Supplemental Figures

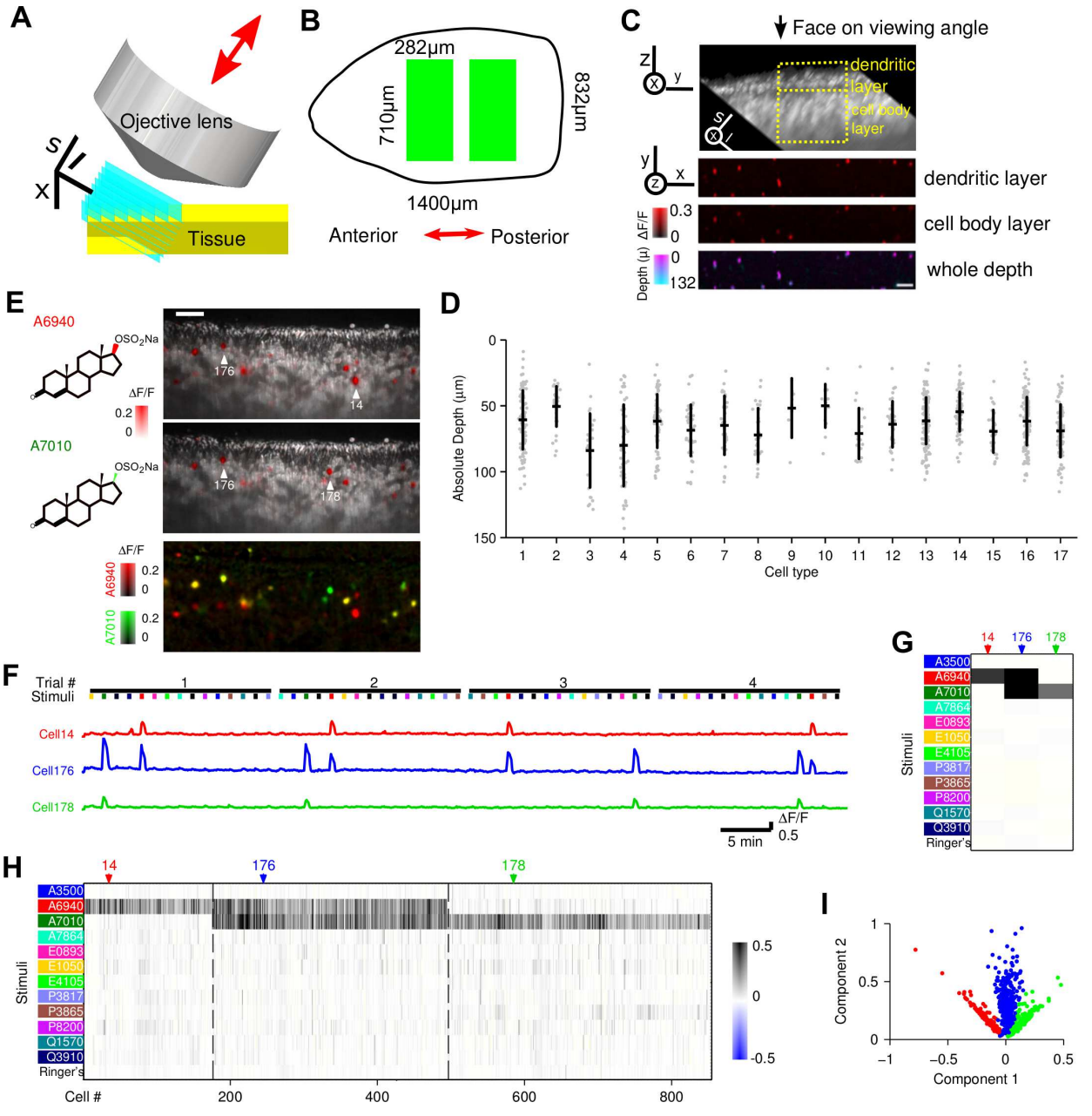

Figure S1. Related to Figure 1.

(A) Schematic of OCPI volumetric scanning. The objective's focal plane is illuminated by a thin sheet of light generated by cylindrical lens (not shown here). Illumination optics are physically coupled to the objective lens such that moving them up and down achieves three-dimensional scanning of the tissue. ℓ: lightsheet propagation direction, s: scan axis.

(B) Geometry of the imaging volume within the whole-mount VNO preparation. In a subset of experiments, two OCPI imaging volumes were collected from a single VNO preparation. Those volumes cover the full tissue depth (125 μm), and two intact regions (710 μm$\times$282 μm rectangles in green) of the tissue when viewed from the top. In a typical intact VNO tissue of 1,400 μm long and 832 μm wide (measurement from a 3-month-old female mouse), each imaging volume covers approximately one-seventh of the whole tissue volume.

(C) Calcium signals at the dendritic knobs and the cell bodies were consistent when whole-mount VNO was stimulated by sulfated steroids. Top, a side view of a VNO imaging volume with surface facing up. In this view, the dendritic knob and cell body of each VSN were aligned vertically. x-ℓ-s represent the coordinate axis related to image acquisition, with ℓ representing the lightsheet propagation direction and s along the scanning axes (see A). x-y-z are coordinate axes for tissue, in which z is along the tissue depth. Rectangles encompass the regions of dendritic layer and cell body layer analyzed in the lower bottom panels. Haze below the cell body layer is a consequence of scattering. Lower panels, maximum intensity projection of cell responses to A3500 at the dendritic layer and cell body layer defined in the top. Red color scale represents the GCaMP2 fluorescence intensity increase ($\Delta F/F$). Bottom, maximum intensity projection of cell responses of the whole VNO depth. Colors code the depth from which each maximum intensity originated, with magenta for the superficial layers (the knobs) and blue for the deep layers (the cell body layer). This representation is independent of any manual segmentation, yet note the consistency of magenta signals with that from manually assigned dendritic layer in the second from the top, and the consistency of blue signals with that from cell body la indicated that most cells showed consistent calcium signals in both layers. While the calcium signals at the dendritic knobs were occasionally more intense, the relatively larger size and more regular shape of cell bodies made the image registration and cell segmentation more reliable. In this study, all the cell responses were measured at the somata.

(D) The depth of cells from an example imaging volume of a male VNO tissue. Each gray

dot represents a single cell. Depth was calculated as the distance of cell body centroid to the tissue surface. A majority of the types situated in the apical zone of the epithelium, except that pregnanolones-responsive types 3 &4 were more basally distributed. Male specific type 8 were close to the apical zone, yet more basal than other apical types. Error bars, standard deviation.

(E–I) Classifying specific VSN functional types using multiple stimulus responses: an example of two steroid enantiomers.

E, a two-dimensional slice showing responses to two stereoisomers, epitestosterone sulfate (A6940, top) and testosterone sulfate (A7010, middle). Red color scale represents the GCaMP2 fluorescence intensity increase ($\Delta F/F$). Bottom, merged image with responses to the two stimuli encoded in different color channels. Cell 14 responded to A6940, cell 178 responded to A7010, and cell 176 responded to both.

F, $\Delta F/F$ as a function of time for the example cells in E. Delivery of different sulfated steroids and Ringer's control is represented by colored bars, with red for A6940 and green for A7010. These cells were specifically activated by A6940 and/or A7010 in all 4 trials. A small fluorescence increase triggered by stimulus delivery, independent of the particular contents of the stimulus, was present in essentially every cell in the recording (uncorrected trace shown; see Experimental Procedures).

G, column representation of the example cells in E and F. For each stimulus, gray intensity is the average of $\Delta F/F$ across 4 trials.

H, the subset of all neurons showing selectivity for either (or both) of these androgens.

I, linear discriminant analysis (LDA) for all the cells in H reveals three largely-separated groupings of cells. Each dot represents a single cell (color indicates cluster identity in G&H); the original 13-dimensional response vectors in H are projected to the first two LDA dimensions.

Scale bar in (C) and (E), 50 μm.

