Opin. Neurobiol.* 25, 107–115.
- Shinoda, K., Shiotani, Y., and Osawa, Y. (1989). “Necklace olfactory glomeruli” form unique components of the rat primary olfactory system. *J. Comp. Neurol.* 284, 362–373.
- Silbering, A.F., and Benton, R. (2010). Ionotropic and metabotropic mechanisms in chemoreception: ‘chance or design’? *EMBO Rep.* 11, 173–179.
- Singer, M.S. (2000). Analysis of the molecular basis for octanal interactions in the expressed rat 17 olfactory receptor. *Chem. Senses* 25, 155–165.
- Sun, L., Wang, H., Hu, J., Han, J., Matsunami, H., and Luo, M. (2009). Guanylyl cyclase-D in the olfactory CO2 neurons is activated by bicarbonate. *Proc. Natl. Acad. Sci. USA* 106, 2041–2046.
- Terakita, A. (2005). The opsins. *Genome Biol.* 6, 213.
- Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., and Bargmann, C.I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* 83, 207–218.
- Tsai, L., and Barnea, G. (2014). A critical period defined by axon-targeting mechanisms in the murine olfactory bulb. *Science* 344, 197–200.
- van Giesen, L., Hernandez-Nunez, L., Delasoie-Baranek, S., Colombo, M., Renaud, P., Bruggmann, R., Benton, R., Samuel, A.D., and Sprecher, S.G. (2016). Multimodal stimulus coding by a gustatory sensory neuron in *Drosophila* larvae. *Nat. Commun.* 7, 10687.
- Wang, F., Flanagan, J., Su, N., Wang, L.C., Bui, S., Nielson, A., Wu, X., Vo, H.T., Ma, X.J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* 14, 22–29.
- Wooding, S. (2011). Signatures of natural selection in a primate bitter taste receptor. *J. Mol. Evol.* 73, 257–265.
- Young, J.M., Waters, H., Dong, C., Fülle, H.J., and Liman, E.R. (2007). Degeneration of the olfactory guanylyl cyclase D gene during primate evolution. *PLoS ONE* 2, e884.
- Zhang, J., and Webb, D.M. (2003). Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl. Acad. Sci. USA* 100, 8337–8341.
- Zuccolo, J., Bau, J., Childs, S.J., Goss, G.G., Sensen, C.W., and Deans, J.P. (2010). Phylogenetic analysis of the MS4A and TMEM176 gene families. *PLoS ONE* 5, e9369.

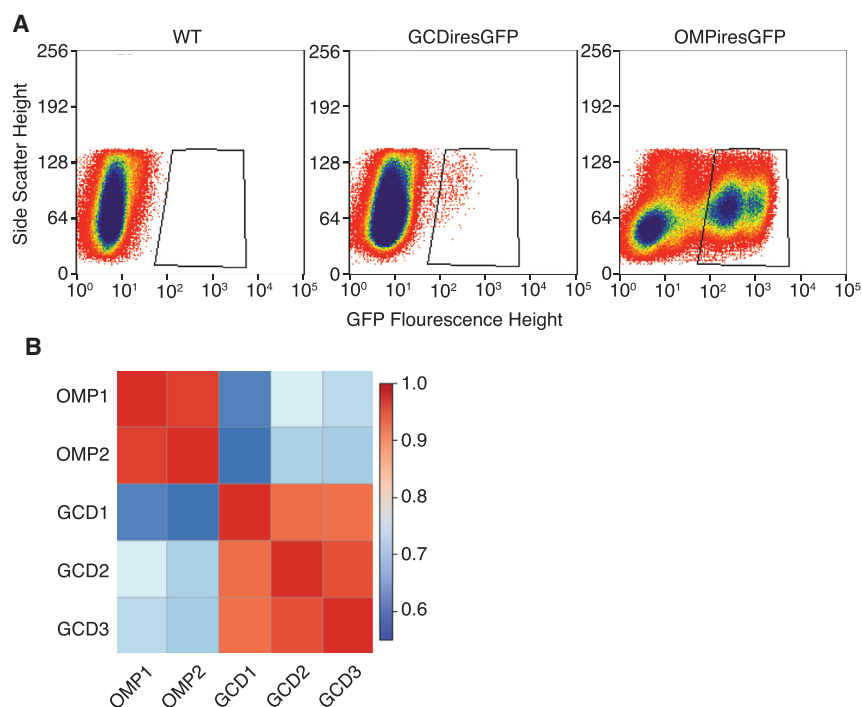
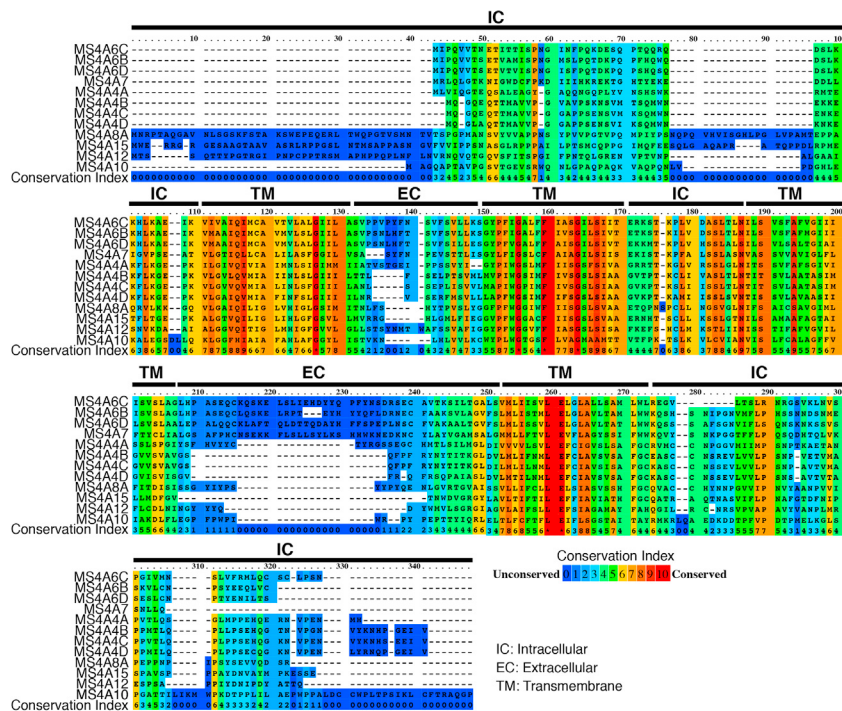
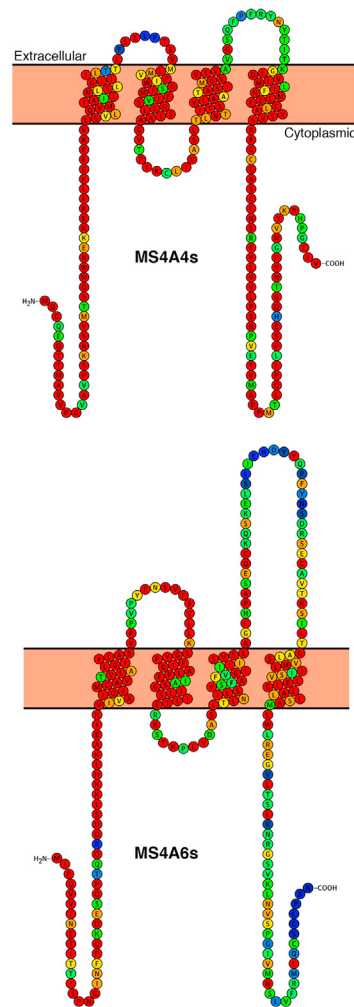
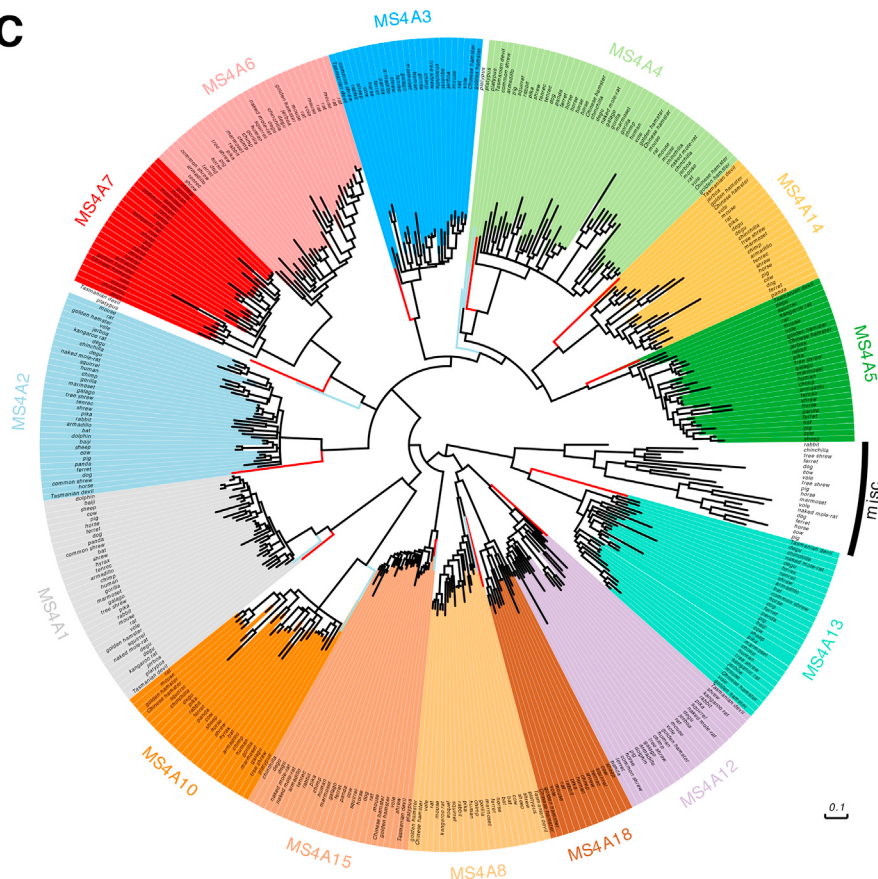


Figure S1. Related to Figure 1

(A) Scatter plots of FAC sorted, dissociated olfactory epithelial cells from wild type mice (left), mice harboring the *Gucy2d-IRES-TauGFP* allele (middle), or mice expressing the *Omp-IRES-GFP* allele (right). The gate used to isolate ~100% pure populations of fluorescent necklace OSNs or canonical OSNs is indicated. (B) Heat map of the correlation of gene expression between RNA-seq samples, with warmer colors corresponding to more highly correlated gene expression.

A**B****C**

(legend on next page)

Figure S2. Related to Figure 2

(A) Multiple sequence alignment of the mouse MS4A proteins expressed in GC-D cells. Residues that are more conserved are shown in warmer colors, whereas residues that are less conserved are depicted in colder colors (conservation scores were determined using PRALINE, see methods). The intracellular (IC), extracellular (EC), and transmembrane (TM) regions of the proteins are indicated, and reveal the greatest sequence diversity in the extracellular domains (with additional diversity in the intracellular C-terminus).

(B) Multispecies alignments of MS4A proteins from either the MS4A4 or MS4A6 subfamilies. Amino acid conservation was determined and heat-mapped as in A. As with alignments of all GC-D-expressed MS4As, the extracellular domains within these subfamilies are more diverse than other regions of the MS4As.

