



Published in final edited form as:

J Comp Neurol. 2012 August 15; 520(12): 2608–2629. doi:10.1002/cne.23052.

Genomics of Mature and Immature Olfactory Sensory Neurons

Melissa D. Nickell¹, Patrick Breheny^{2,3}, Arnold J. Stromberg³, and Timothy S. McClintock^{1,*}

¹Department of Physiology, University of Kentucky, Lexington, Kentucky 40536-0298

²Department of Biostatistics, University of Kentucky, Lexington, Kentucky 40506-0027

³Department of Statistics, University of Kentucky, Lexington, Kentucky 40506-0027

Abstract

The continuous replacement of neurons in the olfactory epithelium provides an advantageous model for investigating neuronal differentiation and maturation. By calculating the relative enrichment of every mRNA detected in samples of mature mouse olfactory sensory neurons (OSNs), immature OSNs, and the residual population of neighboring cell types, and then comparing these ratios against the known expression patterns of >300 genes, enrichment criteria that accurately predicted the OSN expression patterns of nearly all genes were determined. We identified 847 immature OSN-specific and 691 mature OSN-specific genes. The control of gene expression by chromatin modification and transcription factors, and neurite growth, protein transport, RNA processing, cholesterol biosynthesis, and apoptosis via death domain receptors, were overrepresented biological processes in immature OSNs. Ion transport (ion channels), presynaptic functions, and cilia-specific processes were overrepresented in mature OSNs. Processes overrepresented among the genes expressed by all OSNs were protein and ion transport, ER overload response, protein catabolism, and the electron transport chain. To more accurately represent gradations in mRNA abundance and identify all genes expressed in each cell type, classification methods were used to produce probabilities of expression in each cell type for every gene. These probabilities, which identified 9,300 genes expressed in OSNs, were 96% accurate at identifying genes expressed in OSNs and 86% accurate at discriminating genes specific to mature and immature OSNs. This OSN gene database not only predicts the genes responsible for the major biological processes active in OSNs, but also identifies thousands of never before studied genes that support OSN phenotypes.

Keywords

smell; development; neural differentiation; gene expression; microarray

The main olfactory epithelium is an extraordinary model for the study of neuronal differentiation and the events that regulate it (Cau et al., 2002; Schwob, 2002). To come in contact with volatile odorant chemicals, mammalian olfactory sensory neurons (OSNs) are

© 2012 Wiley Periodicals, Inc.

*CORRESPONDENCE TO: Timothy S. McClintock, Department of Physiology, University of Kentucky, 800 Rose St., Lexington, KY 40536-0298. mcclint@uky.edu.

Additional Supporting Material may be found in the online version of this article.

necessarily exposed to external stressors such as absorbed chemicals (including the odorants themselves), pathogens, and inspired particulate matter (Hinds et al., 1984; Monath et al., 1983). Although OSNs have some capacity to defend themselves against chemical stressors (Sammata and McClintock, 2010) and neighboring sustentacular cells express biotransformation enzymes that are thought to detoxify and help clear absorbed chemicals (Breipohl et al., 1974; Dahl et al., 1982, 1988; Rafols and Getchell, 1983), this tissue's ultimate defense resides in its ability to replace damaged OSNs. Because OSN replacement is lifelong and continuous, the olfactory epithelium is a uniquely dynamic place where mature OSNs always coexist with immature OSNs and the proliferative basal cells that serve as their progenitors (Graziadei et al., 1978; Graziadei and Metcalf, 1971).