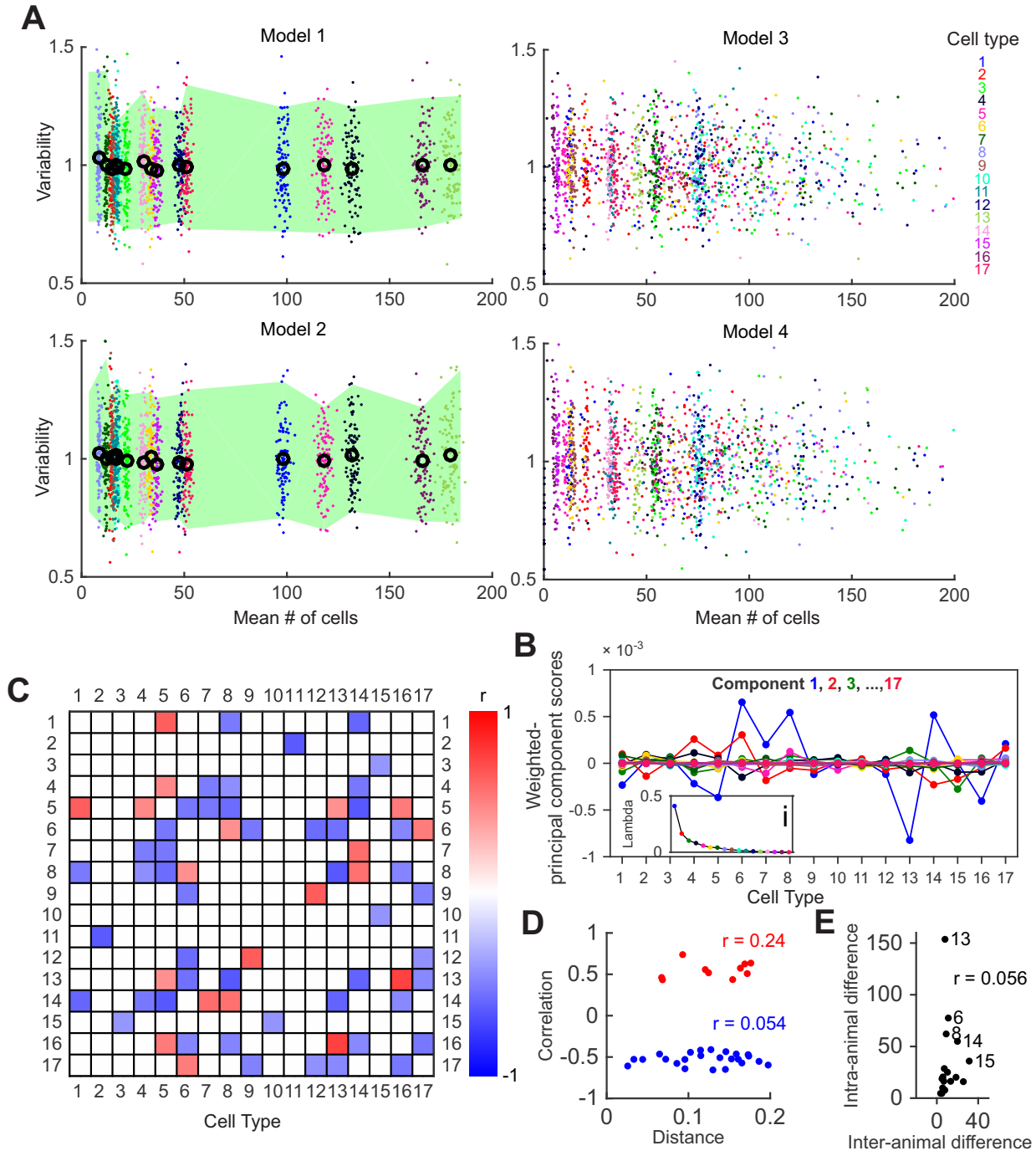
A
Model 1
Model 2
Model 3
Model 4
Variability
Mean # of cells
Cell type
B
Component 1, 2, 3, ...,17
Weighted-principal component scores
Lambda
i
Cell Type
C
r
Cell Type
D
r = 0.24
r = 0.054
Correlation
Distance
E
r = 0.056
Intra-animal difference
Inter-animal difference

Figure S2. Related to Figure 2.

(A) A comparison of multiple models for random sampling of cell counts.

We used simulations to explore 4 models of random sampling:

1. A model in which each of 17 types is Poisson-distributed, with the mean numbers of each type chosen at random.

2. A multinomial model in which neurons are drawn from a large pool of 10,000 neurons (the approximate size of our imaging volume), with 17 "classified" types accounting for an expected (but variable) $\approx 10\%$ of the neurons and "unclassified" neurons accounting for the remaining neurons. (This model adopts the constraint that the cumulative probability of all cell types is 1. It corresponds to the model used in analyses in the main text.)

3. A multinomial model like the previous one, except that we now posit that our 17 types really correspond to 30 different receptor genes, some of which are "pooled" together to make our 17 types.

4. A multinomial model like the previous one, in which we also allow the number of neurons per imaging volume to vary according to a Poisson distribution (so some "experiments" will contain more neurons than others).

In all 4 models, we used the cell type probability from the experimental data to perform the simulation of cell counts. Model 3 and 4 used the same probability per receptor gene that were randomly generated. Hence model 1 and 2 generated very similar mean number for each type, as did model 3 and 4.

For each model, variability (defined in Equation 1 in the main Experimental Procedures) is calculated from 26 simulations to match the 26 experimental data sets in Figure 2. Shown is the variability versus mean number of cells for each type. With the same model parameters, we generated 1000 simulated data sets for each model, and presented the mean ($\circ$) and 95%

confidence interval (green patch) of variability, together with a scatter plot of variability from single simulated data sets. To avoid crowding, only 100 out of 1000 simulated data sets were shown with individual data points. Mean and variance were not calculated for models 3 and 4, because the expected number of cells in each of the 17 types varied with each simulated data set.

Note that the 4 random sampling models produced consistent results for variability in cell counts. The variability is approximately 1, regardless of the mean number of cells of each type and regardless of the details of the particular model.

(B) PCA analysis of normalized cell counts in all types. Principal components are color coded, and scaled by the corresponding λ shown in the inset. Note that the cell counts were normalized by total cell counts in each imaging volume, in order to eliminate variability in the total number of responsive cells from one preparation to the next.

(C) Matrix of correlation coefficients for all 17 cell types, based on normalized cell counts as in B. Pairs with significant correlation ($p < 0.05$) are shown. Correlation/anti-correlation of specific pairs of types are consistent with the largest principal component in B.

(D) The correlation strength in C was not explained by the similarity between cell types. Cell type similarity was measured by distance between the average response of all cells of a given type. There was no correlation (r : Pearson's correlation coefficient, $p = 0.49$ and 0.87) between the pairwise response similarity and the correlation in cell counts, indicating that response similarity does not necessarily determine cell fate.

(E) Inter-animal variability and intra-animal variability were not correlated across cell types. r: Pearson's correlation coefficient ($p = 0.83$).