(C) A phylogenetic tree of the mammalian *Ms4a* gene family was generated using every *Ms4a* gene found in 37 representative taxa, which were selected to cover all major mammalian lineages (Supplemental Table 1). Every *Ms4a* gene was assigned to an MS4A subfamily (see methods), and each subfamily is represented with a unique color to facilitate visualization within the circular phylogenetic tree. Each *Ms4a* gene is represented as a line within this plot where the length of the line corresponds to the degree of evolutionary change within a lineage over time (the scale bar represents the number of substitutions per site). The *Ms4a* gene family cluster diversified through tandem duplication early in the evolution of mammals as illustrated by the presence of 10 homologs in the monotreme (platypus, light blue lines) and marsupial (Tasmanian devil, red lines) representatives, which contrasts the single copy of MS4A15 found in bird genomes (Zuccolo et al., 2010). Further extension of the family occurred during the evolution of placental mammals, with human and mouse genomes harboring 16 and 17 genes, respectively. The majority of MS4A subfamilies exhibit one-to-one orthologous pairs across species. By contrast, the MS4A4 and MS4A6 subfamilies, which are highly enriched in GC-D neurons, demonstrate complex one-to-many and many-to-many paralogous relationships between species. It is noteworthy that 50% of the genes present in the bovid representatives are either lost or pseudogenized in cetacean lineages suggesting rapid gene turnover throughout evolution.

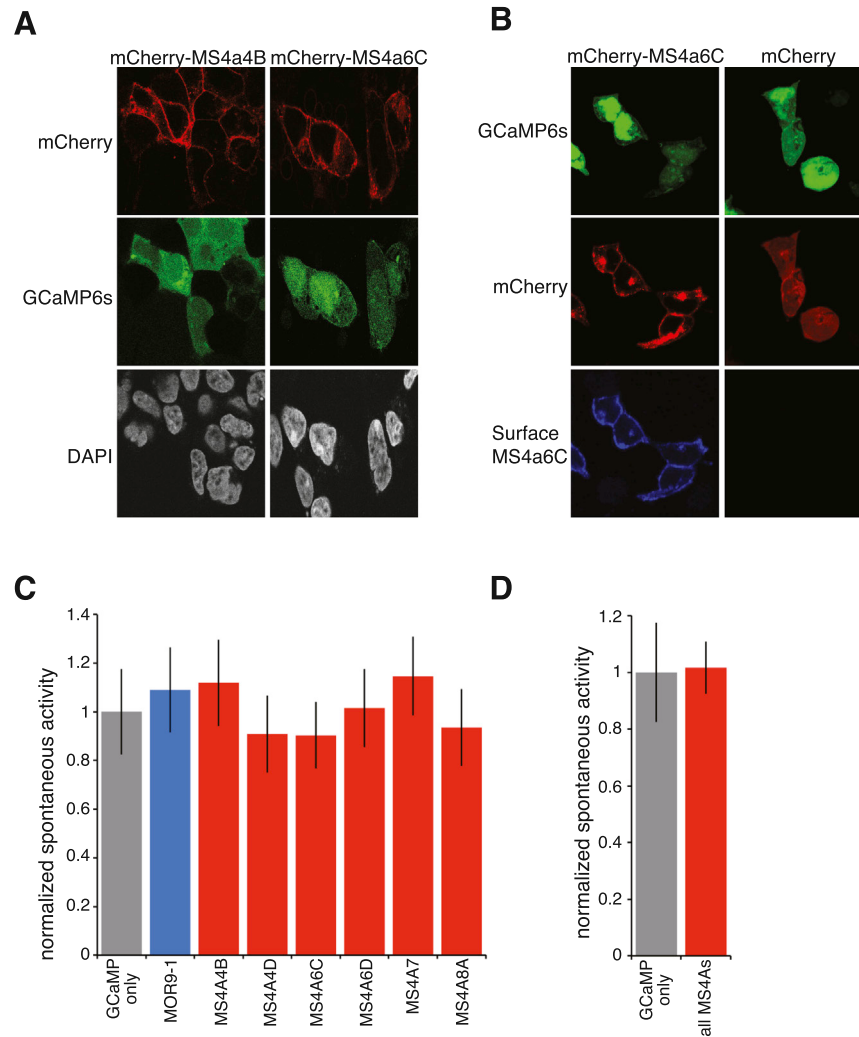


Figure S3. Related to Figure 3

(A) Representative confocal images of HEK293 cells transfected with plasmids encoding GCaMP6s (green) and N-terminal mCherry-fusion proteins of the indicated MS4A protein (red), revealing the presence of mCherry-MS4A fusions at the plasma membrane.

(B) HEK293 cells transfected with GCaMP6s (green) and either mCherry alone or mCherry-MS4A6C (red) were immunostained under non-permeabilizing conditions with an extracellularly-directed anti-MS4A6C antibody, revealing specific labeling of MS4A6C proteins (blue) indicating that MS4A6C is efficiently trafficked to the plasma membrane and adopts the predicted topology.

(C) The spontaneous activity rate of HEK293 cells expressing MOR9-1 or the indicated MS4A was calculated and then normalized to control cells that expressed GCaMP6s alone (see methods). Data are presented as mean $\pm$ SEM from 24 independent coverslips per condition. The normalized spontaneous rate of 1 for GCaMP corresponds to a response percentage of 1.4%. No significant differences were observed amongst conditions (unpaired Student's T-test).

(D) As in C except HEK293 cells expressing any MS4A protein are considered as a single group.

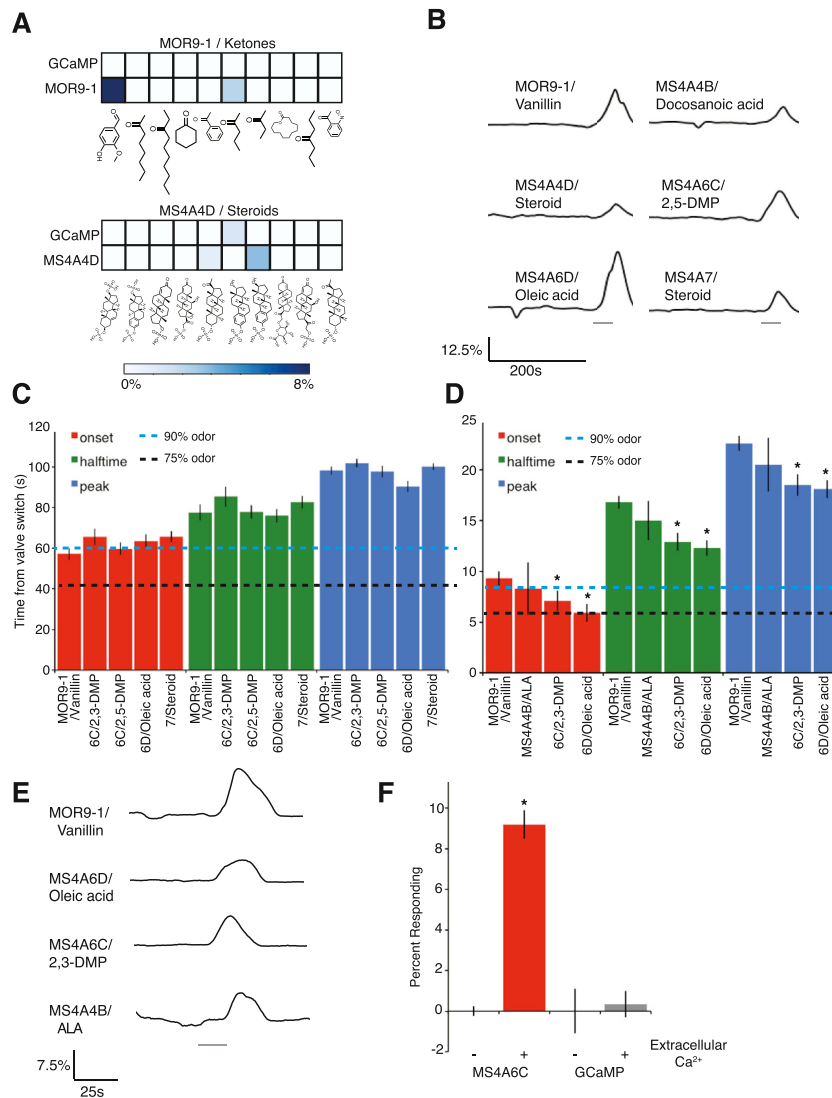


Figure S4. Supplement to Figure 3

(A) Deconvolution of selected odorant mixtures reveals monomolecular compounds that specifically activate a conventional odorant receptor (MOR9-1) or MS4A4D. Individual odors were delivered at 50 μ M in liquid phase to cells co-expressing GCaMP6s and the indicated MS4A receptor or mOR (bottom) or GCaMP6s alone (top). The aggregate percent of cells that responded to each chemical across three independent experiments is color-mapped as indicated, where only statistically significant responses are plotted (see [Experimental Procedures](#)).

(B) Traces of dF/F averaged across all cells that responded to the best monomolecular odorant for each MS4A/mixture pair.

(C) The timing of responses of four specific MS4A/odor pairs was assessed by comparing time of response onset (red), time to half-maximal response (green), and time to peak response (blue) to the same parameters for cells expressing the control odorant receptor MOR9-1 and its optimal ligand (vanillin). No statistically-significant differences were observed between any of the MS4A/odor pairs and MOR9-1/vanillin (Student's T-test, Bonferroni corrected for multiple comparisons). Data are presented as mean $\pm$ SEM. The black and blue dashed lines indicate the estimated time when odor concentration reaches 75% and 90% of maximum, respectively (determined by dye experiments, see methods). Between 29 and 65 responding cells across three independent coverslips were analyzed per condition.

(D) As in C except calcium was reported using GCaMP6f, images were acquired at 2 Hz, and odorants were delivered with a stimulus pencil that was positioned directly above the cells. Between 129 and 700 responding cells from at least 4 independent coverslips were analyzed per condition. * $p < 0.01$, Student's T-test.

(E) Traces of dF/F from individual cells that responded to stimulus pencil-delivered odorant for each MS4A/ligand pair for which dose response curves were generated.

(F) The requirement of extracellular calcium for MS4A ligand responses was assessed by stimulating HEK293 cells expressing GCaMP6s alone or co-expressing MS4A6C and GCaMP6s with 2,3-DMP in the presence or absence of extracellular calcium. Data are presented as mean $\pm$ SEM from at least 4 coverslips per condition. * $p < 0.01$, Z-test of proportions.