The differentiation of OSNs requires conversion of morphologically simple progenitor cells into ciliated bipolar neurons whose exquisitely specific synaptic connections in the glomeruli of the olfactory bulb are determined by the single odorant receptor protein expressed by each OSN (Mombaerts, 2006). The OSNs arise out of a complex population of progenitor cells whose neural fate has been linked to expression of the transcription factors Meis1, Meis2, Sox2, Pax6, and Ascl1 (Mash1), and then further supported by expression of Neurog1 and Neurod1 (Cau et al., 2000, 2002; Gokoffski et al., 2011; Guo et al., 2010; Tucker et al., 2010). Late-stage Neurog1⁺ progenitor cells transition into nascent OSNs, marked by expression of Cxcr4 and Dbn1, which then quickly begin to express Gap43, the canonical marker of immature OSNs (McIntyre et al., 2010). Further differentiation of immature OSNs should be driven by additional changes in patterns of gene expression, consistent with the developmental defects observed after null mutations of several transcription factors (Bishop et al., 2003; Cheng and Reed, 2007; Hirata et al., 2006; Hirota and Mombaerts, 2004; Kajimura et al., 2007; Kolterud et al., 2004; Laub et al., 2005; Levi et al., 2003; Long et al., 2003; McIntyre et al., 2008; Sammata et al., 2010b; Shimizu and Hibi, 2009; Theriault et al., 2005; Wang et al., 1997, 2004; Watanabe et al., 2009). These data are still far short of a complete understanding of the differentiation process, however. A significant step toward a complete understanding would be the identification of all genes expressed by OSNs and determining the differential expression of these genes between immature and mature OSNs.

Beginning with mature and immature OSNs, we are using the olfactory epithelium as a model for mapping the expression of every gene in every major cell type in a neural tissue. This strategy emphasizes cells over genes, seeking to understand the cells as systems by identifying all expressed genes and allowing biological processes used by the cells to emerge from analyses of the known functional relationships between the products of these genes. Furthermore, by associating genes with a cell type and therefore also with the functions of the cell, this approach leads to hypotheses about the functions of genes whose products have never been studied (McClintock et al., 2008). Our approach is a logical extension of strategies that focus on genes by physically mapping the expression of all genes but lack the sensitivity or capacity to produce a comprehensive catalog of the genes expressed by specific cell types (Diez-Roux et al., 2011; Dougherty et al., 2010; Doyle et al., 2008; Lein et al., 2007).

Previous analysis of mature OSNs (Sammeta et al., 2007) identified expression of ~10,000 genes whose functional relationships further identified biological processes that were either wholly consistent with known functions of OSNs (e.g., ciliogenesis, ion transport, and synaptogenesis) or indicative of understudied properties of OSNs (e.g., chromatin remodeling, transcriptional regulation, and cell adhesion-like functions). In order to determine which of these gene products and the biological processes they form were developmentally regulated and distinctly associated with OSN maturity or immaturity, we have extended our approach to immature OSNs. Only a few dozen genes expressed in immature OSNs were previously known, but we have now identified 8,573 genes expressed in immature OSNs, including 847 genes whose expression was specific to immature OSNs. Similarly, we identified 7,045 genes expressed in mature OSNs, including 691 whose expression was specific to mature OSNs. These mature OSN-specific mRNAs tend to encode proteins with functions related to odor detection, ion transport and ciliogenesis, whereas functions overrepresented in immature OSN-specific mRNAs were related to axonogenesis, the regulation of gene expression, and RNA processing.

MATERIALS AND METHODS

Mice

Two strains of mice were utilized for these experiments. TgN1-2G transgenic mice were obtained from Dr. Jane E. Johnson, University of Texas Southwestern Medical School (Nakada et al., 2004). These mice express enhanced green fluorescent protein (GFP) under the control of a 7.5-kb fragment of the *Neurog1* promoter, causing fluorescent labeling of cells in the OSN lineage. C57BL/6 mice (Harlan Laboratories, Frederick, MD) were used for in situ hybridization mapping and for quantifying gene expression, in part to control for the possibility of strain-specific gene expression patterns. Handling and usage of animals was in accordance with NIH guidelines and approved Institutional Animal Care and Use Committee protocols.

Cell dissociation, fluorescence-activated cell sorting (FACS), and RNA isolation

Cells were dissociated from the olfactory epithelia of 32 postnatal day 5–8 (P5–P8) TgN1-2G mice (from five litters) and prepared for FACS as previously described (Sammeta et al., 2007; Yu et al., 2005). Dissections were biased toward the dorsal olfactory epithelium, although not exclusive to it, because the *Tgn1-2G* transgene tends to express more reliably in the dorsal epithelium. Cells were sorted into fetal bovine serum and then collected by centrifugation (500 rcf for 5 minutes at 4°C). In total, 6.8×10^6 GFP-negative cells, 1.3×10^6 weakly GFP-positive (GFP+), and 1×10^6 highly GFP-positive cells (GFP+++), were collected via FACS. Total RNA was isolated from these three fractions by using the TriReagent protocol (Molecular Research Center, Cincinnati, OH).