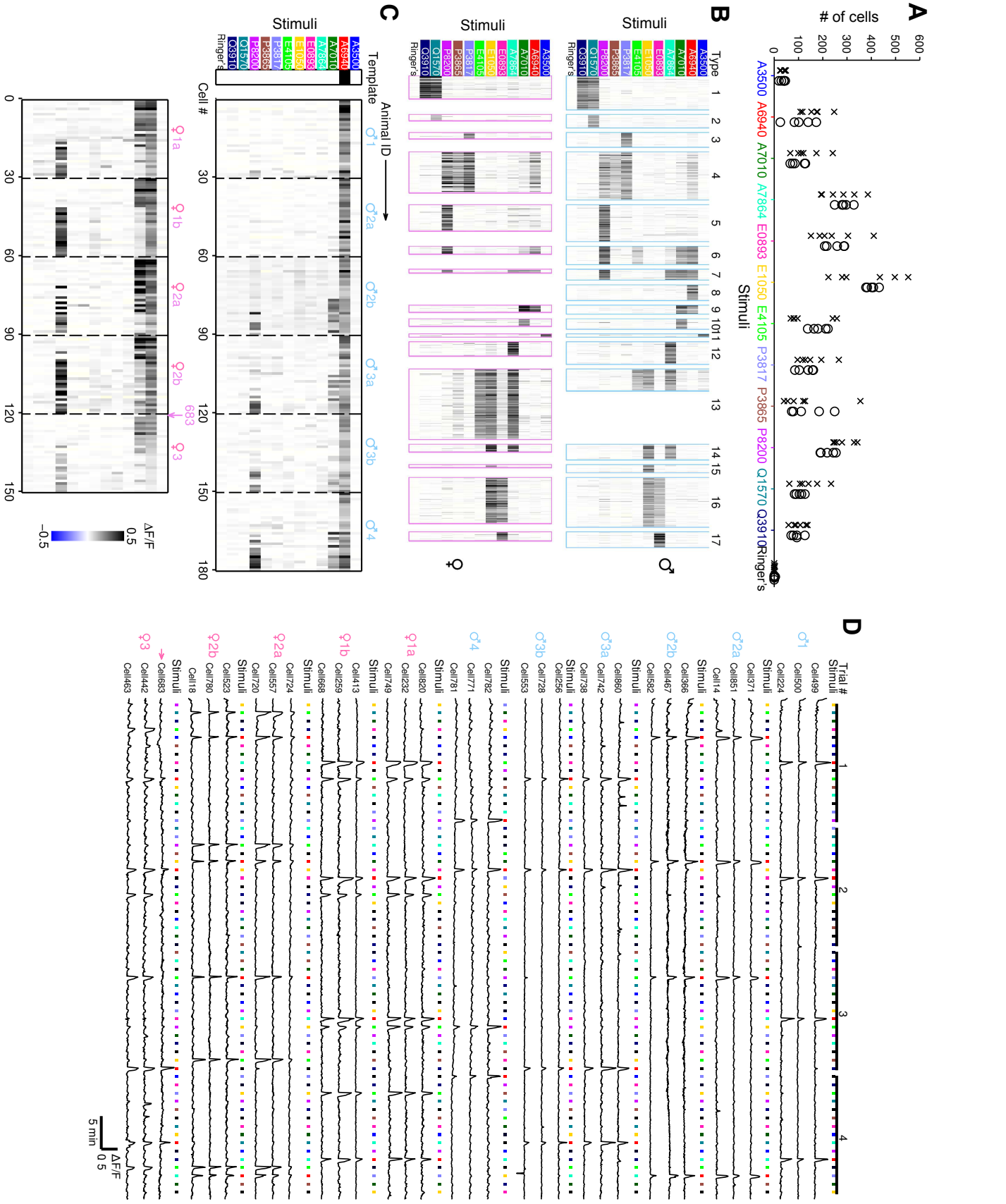
A
# of cells
A3500 A6940 A7010 A7864 E0893 E1050 E4105 P3817 P3865 P8200 Q1570 Q3910 Ringer's
Stimuli
B
Type
Stimuli
♂
♀
C
Stimuli
Template
Animal ID
♂1
♂2a
♂2b
♂3a
♂3b
♂4
Cell #
♀1a
♀1b
♀2a
♀2b
683
♀3
ΔF/F
0.5
−0.5
D
Trial #
Stimuli
♂1
♂2a
♂2b
♂3a
♂3b
♂4
♀1a
♀1b
♀2a
♀2b
♀3
5 min
ΔF/F
0 5

Figure S3. Related to Figure 3.

(A) Number of cells responding to each stimulus in each male (×) and female (○) VNO imaging volume. Each marked point represents one entire imaging volume. On this per-ligand basis, there were no significant differences between male and female VNO imaging volumes (Students' t-test with Ŝidák correction for multiple comparison, familywise significance level $\alpha = 0.05$). Significant sex differences were seen only via decomposition by VSN type (see Figure 3 B).

(B) Physiological types of VSNs in single male (upper) and female (lower) imaging volume. Each thin column of the heatmap represents a single cell; gray intensity is the average induced fluorescence change $\Delta F/F$ across 4 trials. For the 17 main physiological types, all but type 8 were present in both and female mice. Type-8 VSNs were only detected in the male VNO (see Figure 3 B–E).

(C &D) A clustering-free analysis shows that epitestosterone sulfate (A6940)-selective VSNs were specific to male mice.

C, upper Left: a template "artificial cell" responding exclusively to A6940; this template was used to "fish out" the 30 cells with most similar response pattern (smallest Euclidean distance between normalized responses) from each imaging volume. Within each imaging volume (separated by dashed lines), cells were arranged in decreasing order of similarity to the template. Each column represents a single cell; gray intensity is the average $\Delta F/F$ across 4 trials. Each imaging volume is labeled with the animal's identity; multiple imaging volumes collected from the same animal are differentiated by letters. Top, all 6 experiments from male mice (N = 4 animals) contain cells that responded selectively to A6940. Bottom, in females only a single neuron, indicated by the arrow, showed a similar response profile from all the 5 female VNO volumes (N = 3 animals).

D, the $\Delta F/F$ traces of the top three cells in C. Color bars represent different stimuli, red for epitestosterone sulfate (A6940). Cells "fished" from male imaging volumes responded to A6940 exclusively, reproducibly in all 4 trials. Spontaneous activity was detected as small

peaks that were not typically coincident with stimulus onset.

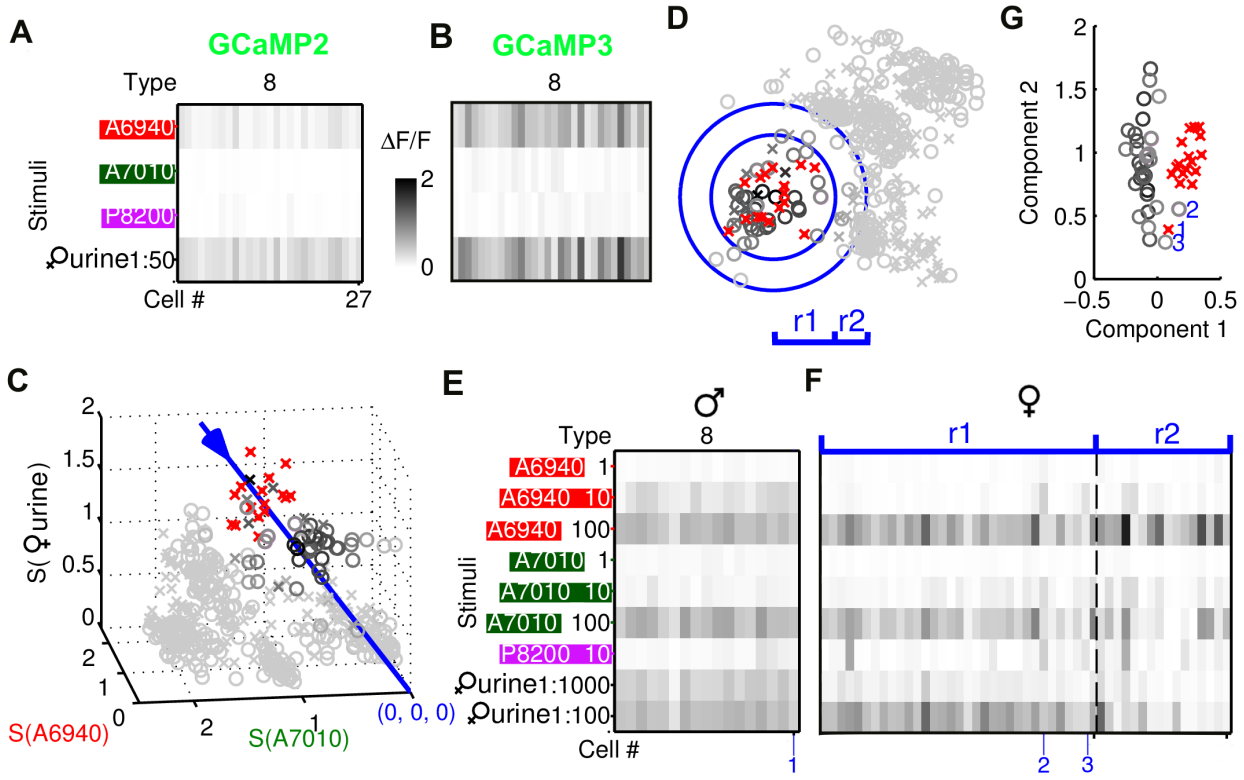

Figure S4. GCaMP3, like GCaMP2, detected type-8 VSNs in male but not female mice, tested with a concentration series of sulfated steroids and female urine. Related to Figure 3.

(A &B) Type-8 VSNs recorded from a male GCaMP2 VNO (A) and a male GCaMP3 VNO (B). The GCaMP2 VSNs came from Figure 1 E. The GCaMP3 VSNs were classified as type 8 according to their responsiveness to A6940 and non-responsiveness to A7010 and P8200, all at 10 μM (see main Experimental Procedures).

(C) For all A6940-responsive cells from GCaMP3 male (×) and female (○) VNOs, sensitivity (S, sum of normalized responses across concentrations 1 μM, 10 μM and 100 μM) for A6940, A7010 and female urine (1:100, and 1:1000 dilution) were plotted in three-dimensional space. Type-8 cells were marked in red (×). Putative type-8 cells, if present, would be close to the marked cells; or putative type-8 cells with lower sensitivity, if present, would be located close

to the blue line that goes through the center of type-8 cluster and the origin. The distance of each cell to the blue axis is coded by the darkness of the marker.