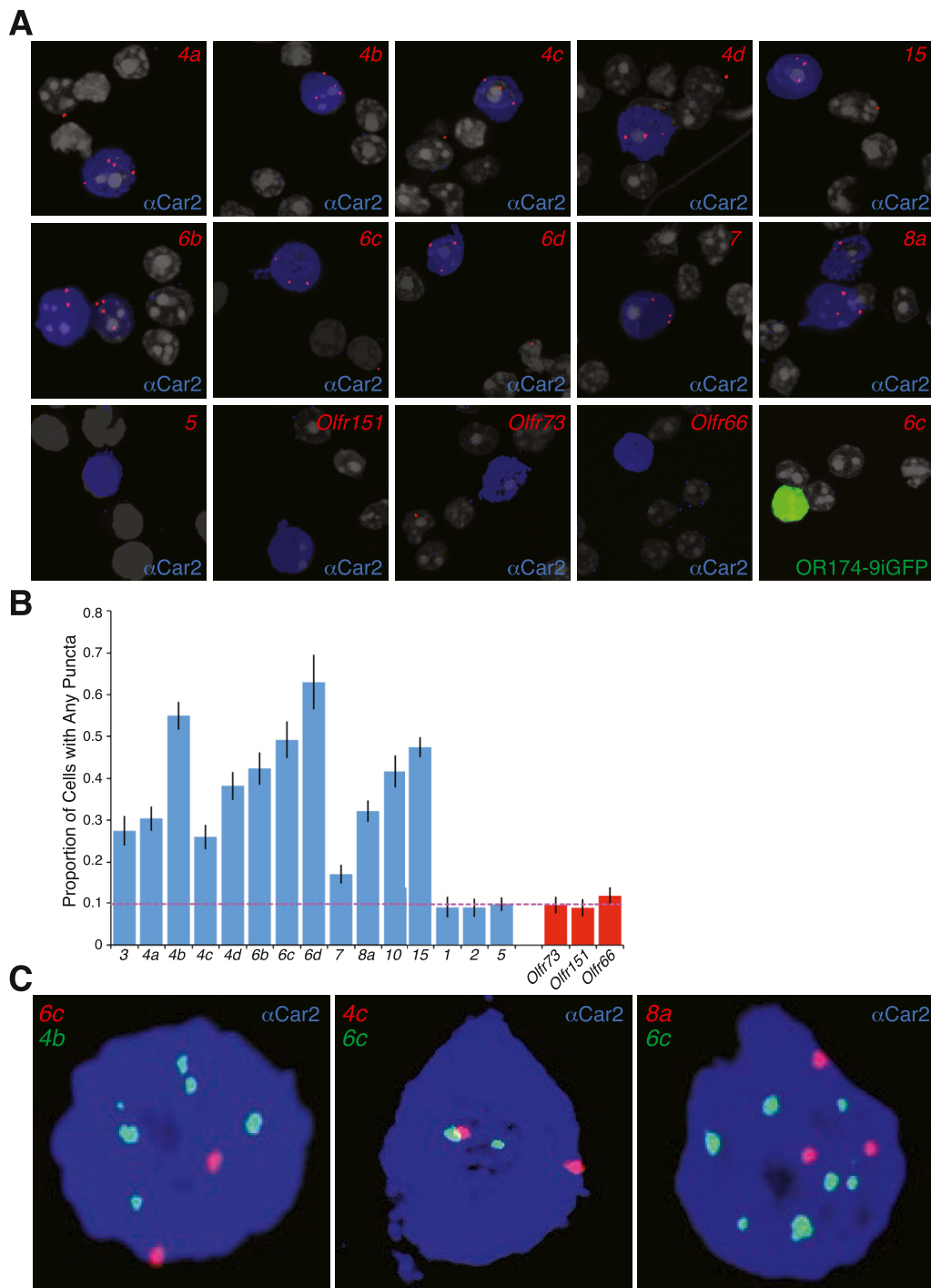


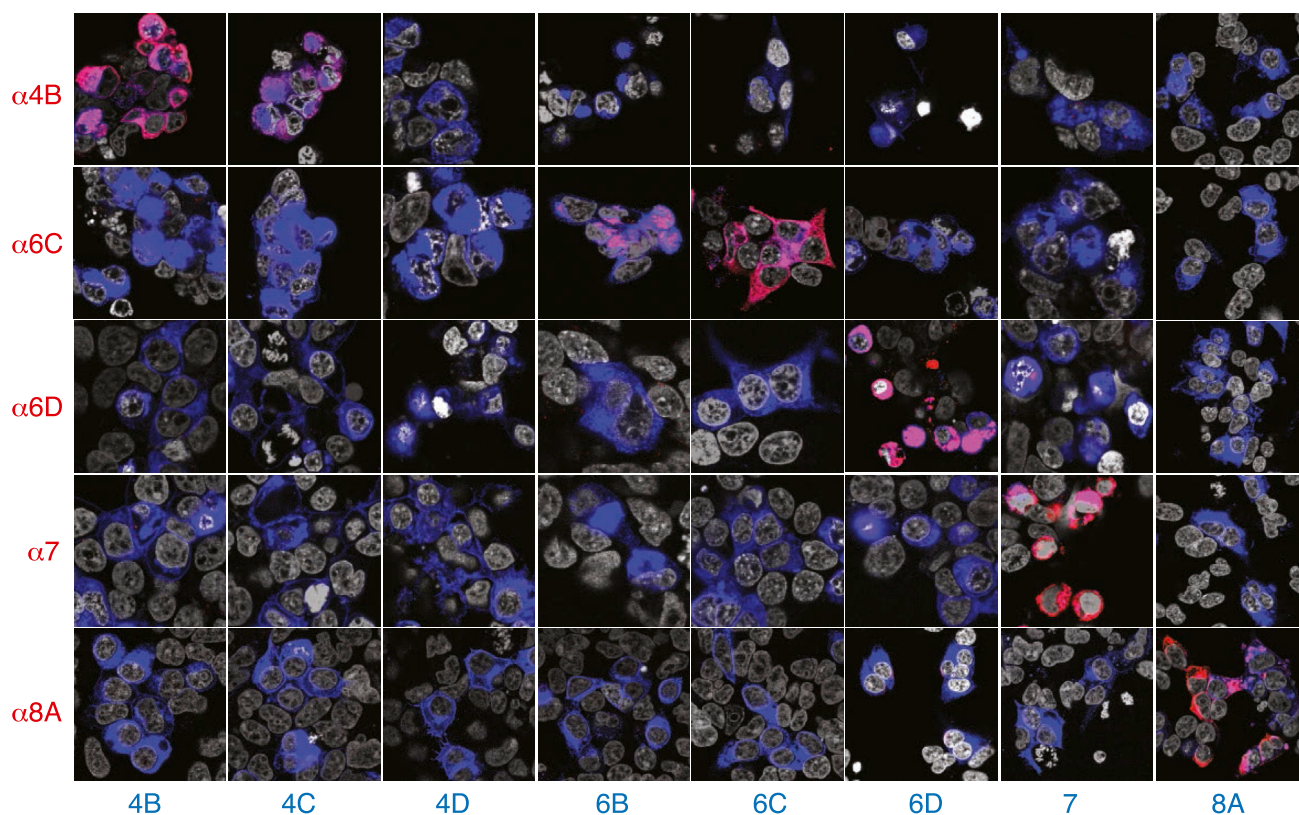
Figure S5. Related to Figure 4

(A) Representative images from RNAscope assays of dissociated olfactory epithelial cells. Necklace cells were identified with an antibody against Car2 (blue), and puncta from probes against an *Ms4a* or *Olfr* family member are in red. DAPI stain is represented as grayscale. Note that the Car2 signal was intentionally saturated to identify the volumes of individual GC-D cells, enabling unambiguous assignment of red puncta to individual GC-D cells. DAPI stain is represented as grayscale. *Ms4a6c* puncta are not found in GFP+ cells from dissociated OR174-IRES-GFP epithelia (bottom right).

(B) Proportion of necklace OSNs (identified as Car2+) with one or more fluorescent puncta for each *Ms4a* and *Olfr* probe (n=3 experiments, between 150-750 cells/probe, error bars are standard error of the proportion). Dashed red line represents the average value of negative controls (*Ms4a1*, *Ms4a2*, *Ms4a5*, and the *Olfr* genes).

(C) Representative images of Car2+ (blue) cells co-labeled with additional *Ms4a* probe pairs.

A



B

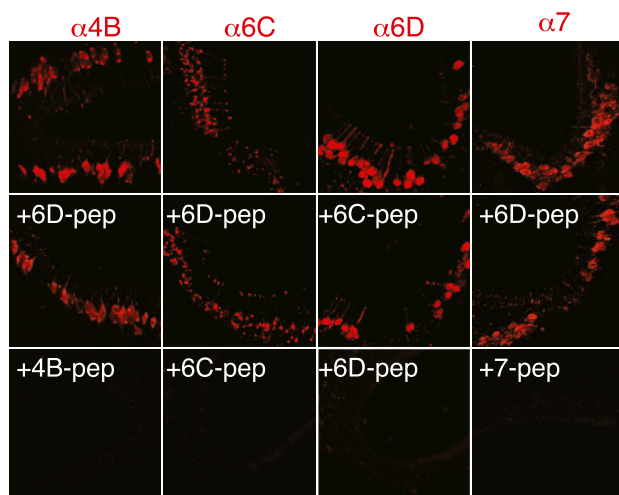


Figure S6. Related to Figure 5

(A) Representative images of HEK293T cells transfected with a plasmid encoding a single mCherry-MS4A fusion protein and stained with the indicated anti-MS4A antibody; antibodies are specific, although under conditions of overexpression anti-MS4A4B cross-reacts modestly with the closely related MS4A4C, as does anti-MS4A6C with MS4A6B. anti-MS4A immunofluorescence (red) overlaps almost completely with appropriate mCherry fluorescence (blue) and is not seen in cells that do not express the plasmid at appreciable levels, visible by nuclear counterstain (grayscale).

(B) Representative images of cul-de-sac tissue sections immunostained in the presence of peptide competitor (~1000-fold molar excess, see [Experimental Procedures](#)). Only the antigenic peptide, and not a peptide from a different MS4A protein, blocks staining of necklace cells by a given antibody.

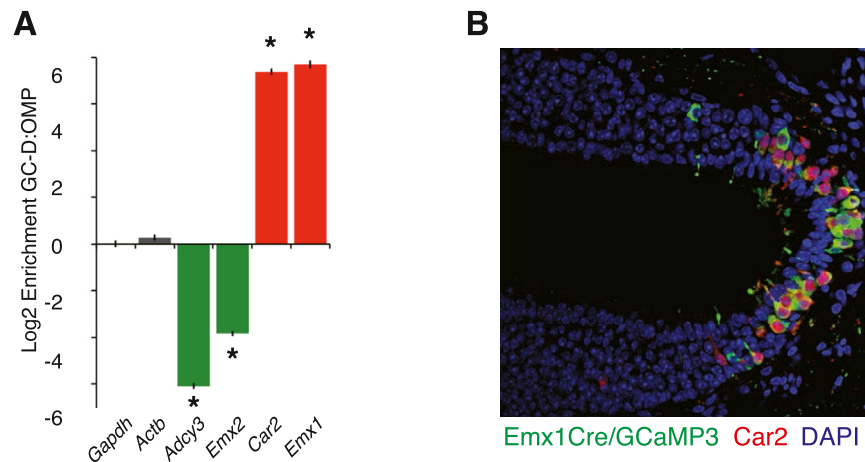


Figure S7. Related to Figure 6

(A) Quantification of mRNA expression in GC-D cells relative to OMP cells using the single-molecule detection method Nanostring. Marker genes for OMP cells such as *Adcy3* (green bars) and GC-D cells like *Car2* (red) are enriched in the appropriate populations. This analysis revealed that whereas OMP cells express the transcription factor *Emx2*, GC-D sensory cells exclusively express *Emx1*. Data are presented as mean $\pm$ SEM from three independent biological samples. * $p < 0.05$, paired t-test.

(B) Immunohistochemical analysis of sections prepared from the nasal epithelium of mice co-expressing an *Emx1-cre* allele and a Cre-dependent *GCaMP3* reporter using antibodies against GCaMP (green) and the necklace marker CAR2 (red) reveals that a large fraction of GCaMP-expressing cells are necklace cells; note that CAR2 staining tends to be enriched in nuclei whereas GCaMP is enriched in cytoplasm.