GeneChip microarray

For gene expression profiling of the three TgN1-2G sorted fractions, 1.25 µg of RNA per fraction was used for probe synthesis. Probes for the three fractions were hybridized to the M430v2.0 GeneChip (Affymetrix, Santa Clara, CA) by the University of Kentucky Microarray Core Facility. Initial analysis, including normalization of signal intensities and

detection call assignment, was done by using MAS5.0 software. Subsequent manipulation of these data was done in Excel (Microsoft, Redmond, WA). Probe sets with absent calls in all three fractions were ignored. To further ensure that background signals were eliminated, we also ignored probe sets that in all three fractions had signal intensities less than 15% of the mean signal for all probe sets on the array (Sammata et al., 2007).

Cell type enrichment ratio criteria sets

Enrichment ratios for all probe sets were computed from the signal intensity values of the GFP+++ sample/GFP+ sample, termed the High/Weak enrichment ratio. Immature OSN markers were enriched in the GFP+++ sample. Markers of mature OSNs were enriched in the GFP+ sample, but also present in the GFP- sample. Markers of other cell types were often enriched in the GFP- sample, but were typically also abundant in the GFP+ sample. Given these distributions, the High/Weak enrichment ratio proved to be the most useful measure generated, as it reflected enrichment in immature OSNs versus mature OSNs. Along with the Omp-GFP+/- ratio (Sammata et al., 2007), it was used to define criteria that distinguished expression in mature and immature OSNs. Initial criteria sets were generated by using the enrichment ratios of mRNAs known to be specific to cell types. These initial criteria were tested by in situ hybridization by using randomly selected mRNAs and then adjusted to form the final enrichment ratio criteria. Additional mRNAs were subsequently selected for in situ hybridization; along with previously tested mRNAs they constitute a pool of more than 350 mRNAs for which in situ hybridization labeled cells in or near the olfactory epithelium (Supplemental Fig. 1).

Immunohistochemistry and in situ hybridization

For immunohistochemistry, age P7 mice were anesthetized by hypothermia, and then transcardially perfused with 4% paraformaldehyde. The anterior-dorsal region of the head (snout) was dissected free and fixed in 4% paraformaldehyde for an additional 2 hours at 4°C. Snouts were then cryoprotected in a series of sucrose solutions at 4°C (1 hour in 10%, 1 hour in 20%, overnight in 30%) before being embedded in OCT compound (Sakura Finetek USA, Torrance, CA) and stored at -80°C. Coronal cryostat sections, 10-12 µm thick, were cut on a cryostat and mounted on Superfrost Plus slides (Fisher Scientific, Pittsburgh, PA). Slides were washed three times in 1X phosphate-buffered saline (PBS), pH 7.4 (10 minutes, room temperature) and permeabilized in 1% Triton X-100 in 1X PBS (30 minutes, room temperature). Slides were then blocked with 5% donkey serum, 0.4% Triton X-100 in 1X PBS (1 hour, room temperature). Primary antibodies were diluted in the blocking buffer and incubated overnight at 4°C. The following primary antibodies (Table 1) were used: rabbit anti-growth-associated protein 43 (GAP43; 1:200; Milipore, Bedford, MA; AB5220) and goat anti-olfactory marker protein (OMP; 1:1,000; Wako, Richmond, VA; 544-1001).