(D) Two-dimensional projection of all cells along the blue axis in C. The concentric blue circles defined by the type-8 center (center of all ×) encompass potential type 8-like cells.

(E &F) The full responsiveness of classified type 8 from males (E) and potential type 8-like cells in the blue circles from females (F), sorted by ascending distance to the type-8 center. Note this approach detected a cohort of cells (mostly within r1 distance) as the top candidates of type 8-like cells. However, these neurons showed similar sensitivity to male type-8 neurons to A7010, but were of approximately ten-fold lower sensitivity to A6940 and female mouse urine, representing a different cell type as shown in G.

(G) Linear discriminant analysis (LDA) of type-8 cells (×) and female neurons (○) within r1 in F. The two LDA components were obtained after an unbiased k-means clustering ($k = 3$) of cells. Note the almost perfect separation between male type-8 neurons and the candidates in females. The three numbered cells in the overlap region are also indicated in E&F; cells 2 and 3, from females, were not type-8 cells as they responded to P8200 (see F), a difference that was not well-captured by these two LDA components.

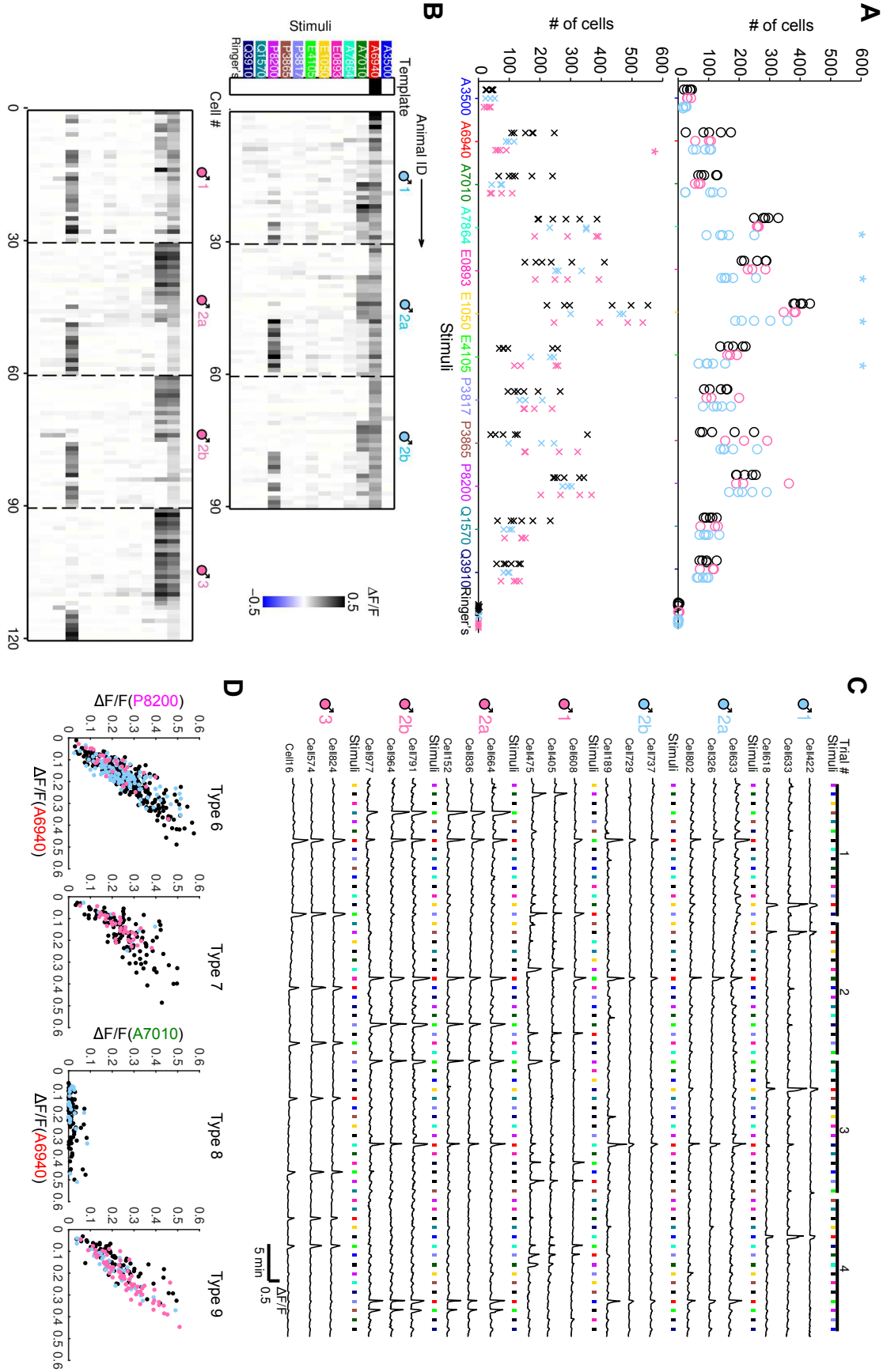
A
# of cells
A3500 A6940 A7010 A7864 E0893 E1050 E4105 P3817 P3865 P8200 Q1570 Q3910 Ringer's
Stimuli
B
Stimuli
Template
Animal ID
Cell #
ΔF/F
C
Trial #
Stimuli
5 min
ΔF/F
D
ΔF/F(P8200)
ΔF/F(A6940)
ΔF/F(A7010)
Type 6
Type 7
Type 8
Type 9

Figure S5. Related to Figure 4.

(A) Full effects of sensory experience were seen only via decomposition by VSN type (see Figure 4 D), with little or no change evident on a per-ligand basis. Number of cells responding to each sulfated steroid in female (upper panel) and male (lower panel) mice as a function of sensory experience. Statistical comparisons were performed among isolation-housed females ($\circ$), females exposed to another female ($\circ$), and females exposed to a male ($\circ$), or among isolation-housed males ($\times$), male-exposed males ($\times$), and female-exposed males ($\times$). $*$: $p < 0.05$, Students' t-test, but none was significant with Ŝidák correction for multiple comparison at familywise significant level $\alpha = 0.05$.

(B–C) Clustering-free analysis shows that male mice exposed to females lost epitestosterone sulfate (A6940)-selective VSNs.

B, upper Left: a template artificial cell, as described Figure S3 C &D. In contrast with isolation-housed males (Figure S3 C) and male-exposed males (upper, 2 animals, 3 imaging volumes), males that had been exposed to females (lower, 3 animals, 4 imaging volumes) did not show A6940-selective VSNs (type-8 VSNs).

C, the $\Delta F/F$ fluorescence traces of the top three cells from each imaging volume. Color bars represent different stimuli, red for epitestosterone sulfate (A6940). Top cells from imaging volumes of male-exposed males exclusively responded to A6940 (similar to males in Figure S3), but top cells from female-exposed males responded to more than just A6940 (as with females in Figure S3 C &D).

(D) Individual cell responses to A6940 and related stimuli, for all A6940 responsive cell types. Sensory experience changed the abundance of type-8 cells, but did not alter the response strength to A6940 relative to other stimuli among cells that responded to more than one ligand. Color code is as described in panel A.