Cell, Volume 165

Supplemental Information

A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction

Paul L. Greer, Daniel M. Bear, Jean-Marc Lassance, Maria Lissitsyna Bloom, Tatsuya Tsukahara, Stan L. Pashkovski, Francis Kei Masuda, Alexandra C. Nowlan, Rory Kirchner, Hopi E. Hoekstra, and Sandeep Robert Datta

Supplemental Experimental Procedures

Mice:

GCD-IRES-tauGFP and *OMP-IRES-GFP* mice were obtained from the Jackson Laboratory (strain B6;129P2-Gucy2d^{tm2Mom}/MomJ, stock number 006704 and strain B6;129P2-Omp^{tm3Mom}/MomJ, stock number 006667, respectively). *OR174-9-IRES-tauGFP* mice were obtained from the Axel Laboratory (Sosulski et al., 2011). *Emx1-IRES-Cre* and *GCaMP3* mice were obtained from the Jackson Laboratory (strain B6.129S2-Emx1^{tm1(cre)Krtj}/J, stock number 005628 and strain B6;129S-Gt(ROSA)26Sor^{tm38(CAG-GCaMP3)Hze}/J and stock number 014538, respectively). Unless otherwise noted, all experiments on wild type mice were performed on 6-8 week old C57/BL6 male mice (Jackson Laboratory). All mouse husbandry and experiments were performed following institutional and federal guidelines and approved by Harvard Medical School's Institutional Animal Care and Use Committee.

Epithelial single cell dissociation:

Mice were euthanized with a lethal dose of Xylazine (~50 mg/mouse, Lloyd). Using 1X phosphate-buffered saline (PBS, VWR) to keep the epithelium moist during dissection, the head of the mouse was severed and the olfactory epithelium was rapidly dissected and placed in PBS. The epithelium was then transferred to a round-bottomed glass dish containing 1 mL of papain solution (one vial of Papain, (Worthington) dissolved in 5 mL of Earle's Balanced Salt Solution (Worthington) and then equilibrated 10 minutes at 37 °C, 5% CO₂) and 100 µL of DNase solution (one vial of DNase (Worthington) dissolved in 1 mL of Earle's Balanced Salt Solution (Worthington)). Bone was removed from the epithelium under a dissecting microscope (Leica MZ75) and the resultant tissue was placed in a 5 mL Falcon tube (Becton Dickinson) with an additional 1 mL of Papain solution and 100 µL of DNA solution and rocked gently for 30 minutes at 37 °C. The tissue was then gently triturated with a 5 mL stripette 10-15 times and the non-dissociated pieces of tissue were allowed to sediment by gravity for approximately 3 minutes. The supernatant was transferred to a fresh 5 mL Falcon tube and centrifuged 5 minutes at 300 x g. The supernatant was decanted, the cells were washed twice with 5 mL DMEM, high glucose, minus glutamine (Life Technologies)/10% heat inactivated Fetal Bovine Serum (FBS, Life Technologies) and resuspended in 1 mL DMEM/10% FBS to use for experimentation.

FACS:

Olfactory epithelial cells were dissociated as described above. To label hematopoietic cells, cell suspensions were incubated with PE-Cy5-conjugated rat anti-mouse CD45 antibody (1:2000, BD Pharmingen) for one hour at room temperature and then were washed twice with 1 mL DMEM/10% FBS. The cells were then incubated with Hoechst to allow the differential labeling of dead cells, before being placed on ice until they were sorted. Cells were sorted using either an Astrios EQ (Beckman Coulter, Fort Collins, CO) or a FACSARIAII (Becton Dickinson, San Jose, CA). Following the exclusion of dead cells and CD45 positive immune cells, GFP positive cells were then sorted to >99% purity by setting gates based on the fluorescent profile of cells derived from wild type, non-fluorescent mice. Cells were sorted directly into Trizol (Invitrogen) prior to RNA extraction (see below).

RNA isolation:

Fluorescent cells were sorted directly into 750 µL Trizol (Invitrogen) and samples were then vortexed to homogeneity before being placed on ice for the duration of the FACS session. All subsequent RNA work was performed in a RNA-dedicated PCR workstation (Air Clean Systems) that was UV-irradiated for 20 minutes prior to use and then was subsequently decontaminated with RNaseZap wipes (Ambion). Samples were passed through a 1 mL insulin syringe (Westnet) four times and the samples were allowed to sit for 5 minutes at room temperature. 200 µL of chloroform (Sigma) was then added to each sample, and the tubes were shaken vigorously by hand for 15 seconds and then placed at room temperature for 3 minutes. Samples were centrifuged at 12,000 x g for 15 minutes at 4 °C. The aqueous phase was removed to a nuclease free microcentrifuge tube (Ambion) and 10 µg of RNase-free glycogen (Life Technologies) was added as a carrier along with 500 µL of 100% 2-propanol (Sigma). The samples were inverted several times to homogenize and incubated for 10 minutes at room temperature. The samples were then centrifuged at 12,000 x g for 10 minutes at 4 °C and the supernatant was removed with a pipette. The RNA pellet was washed with 1 mL of 75% ethanol and then centrifuged for five minutes at 4 °C at 7,500 x

g. The supernatant was pipetted off, the pellet was allowed to air dry for five minutes and was resuspended in 5-10 μL of nuclease free water (Ambion) and stored at $-80\text{ }^{\circ}\text{C}$ until further use.

cDNA library generation:

cDNA libraries were generated from isolated RNA using a modified version of the Smart-Seq2 protocol that was optimized for reliability and sensitivity using “spike” RNAs (Ambion) (Picelli et al., 2014). Briefly, 0.6 μL of RNA was added to 8 μL lysis buffer (1.6 mM dNTPs, 2.5 μM oligo(dT)-VN primer (5'-AAGCAGTGGTATCAACGCAGAGTACT₃₀VN-3', where 'N' is any base and 'V' is A, C, or G), 0.2 μL RNase inhibitor (Ambion), 0.1% TX-100 v/v) in a nuclease free PCR tube and the samples were incubated in a Master Cycler ProS thermocycler (Eppendorf) for 3 minutes at $72\text{ }^{\circ}\text{C}$ before being held at $4\text{ }^{\circ}\text{C}$ for > 1 minute. Next, 11.4 μL of reverse transcriptase mix (10 μL Superscript II (Invitrogen), 5 μL RNase inhibitor (Ambion), 40 μL 5X first strand buffer (Invitrogen), 10 μL 0.1M DTT (Invitrogen), 8 μL 5 M Betaine, 37.8 μL nuclease free water (Ambion), 1.2 μL 1M MgCl_2 , 2 μL of TSO (5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3', which contains two riboguanisines (rG) and one LNA-modified guanine (+G) to facilitate template switching) was added to each sample and mixed by pipetting. The samples were incubated in the thermocycler for 90 minutes at $42\text{ }^{\circ}\text{C}$ and then were subjected to 10 cycles of 2 minutes at $50\text{ }^{\circ}\text{C}$ followed by 2 minutes at $42\text{ }^{\circ}\text{C}$, one 15 minute incubation at $70\text{ }^{\circ}\text{C}$, and a final hold at $4\text{ }^{\circ}\text{C}$ lasting at least 1 minute. Next the samples were amplified by adding 30 μL of PCR reaction mix (300 μL Kappa HiFi HotStart ReadyMix (2X KAPA Biosystems), 6 μL ISPCR primer (5'-AAGCAGTGGTATCAACGCAGAGT-3'), and 54 μL nuclease free water). The reactions were mixed by pipetting and incubated at $98\text{ }^{\circ}\text{C}$ for 4 minutes followed by 24 cycles of $98\text{ }^{\circ}\text{C}$ for 20 seconds, $67\text{ }^{\circ}\text{C}$ for 15 seconds, and $72\text{ }^{\circ}\text{C}$ for 6 minutes. A final incubation was performed at $72\text{ }^{\circ}\text{C}$ for 5 minutes before the samples were placed at $4\text{ }^{\circ}\text{C}$ for > 1 minute. The resultant cDNA was isolated using an Agencourt Ampure XP kit (Beckman Coulter) following the manufacturer's instructions and was subsequently stored at $-80\text{ }^{\circ}\text{C}$ until further use.

RNA sequencing:

cDNA libraries were sheared to a size of approximately 250 bases using a Covaris S2. The sheared cDNA was run on a Wafergen Apollo machine using the Kapa Genomic Library Construction Kit (Kapa Biosystems, KK8234). Ten cycles of PCR were run after the samples were removed from the robot and the reactions were cleaned using magnetic beads. Quality control was performed using an Agilent 2200 Tape Station with a D1000 High Sensitivity Tape with ladder provided. The morphology and overall concentration of the samples were assessed and those samples that passed a concentration cutoff were subjected to qPCR to more accurately determine concentration. qPCR was run with SYBR green using the KAPA SYBR FAST Universal 2X qPCR Master Mix reagent (Kapa Biosystems, KK4602) and primers complementary to the P5 and P7 regions of the adapter sequences. Serial dilutions of PhiX were used to generate a standard curve, which was, in turn, used to determine the concentration of cDNA in each sample prior to sequencing. Samples were sequenced on an Illumina 2500 in rapid mode on a single lane of SR50.

Alignment and differential expression:

RNA-sequencing reads were processed using the RNA-seq pipeline implemented in version 0.8.3a-9483413 of the bcbio-nextgen analysis project. Briefly, poor quality bases with PHRED scores less than five (Macmanes, 2014), contaminant adapter sequences, and polyA tails were trimmed from the ends of reads with cutadapt version 1.4.2, discarding reads shorter than twenty bases. A STAR (Dobin et al., 2013) index was created from a combination of the *Mus musculus* version 10 (mm10) build of the mouse genome and the Ensembl release 75 gene annotation. Trimmed reads were aligned to the STAR index, discarding reads with ten or more multiple matches to the genome. Quality metrics including mapping percentage, rRNA contamination, average coverage across the length of the genes, read quality, adapter contamination and others were calculated using a combination of FastQC, RNA-SeQC (DeLuca et al., 2012), and custom functions from bcbio-nextgen and bcbio.mnaseq (available upon request). Read mapping to genes were counted using featureCounts (Liao et al., 2014) version 1.4.4, excluding reads mapping multiple times to the genome and reads that could not be uniquely assigned to a gene. Counts were normalized and differential expression between cell types was called at the level of the gene using DESeq2

(Love et al., 2014) version 1.6.3. To identify gene families that were expressed in GC-D cells, the data set was filtered to identify transmembrane protein-encoding genes that were at least ten-fold enriched in GC-D cells with an adjusted p-value of < 0.05 after correcting for multiple comparisons using the methods described by Benjamini and colleagues (Benjamini et al., 2001). Despite using these stringent filtering criteria, with very large gene families one might still observe the expression of an occasional family member by chance and thus the data were filtered further to only consider gene families in which at least three members were ten-fold enriched with an adjusted p-value of < 0.05 . From this list of gene families, the *Ms4a* genes were selected for further study as they exhibited significant molecular diversity to encode chemoreceptors (see Figure 2) and had relatively uncharacterized function.

Nanostring:

10,000 GFP positive cells from *Gucy2d-IRES-GFP* or *Omp-GFP* mice were sorted into Trizol and the RNA was isolated as described above. Three biological replicate RNA samples were hybridized to Nanostring probes using nCounter Elements reagents according to the manufacturer's specifications. The protocol was modified to perform the hybridization step at 67 °C for 48 hours to maximize the detection of low abundance transcripts. RNA molecules that hybridized to probe were captured and quantified using an automated Nanostring prep station following the manufacturer's instructions. The resultant data were analyzed using nSolver software. Briefly, the average number of detected molecules for six internal negative control probes (whose complementary sequences are not present in the mouse genome) was used to calculate a rate of non-specific hybridization. After subtracting the amount of binding resulting from non-specific interactions, the number of molecules of each RNA transcript found in GC-D samples and OMP samples was compared using Student's t-test. See Supplemental Table 2 for probe sequences used in these experiments.