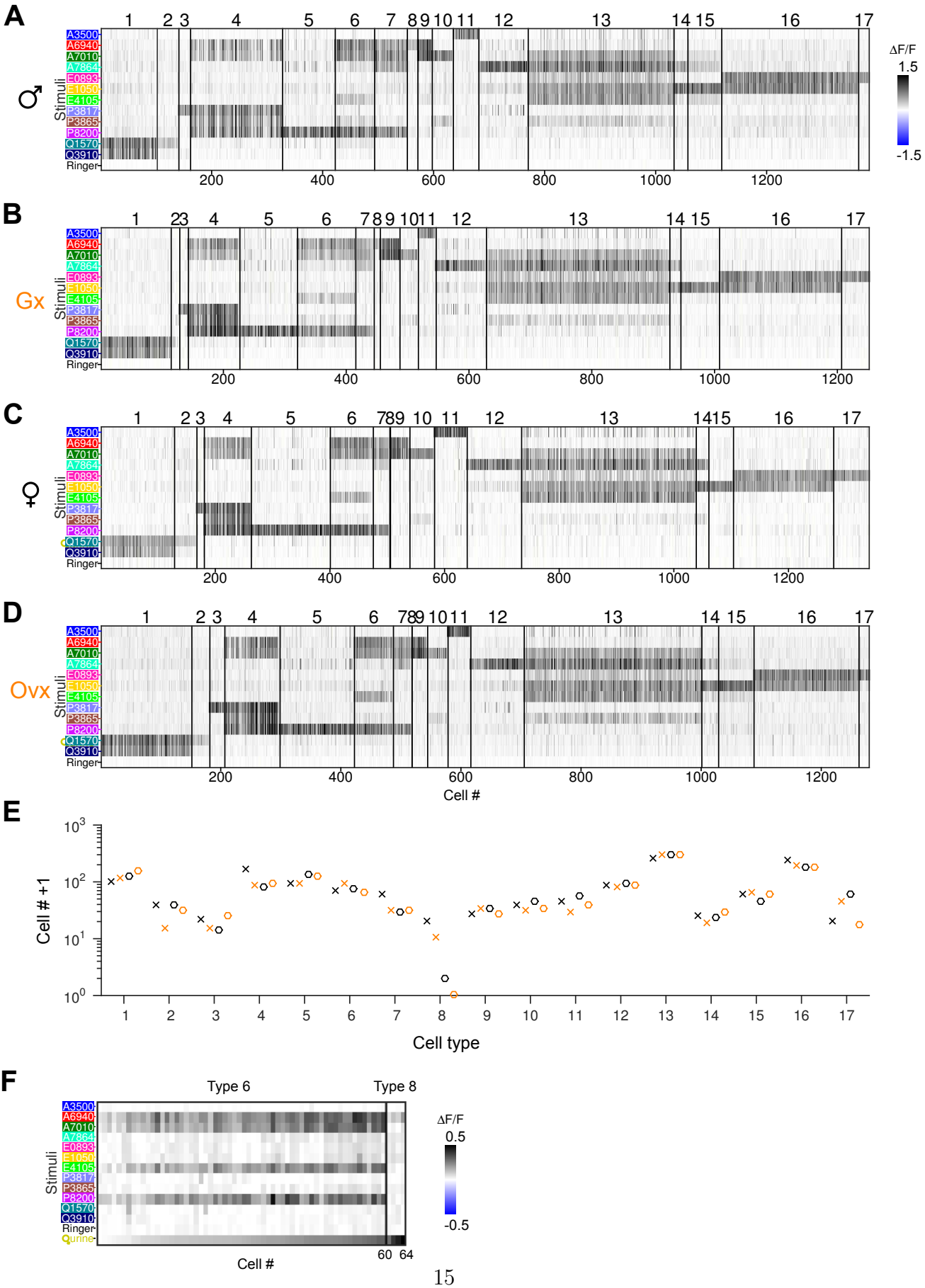
A
B
C
D
E
F
♂
Gx
♀
Ovx
Stimuli
A3500
A6940
A7010
A7864
E0893
E1050
E4105
P3817
P3865
P8200
Q1570
Q3910
Ringer
Qurine
ΔF/F
1.5
-1.5
0.5
-0.5
Cell #
Cell # +1
Cell type
Type 6
Type 8
60 64

Figure S6. Gonadectomy did not change functional VSN types in male or female mice, and a modest number of type-8 cells were detected in females with naris occlusion. Related to Figure 5.

(A-D) Heatmap representation of VSN cell types in normal and gonadectomy GCaMP3 mice. The imaging data from an example animal for each group was fully segmented to assess all 17 VSN types. All the 4 groups possessed all 17 cell types, except both the normal female (C) and the ovariectomized female (D) have 1 or 0 type 8 neurons, respectively.

(E) A scatter plot representation of cell counts in A-D. Note that none of the cell types exhibited dramatic change after hormonal manipulation (see also Figure 5).

(F) Type 6 and 8 cells detected from a female mouse with unilateral naris occlusion. On the closed side, 4 type-8 cells were detected, a modest but noteworthy increase from the open side or control females.

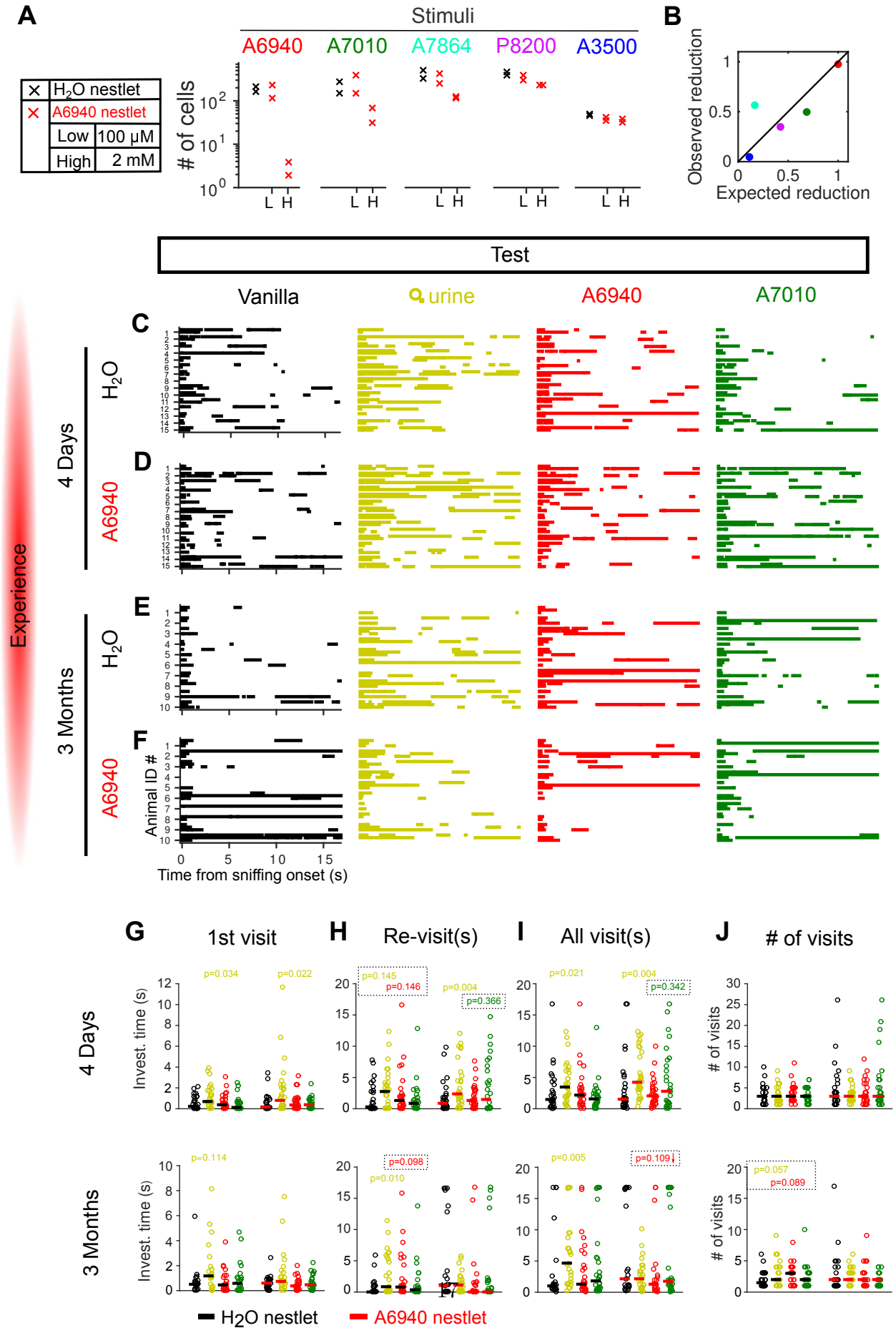
A
Stimuli
A6940
A7010
A7864
P8200
A3500
H2O nestlet
A6940 nestlet
Low 100 µM
High 2 mM
# of cells
L H
B
Observed reduction
Expected reduction
Test
Vanilla
urine
A6940
A7010
Experience
4 Days
3 Months
H2O
A6940
C
D
E
F
Animal ID #
Time from sniffing onset (s)
G
1st visit
H
Re-visit(s)
I
All visit(s)
J
# of visits
Invest. time (s)
p=0.034
p=0.022
p=0.145
p=0.146
p=0.004
p=0.366
p=0.021
p=0.004
p=0.342
p=0.114
p=0.098
p=0.010
p=0.005
p=0.109
p=0.057
p=0.089
H2O nestlet
A6940 nestlet

Figure S7. Physiology and behavior of mice exposed to A6940-soaked nestlets. Related to Figure 7.

(A) The number of cells responsive to each stimulus after long-term exposure to low (100 μM) or high (2 mM) dosage of A6940. Not only did the total number of A6940-responsive cells decrease, but stimuli that co-activated A6940-responsive cells (see Figure 1 E) also decreased in number.

(B) Comparing the reduction in numbers of ligand-responsive neurons (panel A) against the expected reduction from the proportion of A6940-overlapping cells in non-exposed mice, as would be predicted from a complete elimination of all A6940-responsive neurons. Color codes the identity of stimuli as in A.

(C-F) Olfactory investigation of A6940-exposed males to 4 stimuli: control (vanilla only), female urine solid-phase extracts, epitestosterone (A6940), and testosterone (A7010). C-F shows investigatory episodes of animals in each group to the 4 stimuli. For 3-month exposure, there were 10 animals per group, and each animal was tested in 3 trials; for 4-day exposure, there were 15 animals per group and each tested in 2 trials. All the trials were aligned to the onset of the first sniff ($t = 0$). Blocks represent continuous periods of investigation. Blank trials were "failure trials" in which the animals never investigated the cotton swab.

(G-J) Quantification of first visit to the cotton swab (G), re-visit (all visits after the first one, H), all visits (I), and total number of visits (J) for each group of animals in C-F. Trials for all animals were pooled, and plotted as individual data points (note that circle markers are used here only for visualization of overlapping data points and do not indicate sex). Most of the distributions were highly skewed instead of Gaussian, so we adopted statistics robust to outliers. Specifically, the median was calculated and represented as a short thick line for each group, and the Wilcoxon rank-sum test was performed against the control stimulus. All the p-values for significant changes ($p < 0.05$) are labeled on the top of each group; note that p-values in dashed boxes were not significant.

Supplemental Movies

Movie S1: Three-dimensional rendering of a VNO experiment. Related to Figure 1.

Raw fluorescence over time during chemical stimulation of the VNO. Stimulus presentation marked by display of the stimulus name, with periods between flushed with Ringer's. Both spontaneous and stimulated activity can be seen. Red indicates the fluorescence change $\Delta F/F$ upon stimulation, normalized to the baseline period for each stimulus, on a colorscale from 0.03 to 0.3. $\Delta F/F$ was corrected by subtraction of a small global fluorescence change induced by one of the two Ringer's control (labeled with Ringer's_1). An additional inactive stimulus Q4765 was included for this VNO imaging volume. A single trial of each stimulus is shown, corresponding to the second trial in Figure S1 F, which shows 3 single cells from this imaging volume). Note the tissue was stationary during experiment; the "wobble" was introduced computationally only to aid perception of the three-dimensionality of the tissue.

Movie S2: Visualization of male-specific VSNs. Related to Figure 3.

A male VNO imaging volume was rendered as in Movie S1. Red indicates $\Delta F/F$ induced by epitestosterone sulfate (A6940), and green indicates the response to A7010 and P8200. The male-specific physiological type-8 VSNs responded exclusively to A6940, and thus remained red instead of green or yellow after responses to all three stimuli are overlaid. Both red and green were scaled from 0.03 to 0.3.

Movie S3–6: Two-dimensional slices of whole VNO imaging volumes from male (Movie S3), female (Movie S4), male exposed to female cues (Movie S5) and male exposed to male cues (Movie S6). Related to Figure 3 and Figure 4.

Colorization is as in Movie S2, therefore the male-specific epitestosterone sulfate (A6940)-selective cells show as red. All two-dimensional slices, with a step size of 5 μm, were shown, except for edge frames corrupted by drift during the 75-minute recording (motion corrected by registration, but missing data cannot be recovered). All 4 trials and the trial-averaged

images are shown. Note that spontaneous activity also leads to colorization, but is not reproducible across trials. Red cells representing male-specific type (type-8) VSN were only reliably detected in the isolation-housed male mice and males exposed to male cues, not in female mice or males exposed to female cues. Scale bar, 50 μm.

Experimental Procedures

Animals

GCaMP2- and GCaMP3- expressing VNOs (He et al., 2008; Zariwala et al., 2012) (RRIDs: IMSR_JAX: 017755 cross-bred with IMSR_JAX: 017754, IMSR_JAX: 006668 cross-bred with IMSR_JAX: 006668) were obtained from mice singly housed since postnatal day 21, unless otherwise specified. At 12–16 weeks, male and female littermates were recorded on the same experimental day for strict parallel control. For mice with unilateral naris occlusion, right external nares were sutured and irreversibly closed at the beginning of the exposure period, which ranged from postnatal week 3 to week 7. Bilateral ovariectomy or orchiectomy was carried out in 11-week-old mice, and mice were allowed to recover for at least 4 days before use in subsequent experiments. All experimental protocols were in compliance with NIH guidelines and were approved by the Washington University Animal Care and Use Committee.

Reagents and Stimuli

The VNO tissue was dissected and maintained in carboxygenated Ringer's solution (115 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 25 mM $NaHCO_3$, 10 mM HEPES, and 10 mM freshly added D-(+)-glucose) throughout the experiment. The Ringer's solution was carboxygenated by bubbling with 95% O_2/5% CO_2 in a 37 °C water bath for at least 30 min. VNO stimuli, the sulfated steroids (Steraloids Inc., RI, USA) were dissolved in methanol, 1 : 1 methanol / H_2O, or H_2O to 20 mM as stock concentration (kept at 4 °C), and were diluted to a final concentration 10 μM with freshly prepared carboxygenated Ringer's solution. 12 sulfated steroids were used in this study: A3500, 5β-androstan-3α-ol-11, 17-dione sulfate (ketoetiocholanolone sulfate); A6940, 4-androsten-17α-ol-3-one sulfate (epitestosterone sulfate); A7010, 4-androsten-17β-ol-3-one sulfate (testosterone sulfate); A7864, 5-androsten-3β,17β-diol disulfate; E0893, 1,3,5(10)-estratrien-3,17α-diol 3-sulfate (17α-estradiol sulfate); E1050, 1,3,5(10)-estratrien-3,17β-diol disulfate (17-estradiol disulfate); E4105, 4-estren-17β-ol-3-one sulfate (nandrolone sulfate); P3817, 5α-pregnan-3α-ol-20-one sulfate (allopregnanolone

sulfate); P3865, 5α-pregnan-3β-ol-20-one sulfate (epiallopregnanolone sulfate); P8200, 5β-pregnan-3β-ol-20-one sulfate (epipregnanolone sulfate); Q1570, 4-pregnen-11β, 21-diol-3,20-dione 21-sulfate (corticosterone 21-sulfate); Q3910, 4-pregnen-11β, 17,21-triol-3,20-dione 21-sulfate (hydrocortisone 21-sulfate). The two negative control stimuli were Ringer's solution with 0.05% methanol as a vehicle control. For behavioral tests, A6940 and A7010 were further purified with silica gel thin layer chromatography (TLC) (Sigma-Aldrich Co., MO, USA), dissolved in ddH_2O to stock concentration 20 mM and kept at -80 °C before use. Urine from male and female mice (2- to 5-month-old) of C57BL/6 background were collected over periods of 2 weeks with liquid nitrogen (Nodari et al., 2008) (CeeKay Supply) and pooled by sex. Solid-phase extracts were prepared with Strata-X 33u polymeric reversed phase column (Nodari et al., 2008) (Phenomenex Inc., CA, USA).

Imaging 10, 000 neurons simultaneously from whole-mount VNO by OCPI. Related to Figure 1.

We imaged the whole-mount vomeronasal organ (VNO) at single cell resolution using a light sheet-based imaging technique, Objective-Coupled Planar Illumination (OCPI) microscopy (Holekamp et al., 2008). OCPI microscopy works by illuminating the objective's focal plane with a thin sheet of light; the objective together with the light sheet moves up and down at a titled angle to repeatedly scan the tissue along the objective's z-axis (Figure S1 A). Using OCPI microscopy, we scanned a VNO volume 710 μm$\times$125 μm$\times$282 μm, accounting for approximately one seventh of the total volume of an intact VNO tissue (Figure S1 B). In the cell body layer, the cell density of sensory neurons in a 4-month-old VNO was previously(Hedewig, 1980) measured to be $\sim 1.01 \times 10^6$ /mm^3 (averaged across sexes)(Wilson and Raisman, 1980). We estimated the total number of vomeronasal sensory neurons (VSNs) contained in a single OCPI imaging volume as 1.01×10^6 /mm$^3 \times 710\,\mu$m$\times 125\,\mu$m$\times 282\,\mu$m$\times 0.9 = 18,200$ (a factor of 0.9 came from the exclusion of edges of imaging volumes during image registration). Because 1) there is a 43% increase in sensory cell number between 1 and 4 months of age(Wilson and Raisman, 1980), and 2) males have a larger VNO volume and

number of sensory neurons than females (Segovia and Guillamón, 1982), we estimated that our single imaging volumes on average provided over 10,000 VSNs for analysis.