Multi species alignment of Mus musculus MS4A proteins:

FASTA format sequences of the indicated *Mus musculus* MS4A proteins were downloaded from the NCBI protein database and aligned using the PRALINE sequence alignment program on the Centre for Integrative Bioinformatics VU website using the default settings. Amino acid conservation across family members was scored using the PRALINE default settings where the least conserved amino acids were given a 0 score and the most conserved amino acids were assigned a 10 (Simossis and Heringa, 2005). TOPCONS was used to determine the predicted topology of the MS4A family member that was used on the top line of the alignment. All topographical representations were generated using the Protter program and manually entering the topographical orientation of the MS4A protein as predicted by TOPCONS.

Phylogenetic and selection analyses with MS4A genes:

MS4A sequences were retrieved from both Ensembl and NCBI databases and imported into Geneious v8 (Biomatters Ltd). We chose 37 representative taxa from all the major mammalian lineages (see Supplemental Table 1 for full list). When a gene had more than one predicted isoform, the sequence that contained the longest open-reading frame was selected. Coding DNA sequences were translated, aligned with MAFFT v7.017 (Katoh and Standley, 2013a, b) using the E-INS-i algorithm, the BLOSUM80 scoring matrix, and a gap-opening penalty of 1. Sequences were then back-translated into codons. For phylogenetic reconstruction of the multigene family tree, the OpenMPI version of MrBayes v3.2.1 (Ronquist et al., 2012) and the GTR+I+G model as determined by jModelTest 2.1.7 (Darriba et al., 2012) were used. The final dataset consisted of 411 sequences and 447 characters corresponding to sites present in at least 75 percent of the aligned sequences. For individual gene tree reconstructions and evolutionary analyses, sequence subsets were extracted based on their group membership as predicted based on the multigene family tree. Sequences were then realigned corresponding to each subset as above, the resulting alignments were trimmed to remove positions that contained gaps in the majority of sequences. The phylogenetic reconstruction was carried out using the OpenMPI version of RAxML v8 (Stamatakis, 2014).

To identify branches under episodic positive selection, the random-effects likelihood branch-site method (BS-REL) (Kosakovsky Pond et al., 2011) was used as implemented in the HyPhy package (Pond et al., 2005). The branch-site models allow the nonsynonymous to synonymous substitution rate ratio ω (d_N/d_S) to vary both among amino acid sites in the protein and across branches on the tree to detect positive selection affecting specific sites along particular lineages (Anisimova and Yang, 2007). We identified evidence for site-specific positive selection in MS4A homologs using the codeml program in the PAML

v4.8 software package (Anisimova and Yang, 2007). In brief, we compared different site models, in which the evolutionary rate ω is allowed to vary among sites. Here, we focused on comparison of model pairs: M1a (neutral; codon values of ω fitted into two discrete site classes between 0 and 1) versus M2a (positive selection; similar to M1a but with one additional class allowing $\omega > 1$); M7 (neutral; values of ω following a beta distribution with $\omega = 1$ maximum) versus M8 (positive selection; similar to M7 but with one additional class allowing $\omega > 1$); and M8a (neutral; similar to M7 but with one fixed class with $\omega = 1$) versus M8. Multiple starting values of ω were used, and either the F3x4 or F61 model of codon frequencies. To evaluate whether the models allowing positive selection provided a significantly better fit to the data, likelihood ratio tests were used. Notably, the M1a-M2a comparison is more stringent and can lack power to detect signatures of diversifying selection compared to the M7-M8 models, which impose less constraints on the distribution of ω . Finally, the M8a vs. M8 comparison can be used to contrast the potential role of reduced purifying selection (or relaxation) versus positive selection. When the null model is rejected, the empirical Bayes procedure was implemented under model M8 to identify sites under positive selection (posterior probability ≥ 0.90). To identify sites that have experienced purifying selection (posterior probability ≥ 0.90), the Fast Unconstrained Bayesian Approximation (FUBAR) (Murrell et al., 2013) was used as implemented in the HyPhy package. Consensus topology predictions were made using a standalone version of TOPCONS2.0 (Tsirigos et al., 2015). All computational analyses were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

Plasmids:

pCI-MOR9-1 was a gift from Hiroaki Matsunami (Addgene plasmid # 22331) (Saito et al., 2009). pGP-CMV-GCaMP6s was a gift from Douglas Kim (Addgene plasmid # 40753) (Chen et al., 2013). The GNAI15 expression plasmid was kindly provided by Steve Liberles. DNA sequences encoding mCherry-MS4A were cloned into the tetracycline inducible mammalian expression plasmid, pcDNA5-FRT-TO using standard molecular biology methods.

Heterologous expression of MS4A proteins:

Flp-In-T-Rex HEK293 cells (Invitrogen R78007) were maintained in complete media (DMEM, high glucose, no glutamine (Life Technologies) supplemented with 10% tetracycline free Fetal Bovine Serum (Clontech), penicillin, streptomycin, and glutamine (Life Technologies)) with 5% CO₂ in a 37 °C humidified tissue culture incubator (NuAire) on 10 cm tissue culture plates (BD #430167). Approximately four hours prior to transfection, cells were washed once with 10 mL plain DMEM and then incubated with 2 mL of Trypsin-EDTA solution (ATCC) for approximately 3 minutes at 37 °C. 10 mL of complete media was then added to the plate and the cells were triturated 5-10 times vigorously to generate single cell suspensions. After centrifugation at 300 x g for 5 minutes, the supernatant was aspirated and the cells were resuspended in complete media before they were plated on Round German Glass 15 mm coverslips (Bellco Biotechnology) in 12 well plates, which had been incubated for at least 24 hours with 0.02 mg/mL poly-D-lysine hydrobromide (Sigma) before being washed twice with ddH₂O. After four hours, cells were transfected with calcium phosphate. For each coverslip, a 50 μ L reaction mix consisting of 250 mM CaCl₂, approximately 2.5 μ g GCaMP6s encoding plasmid, and 1 μ g MS4A encoding plasmid. This mix was homogenized by pipetting 4 times. To this reaction mix, 50 μ L of 2X HeBs (274 mM NaCl, 10 mM KCl, 1.4 mM Na₂PO₄·7H₂O, 15 mM D-glucose, 42 mM Hepes (free acid)) pH 7.04-7.10 was added in a swirling motion from the bottom of the tube and bubbled briefly with air. This mixture was incubated for five minutes at room temperature, pipetted once to mix, and added to the well in a drop-wise manner. The 12 well plate was then shaken 5 times along each major axis before being placed in the incubator. Approximately one hour later tetracycline (Sigma #T7760) was added to a final concentration of 1 μ g/mL and cells were allowed to express MS4A proteins overnight.

MS4A6C surface expression:

HEK293 cells were transfected with a plasmid encoding GCaMP6s and either a control plasmid or a plasmid encoding a mCherry-MS4A6C fusion protein as described above. 16 hours following transfection, 1 μ g of purified anti-MS4A6C antibody was added to the cells. The cells were incubated with antibody for 1 hour at room temperature in the dark, and then washed 3X with complete media. Cells were subsequently fixed with 4% paraformaldehyde/1X PBS for ten minutes and then washed three times with 1X PBS. Cells were re-blocked with PBS/0.3% Triton X-100/5% donkey serum for 30 minutes and then incubated with Alexa633 goat anti-rabbit secondary antibody (1:300) for 45 minutes. The fixed and stained

cells were washed three times with block, and the coverslips were mounted on slides using VECTASHIELD Mounting Medium with DAPI (Vector laboratories).

Odors:

Steroids were purchased from Steraloids (Newport, Rhode Island). Additional compounds were purchased from Sigma and were obtained at the highest purity possible.

In vitro odor screen:

HEK293 cells were co-transfected as described above with a plasmid encoding GCaMP6s and either a plasmid encoding one of the MS4A proteins or plasmids encoding the odorant receptor MOR9-1 and the G protein GNAI15, (which couples the exogenous GPCR to intracellular calcium stores, see Ukhanov et al., 2014). Coverslips were washed with Ringers solution and then secured in a perfusion/imaging chamber (Warner Instruments) using High Vacuum Grease, Dow Corning (VWR). Ringers solution was constantly perfused over the cells at a rate of ~10 mL/minute using an 8-channel valve-controlled gravity-driven perfusion system (Warner Instruments) and images were acquired using the MetaFluor software package using an Olympus IX83 microscope, a Sutter Lambda DG4 Light System with a Xenon arc lamp, and an Andor Neo 5.5 sCMOS camera. A 20X NA 0.45 air objective (Olympus) was used to acquire images at ~0.67 Hz at 3 x 3 pixel binning to reduce the exposure times required to obtain images thereby limiting photobleaching and phototoxicity. Each imaging epoch consisted of 900 seconds. For the first 300 seconds images were acquired but not saved as a significant amount of GCaMP6s signal decay occurred within this window. Subsequently the experiment consisted of 150 seconds of Ringers, valve switch and 75 seconds of Ringers (from a separate reservoir to mimic odor stimulation), valve switch and 150 seconds of Ringers, valve switch and 75 seconds of odorant, followed by a final valve switch and 150 seconds of Ringers. Odorants were delivered as mixtures of chemicals diluted in Ringers solution such that each individual constituent was present at 10 μ M final concentration. Each coverslip was exposed to a single imaging epoch to control for response adaptation. To assess the time required to clear dead volumes and to estimate mixing delays, 100 nM RhodamineB dye was flowed instead of odor and dye saturation kinetics were determined. In figures 3A and S4B stimulus bars begin when odor is estimated to reach 90% saturation.

To identify monomolecular compounds that activate HEK293 cells expressing individual MS4A family members, several odorant mixes that elicited statistically significant responses in cells expressing a particular MS4A protein were selected for deconvolution analysis. The final concentration of each odorant in the deconvolution analysis was 50 μ M except for the polyunsaturated fatty acid constituents, which were delivered at 10 μ M, as at higher concentrations some of these molecules non-specifically activated HEK293 cells.

Analysis of in vitro screen data:

Analysis was performed using fully automated custom scripts in iPython that employed NumPy, SciPy, Pymorph, Pandas, and Mahotas source code packages. Each fluorescent image of the time series was locally smoothed and downsampled 3-fold in both dimensions to facilitate subsequent processing. The downsampled image stack was segmented into cell areas using a watershed transform, which were then used to extract the fluorescence values of individual cell areas over time. The resulting traces were baseline-corrected using a wide (100 frames) moving median filter to remove long-timescale drifts such as fluorescence decay due to photobleaching and smoothed using a local (9 frames) moving average filter. A total of seven coverslips from the mixture screen and five coverslips from the deconvolution screen were removed from the analysis for failing to meet quality control.

Processed traces were normalized by calculating an average fluorescence value across the period prior to odor stimulation and then dividing the entire trace by this value (thereby generating a dF/F trace). Responses were identified as fluorescent peaks within a twenty frame window centered at the time point at which maximal odor concentration occurred (determined empirically using RhodamineB dye) that were five standard deviations above the noise observed in the baseline period. The spontaneous activity response rate (determined as responses observed during twenty frames of the Ringers only period) was subtracted from the odor-evoked response rate; this rate did not change based upon the specific position of the window chosen in the trace (as long as it did not overlap with odor delivery). The proportion of cells co-expressing a chemoreceptor and GCaMP6s that responded to a given stimulus was compared by Z-test to the

proportion of cells expressing only GCaMP6s that responded to the same stimulus. Only response tallies with a P-value of < 0.05 following FDR correction for multiple comparisons are depicted.

In vitro functional experiments with focal stimulus delivery pencil:

To characterize ligand-induced response kinetics with greater temporal precision than in the bulk calcium assay, HEK293 cells were transfected with a plasmid encoding the fast version of GCaMP6 (GCaMP6f) and plasmids encoding either an MS4A or MOR9-1/GNAI15. The cells were perfused with Ringer's solution and odorants were delivered in liquid phase with a six channel gravity-controlled perfusion manifold (Warner Instruments) through a custom-made "stimulus pencil" made of quartz tubing focally positioned relative to the field of view. Images were acquired as described for the *in vitro* odor screen with an imaging rate of 2 Hz. The delay between valve switch and odor delivery was determined using RhodamineB dye as in the bulk calcium assay.