Calcium Imaging of Whole-mount VNO by OCPI Microscopy

Mice were euthanized with CO_2 and decapitated, and the VNO was quickly removed into ice-cold carboxygenated Ringer's solution. Under a stereomicroscope, the two VNO tissues were removed from the bony capsule, and each intact neuroepithelium was separated from the blood vessel. The flatted VNO epithelium was then adhered to a nitrocellulose membrane (0.45 μm, Millipore Co., MA, USA), with the dendritic layer facing up. This whole-mount VNO preparation could be kept in ice-cold carboxygenated Ringer's solution for up to 3 hrs before physiological imaging as described (Xu and Holy, 2013).

Imaging was performed on a custom objective-coupled planar illumination (OCPI) microscope (Holekamp et al., 2008) with 488 nm light sheet (thickness: 5 μm) illumination. To reduce phototoxicity, laser power was minimized to a level corresponding to ~10,000 photons per cell in each stack to achieve a shot noise-limited sensitivity to $\Delta F/F$ ~1%. Typically, this resulted in $<$ 0.05 mW laser power when imaging GCaMP2 VNO tissue. The VNO preparation was placed in a custom imaging chamber with continuous superfusion of carboxygenated Ringer's solution (temperature: 35 °C; flow rate: 2.2 ml/min) alternating with stimuli. The stimuli and Ringer's flush were delivered through a 16-channel superfusion system (Automate Scientific Inc., CA, USA) pressurized with helium-balanced 40% O_2 and 3% CO_2. The superfusion system was computer controlled by custom software to synchronize with the imaging acquisition system (*Imagine*, http://holylab.wustl.edu/software.htm). Before recording, the VNO was superfused with Ringer's for at least 30 min to acclimate. We found that stimuli resident in the superfusion system's PTFE/silicone tubing for several tens of minutes induced a global increase in fluorescence intensity of the VNO tissue, presumably due to loss of gases and the accompanying pH changes; to reduce this artifact, we expelled a small amount of each solution right before recording. During recording, each stimulus was applied to the tissue for 25 s, following by Ringer's flush for 50 s. The order of 12 stimuli

together with 2 Ringer's controls were randomized within a cycle ("trial"). All stimuli were repeated in 4 cycles, together amounting to ~ 75 min of recording on each VNO imaging volume. The dimension of one tissue volume was $710\,\mu\text{m} \times 125\,\mu\text{m} \times 282\,\mu\text{m}$, and OCPI scanned the volume every 2 s, with a 3 s pause to save the data.

Image Registration and Segmentation

VNO tissue normally exhibited small but detectable shifting and deformity (swelling, shrinking, warping, etc.) during long hours of recording. Such tissue movement was corrected by custom three-dimensional nonrigid registration software written in MATLAB, C++, and Julia (Bezanson et al., 2012). Manual image segmentation was performed in a custom GUI in MATLAB to identify individual cells. Briefly, cells exhibiting induced and/or spontaneous activity were highlighted in images showing the change in fluorescence ($\Delta F/F$). Each highlighted region in the VNO cell body layer, but not the dendritic knob layer, was circled as a region of interest (ROI) representing one cell, avoiding double-counting in adjacent optical sections. Adjacent cells normally were clearly differentiated from each other by their distinctive responses to different stimuli. For a typical VNO sensory neuron, the ROI size was $4\,\mu$m in radius. The primary site of sensory transduction is in the dendritic knob, which in principle makes the knob attractive for quantifying responses; however, the small size and close packing of knobs made them more vulnerable than the cell bodies to image registration artifacts. As all observations indicated that the amplitude of responses in the knobs and somas were very well correlated (e.g., Figure S1 C), we chose to center our ROIs on the cell somata.

Analysis of Cellular Responses

Fluorescence intensity of a particular cell in a particular image stack was measured as mean intensity of all pixels within its ROI. Cellular $\Delta F/F$ responses were calculated as the time-weighted fluorescence change upon stimulation (Turaga and Holy, 2012), using the average of the 20 s pre-stimulus period as the background. The component of the response due to Ringer's artifact was subtracted on each trial by performing robust regression of

all neurons' responses to the response caused by one of the two Ringer's control tubes. The second Ringer's control (shown as "Ringer's" in all figures) was treated identically to the other stimuli, and thus served as an independent negative control. To distinguish spontaneous activity, we tested reproducibility over all 4 trials by the one-tail Student's t-test against the second Ringer's control. Only cells responding to at least one stimulus with 1) p-value < 0.05, and 2) $\Delta F/F > 0.03$ were counted as responsive cells and kept for further analysis.

Clustering of Physiological Cell Types

Physiological cell types were identified by clustering all responsive cells pooled from GCaMP2 imaging volumes collected in this study. Each cell was a data point in the 13-dimensional space with the value of responses to 12 stimuli and a Ringer's control. All the data points were clustered by a variant of the mean shift clustering algorithm (Comaniciu and Meer, 2002), a method with sensitivity high enough to reliably detect clusters with size of no less than 10 data points. Statistical significance of clusters was assessed by performing clustering in 10,000 bootstrapped datasets (sampling with replacement), and mean shift peaks that fell into the same cluster with frequency >= 0.05 were merged. The above strategy robustly detected clusters with fine in-cluster homogeneity and clear between-cluster separation. After clustering, the pooled cells were split back into their individual imaging volumes. Clusters showed substantial consistency across datasets, with the main source of variability being the absolute intensity of responses. No cells were left unclassified.

Cells recorded from GCaMP3 VNOs were classified by reference to the much larger GCaMP2 dataset, assigning each to the GCaMP2 cell with the smallest Euclidean distance in the space of normalized responses. This approach was not effective for any new cell types, but it identified GCaMP3 cells belonging to the known types, like type-8 in Figure S4 B. GCaMP3 cells were also analyzed without prior classification (Figure S4 C–G).

Variability Analyses

For any particular cell type and set of experiments, the variability of each cell type was

calculated as the standard deviation of the square root of the number n of cells of each type:

$$\text{variability}_j = 2\sigma(\{\sqrt{n_{ij}}\}_i), \tag{1}$$

where j is the cell type $(1, \ldots, 17)$, σ is the standard deviation, and n_{ij} is the number of cells of type j in the ith imaging volume. The standard deviation was calculated across imaging volumes. The $\sqrt{n}$ scaling transformation is "variance stabilizing" for a Poisson sampling model (which closely approximates a multinomial model when the number of classes is large, see also Figure S2 A), and for such a model Equation (1) has a value very near 1 independent of the mean value of n (McCullagh and Nelder, 1989). In other words, the Variability as defined by Equation (1) does not grow with n. We adopted the variance-stabilizing transformation specifically because it greatly simplifies the comparison of cell types that contain wildly-differing numbers of neurons (as we observed in Figure 1 D &E).

This $\sqrt{n}$ scaling transformation was used, where appropriate, throughout our analyses of variability as well as statistical significance of cell counts.

Simulating Random Sampling of Cell Counts with Multinomial Model

For each cell type, random sampling was simulated using a multinomial model with probabilities assigned from the mean frequency of each of the 17 functional types, with the number of draws matching the number of experimental points. To obtain the distribution of the variability, each simulation was repeated $1,000$ times and the 95% quantile used to estimate the uncertainty of the variability (Figure 2 B). For plotting 6 different groups at once (Figure 4G), it is impractical to show error bars for 6 categories, so the normalized variability for each group was calculated. It is defined as $(Var_{obs} - 1)/(Var_{ran97.5\%} - 1)$, in which Var_{obs} is the observed variability (from Equation 1), and $Var_{ran97.5\%}$ is the two-tailed 95% quantile upper bound variability from multinomial random sampling. In this way, the normalized variability of any random sampling has a 97.5% chance to be at or below 1.

Analysis of Inter/intra-animal Differences

Inter/intra- animal differences (Figure 2E) were analyzed using non-overlapping pairs

of imaging volumes from the same mice. For each animal, the pair of volumes (data sets) contained one from the anterior VNO and the other from the posterior VNO. The observed intra-animal difference was calculated by summing the absolute differences in $\sqrt{n}$ for each type across all the true pairs (or for the total number of responsive neurons when analyzing variability by total number of cells). In the permutation test, to obtain the distribution of inter-animal differences, VNO data sets from the pairwise data sets were shuffled. Shuffling was conducted across pairs but within anterior and posterior data sets, respectively. This controlled for hypothetical differences between anterior and posterior VNO (no obvious location-dependent differences were noted). Only permutations lacking any true pairs were accepted, resulting in 265 permutations using 6 true pairs. In Figure 2 E, each data point represents the sum of inter-animal pairwise differences for a particular permutation. The rank of the observed intra-animal difference in the 265 permuted values was reported as the p value.

Cell Depth Analysis

The VNO surface of each imaging volume was identified using an automated algorithm (Turaga and Holy, 2012). For each ROI, the distance of ROI centroid to the tissue surface was calculated, and further corrected for the imaging tilt (45 °).

Long-term Chemical Exposure

Custom double-layer stacking cages, modified from conventional mouse cages (*Max 75*, dimensions: 18.4 cm W ×29.2 cm D×12.7 cm H), were used to provide persistent exposure to natural chemical cues from other animals. Briefly, a conventional mouse cage with the bottom replaced by a stainless wire mesh (diameter: 1.6 mm; grid size: 1 cm×1 cm) was stacked on another cage with the normal wire bar lid filled with feed and water gel pack. The two layers were locked together by a pair of screws on the mesh and bar lid, and were easily disassembled for refilling food/water and twice-weekly cage changes. The wire mesh and bar lid forming the inter-layer divider provided sufficient open area to allow urine, feces and other discharge from the upper cage fall into the lower cage, as verified by direct inspection.

However, this inter-layer divider with small grid size and $\sim 2\,\mathrm{cm}$ spacing largely restricted physical contact between animals housed in the two layers. We housed the test animal in the lower layer with another animal of the same or opposite sex in the upper layer. The exposure started from weaning on postnatal day 21 and lasted 9–13 weeks unless otherwise specified. A few experiments in which dirty bedding from male/female mice was supplied manually were also conducted; there were no obvious differences between this procedure and housing in stacking cages. Mice from both paradigms are reported in Figure 4 and Figure S4.

Long-term application of female urine or sulfated steroids to mice were conducted with nestlets (standard size $2\,\mathrm{cm} \times 2\,\mathrm{cm}$, 1/3 of normal thickness). For each animal housed in the normal cage, a piece of nestlet was first soaked in 1 ml of raw female urine or sulfated steroid solution ($100\,\mu\mathrm{M}$, $1\,m\mathrm{M}$ or $2\,m\mathrm{M}$ in $\mathrm{ddH_2O}$), and then placed into a 35-mm petri dish lid inside the cage. Animals vigorously chewed the nestlets, thereby delivering the chemicals into their VNOs as verified by fluorescent dye tracing after a 2-hour presentation (data not shown). A fresh nestlet with stimulus was presented to each animal on a daily basis for 2–3 months.

Behavioral Assay

Sniffing episodes were recorded as described (Guo and Holy, 2007). Animals were acclimatized to the testing environment with a 15 min/day episode in the testing chamber for 3 consecutive days before the first trial. Trials were conducted after lights-out under red illumination. A day's test consisted of 15 min of acclimatization followed by four 210-s trials (for 4 testing stimuli), using a $\sim$20 min gap between trials. Each trial was a presentation of a cotton swab with $20\,\mu\mathrm{l}$ stimulus solution. Swabs, mounted in a custom beam-break sensor (Guo and Holy, 2007), were introduced through a hole in the lid of the chamber at the beginning of the recording. An animal approaching the cotton swab broke the infrared beam and triggered a voltage drop of the TTL signal, which was captured at 1 kHz using custom software (Guo and Holy, 2007). Bouts of investigation lasting less than 10 ms—previously found to be false optical triggers, for example, by detector movement or changes in lighting

(Guo and Holy, 2007)—were discarded in data analysis. Bouts of chewing, tugging, and climbing the detector, when present, tended to occur later in the trial and were not typical within the first 20 s after initial contact.

Because vomeronasal stimuli are nonvolatile and do not attract attention on their own, a neutral volatile cue, vanilla diluted 1:1000 in water, was freshly prepared as a vehicle solution for the testing stimuli. This effectively promoted initial contact to the cotton swab when mice freely explored the enclosed testing chamber. TLC purified A6940 and A7010 stock solutions were diluted with vehicle solution to a final concentration 100 μM. For female urine extracts (prepared by C_{18} extraction as described previously (Fu et al., 2015), vanilla was also freshly added into the 1:1 extracts from raw urine. The order of the stimuli was randomized under the condition that vehicle solution was not the first or last trial.

Statistics

Student's t-tests or multiple independent comparisons were used for specified experiments. Multiple independent comparisons were performed by individual Student's t-tests followed by multiple comparison correction to control for type I errors. A conservative method, Šidák correction, was used. For n independent comparisons, the *corrected threshold* α_* for statistical significance was calculated as $\alpha_* = 1 - (1 - \alpha)^{1/n}$, where α is the naïve (isolated) threshold for significance. Here, $\alpha = 0.05$; for $n = 17$, this results in $\alpha_* = 0.003$. For cell counts n with Poisson or multinomial distribution, $\sqrt{n}$ is approximately Gaussian and of uniform variance, so square-root transformed data was used in Student's t-tests. The Wilcoxon rank-sum test was used for analyzing behavioral data.

Supplemental References

Bezanson, J., Karpinski, S., Shah, V. B. and Edelman, A. (2012). Julia: A Fast Dynamic Language for Technical Computing. CoRR *abs/1209.5145*.

Comaniciu, D. and Meer, P. (2002). Mean Shift: A Robust Approach Toward Feature Space

Analysis. IEEE Trans. Pattern Anal. Mach. Intell. *24*, 603–619.

Fu, X., Yan, Y., Xu, P. S., Geerlof-Vidavsky, I., Chong, W., Gross, M. L. and Holy, T. E. (2015). A Molecular Code for Identity in the Vomeronasal System. Cell *163*, 313–323.

Guo, Z. and Holy, T. E. (2007). Sex selectivity of mouse ultrasonic songs. Chem Senses *32*, 463–73.

He, J., Ma, L., Kim, S., Nakai, J. and Yu, C. R. (2008). Encoding gender and individual information in the mouse vomeronasal organ. Science *320*, 535–8.

Hedewig, R. (1980). [Comparative anatomical studies on the organs of Jacobson of Nycticebus coucang Boddaert, 1785 (Prosimiae, Lorisidae) and Galago crassicandatus E. Geoffroy, 1812 (Prosimiae, Lorisidae).II.Galago Crassicaudatus]. Gegenbaurs Morphol Jahrb *126*, 676–722.

Holekamp, T. F., Turaga, D. and Holy, T. E. (2008). Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. Neuron *57*, 661–72.

McCullagh, P. and Nelder, J. (1989). Generalized Linear Models. 1989 London. Chapman and Hall.

Nodari, F., Hsu, F.-F., Fu, X., Holekamp, T. F., Kao, L.-F., Turk, J. and Holy, T. E. (2008). Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. J Neurosci *28*, 6407–18.

Segovia, S. and Guillamón, A. (1982). Effects of sex steroids on the development of the vomeronasal organ in the rat. Brain Res *281*, 209–12.

Turaga, D. and Holy, T. E. (2012). Organization of vomeronasal sensory coding revealed by fast volumetric calcium imaging. J Neurosci *32*, 1612–21.

Wilson, K. C. and Raisman, G. (1980). Age-related changes in the neurosensory epithelium of the mouse vomeronasal organ: extended period of postnatal growth in size and evidence for rapid cell turnover in the adult. Brain Res *185*, 103–13.

Xu, P. S. and Holy, T. E. (2013). Whole-mount imaging of responses in mouse vomeronasal neurons. Methods Mol Biol *1068*, 201–10.

Zariwala, H. A., Borghuis, B. G., Hoogland, T. M., Madisen, L., Tian, L., De Zeeuw, C. I., Zeng, H., Looger, L. L., Svoboda, K. and Chen, T.-W. (2012). A Cre-dependent GCaMP3 reporter mouse for neuronal imaging in vivo. J Neurosci *32*, 3131–41.