Dose response curves and EC50 calculations:

Dose response curves were determined as in (Mainland 2015). HEK293 cells co-expressing GCaMP6f and MOR9-1/GNAI15 or an MS4A were odor stimulated with a focal stimulus delivery pencil as described earlier. Each field of view was stimulated using the stimulus pencil with odorant spanning six log orders of odorant (from 10 nM to 1 mM) starting with the lowest concentration. Fluorescent traces were extracted and normalized as described above for the *in vitro* odor screen. Because of the tighter stimulus control afforded by this configuration, cells were considered to have responded if they exhibited fluorescent peaks greater than 4 standard deviations above baseline within a 15 second window centered around the time of maximal odor responses. A delay of 150 seconds was included between stimulus presentations. The cumulative fraction of cells that had responded by each odor concentration was determined for both receptor-expressing and control cells. To account for non-specific activation that might occur at higher concentrations of odorant the fraction of control cells (expressing GCaMP only) that was activated was subtracted from the fraction of receptor-expressing cells and sigmoidal dose response curves were then fit to the receptor data. Between four and twenty coverslips for both control and receptor-expressing cells were imaged to construct each dose response curve and each data point represents the mean $\pm$ SEM of the relative cumulative fraction of cells that responded to that odor concentration.

Analysis of calcium transient temporal dynamics:

Analysis was performed by applying fully automated custom iPython scripts to extracted fluorescence traces. Response onset, half-rise time and peak time were identified using a peak and trough finding package in SciPy (SciPy.signal.find_peaks_cwt) (Du et al., 2006). Only cells that had response peaks greater than 4 standard deviations above baseline were included for analysis. Two-tailed Student's T-tests were run on the distributions of data and were Bonferroni corrected for multiple comparisons.

Requirement of extracellular calcium for MS4A ligand responses:

Experiments were performed similarly to those described using the stimulus pencil delivery system. Cells were either perfused with Ringers supplemented with 1 mM calcium chloride (plus calcium) or 1 mM EGTA (minus calcium) to chelate calcium. The percent of cells that responded was determined as described above.

RNAscope Fluorescent In Situ Hybridization:

Fluorescent *in situ* hybridization was performed on dissociated olfactory epithelial cells adhered to glass coverslips. Nasal epithelia from 8-12 week old C57/BL6 male mice, OR174-9-IRES-tauGFP mice, or GC-D-IRES-tauGFP mice were dissociated with the same method used for FACS, except that they were incubated in papain + DNase I for 90 minutes. The final cell pellet from a single mouse epithelium was resuspended in 900 μ L of DMEM + 10% FBS, and 75 μ L was placed on each of twelve 12 mm #1 glass coverslips (Bellco Biotechnology #1943-10012) that had been coated with 30 μ L of 5 mg/mL poly-D-lysine (Sigma Aldrich #P6407), dried, and washed three times with deionized water. Cells were allowed to settle on coverslips for 30 minutes at room temperature, rinsed once with 1X PBS, and fixed with 4% paraformaldehyde (PFA)/1X PBS (diluted from 20% PFA, Electron Microscopy Sciences #15713) for 30 minutes at room temperature. Coverslips were washed three times with 1X PBS and passed through a dehydration series of 50%, 70%, and 100% ethanol (v/v in water) for 5 minutes each. The last wash was replaced with fresh 100% ethanol and cells were stored at -20 °C overnight.

On the second day, coverslips were stained for specific mRNA targets by the RNAscope protocol (Advanced Cell Diagnostics, RNAscope Multiplex Fluorescent Assay For Fresh Frozen Tissues User Manual rev. 20121003). Several modifications to the RNAscope protocol were made to combine FISH for low abundance targets with immunohistochemistry: after rehydration, coverslips were treated with protease ("Pretreat 4") diluted 1:30 in 1X PBS; RNAscope target probes, which are supplied at 50X concentration, were diluted to final 2X concentration in blank probe solution; hybridization lasted for 4 hrs at 40 °C in the HybEZ oven (Advanced Cell Diagnostics #310010); and for experiments using type C1 and C2 probes (see below) the fluorescent reagent Amp4-altB was used, whereas for experiments using C1 and C3 probes we used Amp4-altC (Advanced Cell Diagnostics #320850).

To combine FISH with immunostaining, at the end of the RNAscope protocol, instead of being counterstained and mounted, the coverslips were rinsed in 1X PBS twice and blocked in 1X DAKO Serum-Free Protein Block (DAKO X0909) in PBS for 20 minutes at room temperature. Coverslips were then incubated with primary antibody (rabbit anti-Car2 or chicken anti-GFP as below) in blocking solution for 1 hour at room temperature, washed three times for 5 minutes in PBS, and stained with secondary antibody in PBS for 30 minutes at room temperature. After three final washes in PBS the coverslips were counterstained with ProLong Diamond Antifade Mountant + DAPI (Life Technologies #P36961), fixed to slides with nail polish, and imaged by epifluorescent or confocal microscopy as described below.

To quantify RNAscope signal, coverslips were viewed under epifluorescence with a 63X/NA1.4 oil immersion objective lens (Zeiss Plan-Apochromat 63X/1.40 OIL DIC 440762-9904-000) and fluorescent puncta visible by eye were counted for each immunopositive cell. In initial experiments puncta in random immunonegative cells were also quantified and the background levels for MS4A probes were found to be ~ 1 punctum/5-10 cells (data not shown). In approximately 10% of experiments much higher background (3+ puncta/cell) was observed, possibly due to sample drying. These coverslips were discarded from the analysis. For the remainder of the analysis puncta were quantified only in identifiable cells (i.e., GC-D/CAR2 or OR174-9 expressing cells.)

The target probes used in this assay are custom reagents designed by Advanced Cell Diagnostics and are currently available in the ACD catalog. Mouse target probes used are as follows: Ms4a1-c1 #318671, Ms4a2-c1 #438171, Ms4a3-c1 #438181, Ms4a4A-c1 #427391, Ms4a4B-c1 #314611, Ms4a4C-c1 #426371, Ms4a4D-c1 #427401, Ms4a5-c1 #318641, Ms4a6B-c1 #313801, Ms4a6C-c1 #314581, Ms4a6D-c1 #314591, Ms4a7-c1 #314601, Ms4a8A-c1 #426361, Ms4a10-c1 #438151, Ms4a15-c1 #427381, Olfr173-c1 #313771, Olfr151-c1 #431161, Olfr66-c1 #431171, Ms4a6C-c3 314581-C3, Ms4a4B-c3 #314611-C3, Car2-c2 #313781-C2, Gucy2d-c2 #425451-C2, Pde2a-c3 #426381-C3.

MS4A Antisera:

Peptides derived from mouse MS4A protein sequences were synthesized by Covance, Inc. (Denver, PA) as follows: 6C-pep: CKQSKELSLIEHDYYQ; 6D-pep: SQNSKNKSSVSSESLC; 7-pep: HKREKTGHTYEKEDD; 4B-pep: HQGTNVPNGVYKNHPC; 8A-pep: TAKSWEPEQERLTWC; 5-pep: TTQEYQTTELTATAYNC. These peptides were coupled by terminal cysteine residues to KLH and used to raise an immunoglobulin response in rabbits and guinea pig (6C-pep only). Sera collected by Covance, Inc. from rabbits and guinea pig were tested for ability to stain 293T cells expressing MS4A protein and GC-D cells (guinea pig anti-6C and rabbit anti-6C antibodies yielded the same pattern of staining *in vitro* and *in vivo*); the best lots of serum were purified by passing over protein A/sepharose (Life Technologies #10-1042) and eluted with 100 mM Glycine, pH 2.7 to collect immunoglobulin. A subset of these (anti-6C, 6D, and 7) were further affinity purified by passing over resin cysteine-coupled to the respective antigenic peptides (Thermo Scientific #44999) and fractions eluted with 100 mM glycine, pH 2.7 were tested for greatest specificity on 293T cells. We also raised antisera against GC-D protein with the peptide (Ac-QRIRTDGKGRRLLAC), which was purified by passing over protein A/sepharose and a peptide affinity column as above. For peptide competition experiments, antibody concentration was estimated by reacting with protein assay dye reagent (Bio-Rad #5000006) and measuring 595 nm absorbance with a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). Sample concentrations were estimated by fitting to a line generated with IgG standards; most purified antibodies were between 1.0 and 5.0 mg/mL. Peptides were resuspended in 1X TBS at 1 mg/mL and mixed with diluted antibodies in block at a 10 mg peptide: 1 mg antibody ratio; as the peptides are ~15 amino acid residues and IgGs are ~1500 residues total, it was estimated that this gives ~1000-fold molar excess of peptide. Similar results were obtained with lower peptide concentration (data not shown).

Primary antibodies/concentrations used were as follows: goat anti-Pde2a (1:50, Santa Cruz Biotechnology sc-17227), chicken anti-GFP (1:1000, Abcam #ab13970), rabbit anti-AC3 (1:100, Santa Cruz Biotechnology sc-588), rabbit anti-GC-D (serum 7444, affinity fraction 2, 1:1000), rabbit anti-Car2 (1:500, Abcam #191343), goat anti-CD20 (1:50, Santa Cruz Biotechnology #sc-7735), rabbit anti-MS4A4B (serum 7512, PAS fraction, 1:2000), rabbit anti-MS4A6C (serum 7277, affinity fraction 3, 1:500), anti-MS4A6D (serum 7284, affinity fraction 3, 1:500), anti-MS4A7 (serum 7286, affinity fraction 2, 1:500), anti-MS4A8A (serum 7508, PAS fraction, 1:2000), anti-MS4A5 (serum 7503, PAS fraction, 1:2000), guinea pig anti-MS4A6C (serum 7630, PAS fraction, 1:1000), anti-S6 phosphoSerine240/244 (1:200, Cell Signaling Technologies #22155).

Secondary antibodies/concentrations used were as follows: donkey anti-rabbit-Cy3 (1:300, Jackson ImmunoResearch, #711-167-003), donkey anti-rabbit-CF633 (1:300, Biotium #20125), donkey anti-goat-Alexa633 (1:300, Invitrogen #A21082), donkey anti-goat-Alexa488 (1:300, Jackson ImmunoResearch #705-546-147), donkey anti-guinea pig-Alexa647 (1:300, Jackson ImmunoResearch #706-606-148), donkey anti-chicken-Alexa488 (1:300, Jackson ImmunoResearch #703-545-155).

Cultured Cell Preparation for Immunostaining:

293T cells adhered to 12 mm coverslips and transfected with mCherry-MS4A expression plasmids (see above) were fixed 24-48 hours post-transfection with 4% PFA/1X PBS (Electron Microscopy Sciences) for 10 minutes at room temperature. Coverslips were washed three times with 1X PBS and stored at 4 °C in 1X PBS until immunostaining as described below.

Tissue Preparation for Immunostaining:

All preparation of nasal epithelia and olfactory bulbs was performed as follows, except for assays involving phospho-S6: animals were euthanized as above and olfactory epithelia were dissected out from the skull with olfactory bulbs attached and fixed overnight in 4% PFA/1X PBS at RT. After washing 5 minutes with 1X PBS three times, noses were decalcified overnight in 0.45M EDTA/1X PBS at 4 °C. Noses were washed once more with 1X PBS, then sunk in 20% sucrose for 3 hrs at 4 °C, and finally embedded in Tissue Freezing Medium (VWR #15146-025). 15 micron cryosections were cut onto Superfrost Plus glass slides (VWR #48311703) and stored at -80 °C until staining. For experiments involving anti-phosphoS6 staining, nasal epithelia were instead fixed overnight in 4% PFA/1X PBS at 4 °C.

Immunostaining:

For experiments without anti-phosphoS6, cryosections on slides were removed from the freezer, air dried for 10 minutes, and antigen retrieved by immersion for 10 minutes in a 98 °C solution of 10 mM Sodium Citrate, 0.1% Tween-20, pH 6.5. Slides were rinsed in 1X Tris-buffered Saline (TBS: 50 mM Tris-Cl, 150 mM NaCl, pH 7.5) at room temperature and blocked in 5% Normal Donkey Serum (Jackson ImmunoResearch #017-000-121)/0.1% Triton-X100/1XTBS for 30 minutes at room temperature. All immunostaining was done at 4 °C overnight with antisera diluted in blocking solution.

On the following day, slides were washed three times for 10 minutes in 1X TBS/0.1% TritonX100 at room temperature, and then incubated in secondary antibody solution (antibodies listed above in blocking solution) for 45 minutes at room temperature. Finally, slides were washed three times for 10 minutes in 1X TBS/0.1% TritonX100, counterstained with Vectashield + DAPI (VWR #101098-044), and coverslipped before confocal microscopic imaging.

For experiments involving anti-phosphoS6, staining was performed as above with the following changes: there was no antigen retrieval step before blocking; blocking solution was 1X PBS/0.1% TritonX100/3% Normal Donkey Serum/3% BSA; all washes were done in 1X PBS/0.1% TritonX100; and the primary antibody solution included 50 nM of a phosphopeptide ("3P") derived from the S6 sequence (Knight et al., 2012). This peptide, which contains phosphorylated Serine235/236/240, is meant to compete away nonspecific antibody binding to S6 protein not phosphorylated at its neuronal activity-dependent sites.

Finally, assaying *Ms4a6d-IRES-GFP*-infected noses required costaining with two rabbit primary antisera, one to MS4A6D and the other to phosphorylated S6. To prevent cross-reactivity with secondary antibodies, phospho-S6 staining was performed first as described above; after application of fluorescent secondary antibody, remaining sites on the rabbit primary antibody were blocked by incubation with 5 µg/mL unconjugated donkey anti-rabbit Fab fragments (Jackson ImmunoResearch) in washing solution at

room temperature for 1 hour. After washing out unbound Fab fragments three times with wash solution, tissue sections were stained with anti-MS4A6D overnight as described above and detected with a CF633-conjugated secondary antibody (Biotium). No fluorescence to indicate cross-reactivity was observed, and control experiments without anti-MS4A6D antiserum showed that the later secondary antibody did not interact with the Fab-blocked phospho-S6 primary antibody.

Explant calcium imaging:

Adult offspring of *Emx1-IRES-Cre* and *ROSA-GCaMP3* mice were sacrificed and dissected to expose the lateral olfactory epithelium, adjacent to the olfactory bulb. The epithelium and adjacent skull (including the adjacent olfactory bulb) was embedded in a custom-made perfusion chamber with 5% low-melt agarose (Invitrogen) made from modified Ringer's solution (115 mM NaCl, 5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 25 mM NaHCO₃, 10 mM HEPES, pH 7.4). Small slits were cut anterior to the cul-de-sac regions of exposed olfactory epithelium to allow fluid flow, and for the remainder of the experiment the tissue was superfused with carboxygenated modified Ringer's solution (95% O₂, 5% CO₂). Odorants were delivered to the tissue in liquid phase *via* an 8-to-1 perfusion manifold controlled by solenoid valves (Warner Instruments) through a custom-made "stimulus pencil" positioned near the olfactory epithelium. Each odorant or mixture was diluted in 0.1% DMSO in modified Ringer's solution to a concentration of 100 μ M.

Cul-de-sac regions of the tissue containing GCaMP3+ cells were imaged through the epithelial cartilage by standard multiphoton microscopy (Prairie Technologies) with a Ti-sapphire laser (Coherent) at 965 nm through a 25X/NA1.05 water-immersion objective lens (Olympus XL Plan N). 512x512 fluorescent images were acquired at 1 Hz, with each odor trial consisting of 20 seconds of Ringers, 10 seconds of odorant, and finally 50 more seconds of Ringers. After imaging, a high resolution image of the field of view was acquired to aid cell identification. As was done for the *in vitro* assays, the delay time for odor delivery was assessed using RhodamineB dye; these experiments revealed that the maximal lumenal concentration of dye at equilibrium was approximately 10 percent of the initial dye concentration (as assessed via fluorescence intensity). Image registration was accomplished using custom code, and publicly available Python image processing and OpenCV packages. Time-series images were aligned using a feature-based approach that is robust to regional fluorescence intensity fluctuations over the course of the experiment. Images (for alignment purposes) were first contrast enhanced to enable feature detection, and then all frames from a single experiment were registered to a manually chosen target frame in a pairwise manner. For each frame-target pair, positional features, typically corresponding to cell bodies and blood vessels, were then automatically identified using Harris corner detection. Corresponding features were then used to obtain an optimal homography (i.e., projective transformation) between frame and target, and this transform was then applied to the raw images.

Cells from aligned movies of the raw images were identified using a semi-automated approach. First, the centroids of putative cells (i.e., rounded and convex GCaMP3+ objects, sometimes with nuclear exclusion of GCaMP3) in a time-series average projection were manually specified. Cell masks were then generated using a combination of morphological filtering and region-growing. Each cell mask was further refined based on co-fluctuations in the fluorescence of pixels in the cell's vicinity. Independently covarying groups of pixels were first identified using nonnegative matrix factorization. Pixels associated with the cell of interest were then assigned to the current mask; pixels that correspond to adjacent cells were excluded. In the resultant binary mask, each connected component therefore corresponded to a cell; fluorescence time-courses for each cell were then obtained by averaging the pixels in each connected component on a frame-by-frame basis.

Putative necklace cells were defined as cells within the cul-de-sacs that responded to carbon disulfide with at least a 25% increase in average fluorescence and did not respond to DMSO alone. A cell was then categorized as responding to an odor stimulus if its average fluorescence increased at least 25% relative to the previous 10 frames and reached a peak after odor presentation, but not before. For display, plotted traces were smoothed by convolution with a 3-frame rectangular window. For generating the top panels of Figure 6A, the fluorescent signal following vehicle stimulation was subtracted from the fluorescent signal following odorant stimulation, and the resultant image was heat-mapped and overlaid onto a reference image using a custom Python script. For the purposes of data representation these heatmaps are shown as raw change in fluorescence; 100% on this fluorescence scale corresponds to the full dynamic range of image intensity in this acquisition.

Confocal Microscopy and Image Processing:

Slides were imaged under a 63X/NA1.4 oil immersion planar apochromatic objective lens (Zeiss). Digital images were acquired on a Zeiss LSM 510 Meta confocal microscope (Harvard Neurodiscovery Imaging Center). Cyanine and Alexa fluorophores were excited in sequence with an argon laser (488 nm line), a HeNe laser (543 nm), and a second HeNe laser (633 nm). DAPI was imaged with a tuneable Coherent Chameleon Laser at 740 nm in two-photon excitation mode. Emission was detected with standard dichroic mirrors and filter sets. Using Imaris 8.1 software (Bitplane Inc.), multi-channel z-stacks were maximum-intensity-projected into two dimensions and passed through a median filter to remove debris much smaller than structures being assayed.

Adenoviral Infection:

Six week old male mice (Jackson) were anesthetized with 100 mg/kg of Ketamine, 10 mg/kg of Xylazine, and 3 mg/kg of Acepromazine. Infusions of $3\sim 6 \times 10^8$ FFU of human adenovirus serotype 5 encoding Ms4a6c-IRES-hrGFP or Ms4a6d-IRES-hrGFP in the right nostril of mice were performed as previously described (Holtmaat et al., 1996). Mice were subjected to the odor exposure 6-8 days after injections.

Odor Exposure for Phospho-S6 Immunostaining and Quantification of Positive Cells:

8-10 week old C57/BL6 male mice (Jackson Laboratory) were group housed overnight on a reverse light-dark cycle with 3-5 mice/cage. On the day of the experiment, single mice were habituated in fresh cages for 3 hours in the dark. The odor stimulus was then introduced into each cage as 100 μ L of neat odorant blotted onto Whatman paper in 10 cm petri dishes with slits cut into the lid. After 4 hours of odor exposure, animals were sacrificed and their nasal epithelia were dissected and fixed as described above.

To quantify phospho-S6 immunopositive cells, 10-15 images of Pde2a-positive cell enriched cul-de-sacs were acquired with a confocal microscope (see above) for each slide of odor-exposed olfactory epithelial sections. Laser and acquisition parameters were held constant across each experiment. pS6 immunopositive cells were counted manually from the resulting images; a cell was called positive if it showed smooth pS6 fluorescent signal filling up the soma at levels visibly above adjacent tissue and Pde2a-negative cells. Counts from each image were summed to give an estimate of the proportion of activated necklace cells for each animal. In initial experiments with each odorant, the person imaging and counting cells was blind to the identity of the odorant. Because only sulfated steroids have a vehicle control (DMSO alone) these data were analyzed separately from the data in Figure 7A.

The same method was used to quantify MS4A6C-ires-GFP and MS4A6D-ires-GFP virus-infected cells, except that GFP+ cells were sparse and therefore called positive for pS6 and MS4A6C (guinea pig antibody) by finding a GFP+ cell, imaging it under laser excitation, and recording the cell as positive or negative manually. 20 or more GFP+ cells were counted for 3-5 epithelia per odorant, and images of representative cells were acquired and processed as described above.

Table S1, List of taxa used for the phylogenetic reconstructions, Related to Figure 2 and Experimental Procedures

Group	Order	Family	Species	Common name
Monotremata		Ornithorhynchidae	<i>Ornithorhynchus anatinus</i>	platypus
Marsupalia		Dasyuridae	<i>Sarcophilus harrisii</i>	Tasmanian devil
Placentalia	Soricomorpha	Soricidae	<i>Sorex araneus</i>	common shrew
	Carnivora	Mustelidae	<i>Mustela putorius furo</i>	ferret
		Ursidae	<i>Ailuropoda melanoleuca</i>	giant panda
		Canidae	<i>Canis lupus familiaris</i>	dog
	Chiroptera	Vespertilionidae	<i>Myotis lucifugus</i>	little brown bat
	Perissodactyla	Equidae	<i>Equus ferus caballus</i>	horse
	Artiodactyla	Suidae	<i>Sus scrofa domesticus</i>	pig
		Bovidae	<i>Bos taurus</i>	cow
			<i>Ovis aries</i>	sheep
	Cetacea	Delphinidae	<i>Tursiops truncatus</i>	bottlenose dolphin
		Lipotidae	<i>Lipotes vexillifer</i>	baiji
	Macroscelidea	Macroscelididae	<i>Elephantulus edwardii</i>	elephant shrew
	Afrosoricida	Tenrecidae	<i>Echinops telfairi</i>	lesser hedgehog tenrec
	Hyracoidea	Procaviidae	<i>Procavia capensis</i>	rock hyrax
	Cingulata	Dasypodidae	<i>Dasypus novemcinctus</i>	nine-banded armadillo
		Galagidae	<i>Otolemur garnetti</i>	northern greater galago
		Callitrichidae	<i>Callithrix jacchus</i>	common marmoset
	Primates	Hominidae	<i>Gorilla gorilla</i>	gorilla
			<i>Pan troglodytes</i>	chimpanzee
			<i>Homo sapiens</i>	human
	Scandentia	Tupaiaidae	<i>Tupaia chinensis</i>	tree shrew
	Lagomorpha	Ochotonidae	<i>Ochotona princeps</i>	American pika
		Leporidae	<i>Oryctolagus cuniculus</i>	rabbit
	Rodentia	Heteromyidae	<i>Dipodomys ordii</i>	kangaroo rat
		Dipodidae	<i>Jaculus jaculus</i>	jerboa
		Muridae	<i>Mus musculus</i>	mouse
			<i>Rattus norvegicus</i>	rat
		Cricetidae	<i>Cricetulus griseus</i>	Chinese hamster
			<i>Mesocricetus auratus</i>	golden hamster
			<i>Microtus ochrogaster</i>	prairie vole
		Sciuridae	thirteen-lined ground	
			<i>Spermophilus tridecemlineatus</i>	squirrel
		Bathyergidae	<i>Heterocephalus glaber</i>	naked mole-rat
		Chinchillidae	<i>Chinchilla lanigera</i>	chinchilla
		Octodontidae	<i>Octodon degus</i>	degu
		Caviidae	<i>Cavia porcellus</i>	guinea pig

Table S2, List of probe sequences used in Nanostring experiments, Related to Figure 1 and Experimental Procedures

Ms4a1	GCAACCTGCTCCAAAAGTGAACCTCAAAAGGACATCTTCACTGGTGGGCCCCACA CAAAGCTTCTTCATGAGGGAATCAAAGGCTTTGGGGGCTGTCCAA
Ms4a2	ACAGAAAATAGGAGCAGAGCAGATCTTGCTCTCCCAAATCCACAAGAATCCTCCA GTGCACCTGACATTGAACTCTTGGAAGCATCTCCTGCCAAAGCAG
Ms4a3	CCAGGCTTTCAAGGGTTGCCAATCTTCACCGTCACCTGATGTCTGCATTTCCCTGG GTTCTCATCAGATGGCCTGGTGTCTTTAATGCTGATTCTCACC
Ms4a4a	AACCCAAAATCCTTGGGATTGTGCAGATTGTAATCGCCATCATGAACCTCAGCAT AGGAATTATGATGATAATTGCCACTGTGTGCGACCGGTGAAATACC
Ms4a4b	CCTAGGATATTAACACTTCATTGCACTGGCTTTTGAGGTGAATATTAGATTTACTG TAAGTATGTAAGTCAAGCACTTATTAGGTCAACAACACTTCAAC
Ms4a4c	TGGCAAATCTATCTTCTGAACCACTCATTTCTGTGGTCTTAATGGCTCCAATTTGG GGACCAATAATGTTCAATTGTCTCAGGATCCCTGTCAATTGCAGC
Ms4a4d	ACAACCTGGCACTACCATCGTGGTGAAAACCCAGCTCAAGCATACCCACAAATAGA GTCCACATCGAAACTCCACCACATTACTCAAGGATACTGTTTCT
Ms4a5	TGAATTTACTTAGTGCTCTGGGAGCAGCAGCTGGAATCATTCTCCTCATATTTGGC TTCCTTCTAGATGGGGAATTCATCTGTGGCTATTCTCCAGATGG
Ms4a6b	AAACAAAACATAATACCACAAAAACAAATGGAACATATACCGCAGAAGATATGTC TTCATGATAATGCAGAAATTCACACCATCACAGGGTAGCAATGCTT
Ms4a6c	CATGATTCCACAGGTAGTGACCAATGAGACCATCACACGATTTACCAAATGGA ATCAACTTTCCCCAAAAAGACGAGTCCAGCCTACCCAACAGAGG
Ms4a6d	AGTTTGGCTGCTTTAGAGCCTGCCTTGACGCAATGTAAGCTGGCTTTCACACAAC AGACACAACCCAAGATGCTTATCATTTCTTTAGCCCTGAGCCAT
Ms4a7	GCCTCCAATGTAGCAAGCTCTGTTGTTGCCGTCATTGGCCTCTTCCTCTTCACCTAT TGTCTGATAGCCCTGGGGAGTGCTTTCCCACTGTAACCTCAG
Ms4a8a	TGTCACTACAACAATCCAGGTGTGGTCATTCCAAATGTCTATGCAGCAAACCCAG TGGTCATCCCAGAACCACCAACCCAATACCAAGTTATTCCGAAG
Ms4a10	CCTAAGACCTCTCTGAAGGTTCTCTGTGTGATAGCCAACGTTATCAGCTTGTTCTG CGCACTGGCCGGCTTCTTTGTCAATTGCCAAGGACCTCTTCCTGG
Ms4a13	TTTCATGGCTGCTAACACCTCATGTAGGTGCCCATGAGATTCCCATATAACAAGGC ACACCTCATGCATTTTGTGCAAAAGGAAATTCACAACAAGGTGA
Ms4a15	GTGGGAAATCTTGGCTTCGAGAGGTTTCGGAGGTTGTCTTCAAGATCATTAGC ACGGAGAAGCTCAGAATGTTCCAGAATAGACTGGCATTTCAGAGG
Ms4a18	GAATTCATCCTCACTGCATAGCCTCACATTTTGATGCCAGGCTGTCTGCTGCGC CCATTTTCAGAACATGACAATGTTCCCAACCATATTTGGTGGCA
Pde1c	GCTGTAATCGATGCATTGAAGGATGTGGATACGTGGTCTTCGATGTCTTTTCCCT CAATGAGGCCAGTGAGATCATGCACTGAAGTTCATTTTCTATG
Adcy3	CAACAACGGCGGCATCGAGTGTCTACGCTTCTCAATGAGATCATCTCTGATTTTG ACTCTCTCCTGGACAATCCCAAATTCGGGGTCATCACCAAGATC
Cnga2	GCTTGTGGATAATGGAGATCATGTGGGTTGAATTTCTAAGAGCGTGACCTCCTAA GTCTCACAAAGGAATCAGAGAATAGCTAAATTGTCCTTCCTGAGGC
Actb	CAGGTCATCACTATTGGCAACGAGCGGTTCCGATGCCCTGAGGCTCTTTTCCAGCC TTCCTTCTTGGGTATGGAATCCTGTGGCATCCATGAACTACAT
Gapdh	AGGTTGTCTCCTGCGACTTCAACAGCAACTCCCACTCTTCCACCTTCGATGCCGGG GCTGGCATTGCTCTCAATGACAACCTTGTCAAGCTCATTTCTGT
Pde2a	CCACTAGCTTCTCTTCTGTTTTGTTCCCTATGTGTCGTGGGTGGGGGAGGGGGCCA CCTGCCTTACCTACTCTGAGTTGCCTTTAGAGAGATGCATTTT
Car2	TGCCCAGCATGACCCTGCCCTACAGCCTCTGCTCATATCTTATGATAAAGCTGCGT CCAAGAGCATTGTCAACAACGGCCACTCCTTTAACGTTGAGTTT
Golf	ATCGAAGACTATTTCCCGGAGTATGCCAATTATACTGTCCCTGAAGATGCAACAC CAGATGCGGGAGAAGATCCCAAAGTTACAAGAGCAAAGTTCTTTA
Emx1	CAGGCAAGCGACGTTCCCCAGGACGGGCTGCTTTTGCACGGGGCCTTCGCACGCA AGCCCAAGCGGATTCGCACAGCCTTCTCGCCCTCGCAGCTGCTGC

Table S3. List of odorants used for functional imaging experiments, Related to Figure 3 and Experimental Procedures

Alcohols	1-butanol, 2,5-dimethylphenol, eugenol, guaicol, 1-hexanol, isoeugenol, 1-nonanol, 1-octanol, 2-phenylethanol, thymol
Ketones	acetylanilone, acetophenone, 2-butanone, cyclohexanone, 3-decanone, dodecanolactone, 4-heptanone, 2-octanone, 2-pentanone, vanillin
Sulfurs	2,4,5-trimethyl thiazole, TMT, thiophene, tetrahydrothiophene
Acids	formic acid, hexanoic acid, heptanoic acid, ocantoic acid, tiglic acid, valeric acid, isovaleric acid
Esters	allyl cinnamate, amyl acetate, benzyl acetate, cycohexyl aceatate, ethyl benzoate, ethyl propionate, ethyl valerate, ethyl tiglate, piperidine, propyl butyrate
Aldehydes	p-anise aldehyde, butyl formate, butyraldehyde, benzaldehyde, cinnamaldehyde, ethyl formate, heptanal, octanal, propionaldehyde, heptaldehdye
Nitrogenous	2,5-dimethylpyrazine, 2,6-dimethylpyrazine, 2,3-dimethylpyrazine, indole, nicotine, pyrrolidine, pyridine, quinoline
Steroids	4-Androsten-17alpha-ol-3-one sulphate, 5-Androsten-3Beta 17Beta-diol disulphate, 1,3,5(10)-Estratrien-3 17Beta-diol disulphate, 1,3,5(10)-Estratrien-3 17alpha-diol 3-sulphate, 5alpha-pregnen-3alpha-ol-20-one sulphate, 5beta-pregnen-3beta-ol-20-one sulphate, 4-pregnan-11beta 21-diol-3 20-dione 21-sulphate, 4-pregnen-21-ol-3 20-ione glucosiduronate, 1,3,5(10)-Estratrien-3 17Beta-diol 3-sulphate, 4-pregnen-11beta 17,21-triol3 20-dione 21-sulphate
PUFAs	arachidonic acid, docosohexanoic acid, linoleic acid, linolenic acid, nervonic acid, oleic acid, petroselenic acid
Saturated fatty acids	decanoic acid, docosanoic acid, dodecanoic acid, eicosanoic acid, hexanoic acid, myristic acid, octadecanoic acid, octanoic acid, palmitic acid
Terpenes	R-carvone, 1,4-cineole, citral, cintronellal, R-fenchone, E-beta farnesene, geraniol, alpha-ionone, linalool, +-menthone, gamma-terpinene, 1,3-minus-verbenone

Supplemental References

- Anisimova, M., and Yang, Z. (2007). Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites. *Molecular Biology and Evolution* 24, 1219-1228.
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125, 279-284.
- Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., *et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499, 295-300.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* 9, 772-772.
- DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530-1532.
- 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.
- Holtmaat, A.J., Hermens, W.T., Oestreicher, A.B., Gispén, W.H., Kaplitt, M.G., and Verhaagen, J. (1996). Efficient adenoviral vector-directed expression of a foreign gene to neurons and sustentacular cells in the mouse olfactory neuroepithelium. *Brain Res Mol Brain Res* 41, 148-156.
- Katoh, K., and Standley, D.M. (2013a). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30, 772-780.
- Katoh, K., and Standley, D.M. (2013b). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772-780.
- Knight, Z.A., Tan, K., Birsoy, K., Schmidt, S., Garrison, J.L., Wysocki, R.W., Emiliano, A., Ekstrand, M.I., and Friedman, J.M. (2012). Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* 151, 1126-1137.
- Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delpont, W., and Scheffler, K. (2011). A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. *Molecular Biology and Evolution* 28, 3033-3043.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Macmanes, M.D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Front Genet* 5, 13.
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* 30, 1196-1205.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9, 171-181.
- Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676-679.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology*.
- Simossis, V.A., and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* 33, W289-294.
- Sosulski, D.L., Bloom, M.L., Cutforth, T., Axel, R., and Datta, S.R. (2011). Distinct representations of olfactory information in different cortical centres. *Nature* 472, 213-216.
- Tsirigos, K.D., Peters, C., Shu, N., Käll, L., and Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research* 43, W401-W407.