Slides were washed in 1X PBS (10 minutes, room temperature) and then incubated in secondary antibodies in the dark (1 hour, room temperature). Secondary antibodies, applied at 1:500 dilution in 1X PBS, were a Cy3-conjugated goat anti-rabbit IgG (111-165-144) and Cy3-conjugated donkey anti-goat (705-165-147), both from Jackson ImmunoResearch (West Grove, PA). Slides were washed three times in 1X PBS (10 minutes, room

temperature), counterstained with Hoechst 33258 at 0.0001 mg/ml in 1X PBS (5 minutes, room temperature, gentle agitation), and washed in 1X PBS (10 minutes, room temperature, gentle agitation). Finally, slides were air-dried, mounted with 100 μ l of Vectashield mounting medium (Vector, Burlingame, CA), and sealed with nail polish. The specificity of these antibodies has been previously documented (Baker et al., 1989; Inaki et al., 2004; Koo et al., 2004; McIntyre et al., 2010; Miller et al., 2010; Song et al., 2002).

In situ hybridization was performed as previously described (Sammeta et al., 2007; Shetty et al., 2005; Yu et al., 2005). C57BL/6 mice of two age groups were used: P4–P7 or P21–P24. The epithelium is dominated by immature OSNs at P4–P7 but by mature OSNs at P21–P24, so differentially expressed mRNAs show a visible shift in the number of OSNs labeled between these ages. In brief, mice anesthetized via intraperitoneal injection of ketamine hydrochloride (10 mg/ml) and xylazine (1 mg/ml) in 0.9% saline at age P21–P24 or by hypothermia at P4–P7 were perfused transcardially with 4% paraformaldehyde. Fixation in 4% paraformaldehyde of the dissected snout was then continued overnight. Digoxigenin-labeled riboprobes (prepared from cDNA fragments of 400–700 bp in length) were hybridized in 50% formamide in 10 mM Tris-HCl (pH 8.0), 10% dextran sulfate, 1X Denhardt's solution, 600 mM NaCl, 0.25% sodium dodecyl sulfate, 1 mM ethylenediaminetetraacetic acid, 200 μ g/ml yeast tRNA, and 200 ng of riboprobe, at 65°C in a humidified chamber. An alkaline phosphatase-conjugated antibody to digoxigenin and nitroblue tetrazolium chloride/5-bromo-4-chloro-3'-indolyphosphate p-toluidine (both from Roche Diagnostics, Indianapolis, IN) were used for detection of hybridization. Control sense-strand probes were generated for each mRNA tested and were all devoid of staining. The mRNAs and probe regions that had been tested by the completion of this project are listed in Supplemental Table 1. In situ hybridizations that gave detectable signals were obtained from 352 of these mRNAs (Supplemental Fig. 1).

For immunohistochemistry, digital images were collected by using a Leica SP1 inverted confocal microscope at the University of Kentucky Imaging Facility. For in situ hybridization, digital images were collected by using a Spot 2e camera (Diagnostics Instruments, Sterling Heights, MI) mounted on either a Nikon Diaphot 300 inverted microscope or a Nikon Eclipse Ti-U inverted microscope. Adjustments to size, brightness, and contrast were made to original images in Adobe (San Jose, CA) Photoshop before being collated and labeled in Deneba Canvas.

Quantification of ISH signal position

Individual in situ hybridization images for some mRNAs were imported into NIH ImageJ. An area (~200 μ m in length) that began at the basal lamina and extended to the apical surface was then selected. This portion of the image was analyzed by using the plot profile tool. The resulting output of signal strength versus epithelial depth was imported to Excel for scaling of epithelia depth and plotting of the mean and range of the peak signal.

Statistical analyses

Binomial tests were applied to the in situ hybridization data to test the hypothesis that enrichment ratio criteria predict cell type specificity better than random chance. Confidence

intervals were calculated by the Adjusted Wald method (Agresti and Coull, 1998). To go beyond simple threshold criteria, we used the fine gradations in the enrichment ratio patterns to calculate likelihood of expression in cell types. Mixture discriminant analysis (MDA) (Hastie and Tibshirani, 1996), as implemented in R (<http://www.R-project.org>), was used to estimate probabilities of expression of every gene in designated cell type categories. MDA uses Bayes' rule to calculate the probability that an object belongs to a previously established group. In this regard, it is similar to Fisher's linear discriminant analysis (Fisher, 1936). The difference is that linear discriminant analysis assumes that the data follow a normal distribution for each category, whereas MDA models each category by using a semiparametric mixture of normal distributions. By making fewer assumptions about the underlying distribution of the data, this method can achieve more flexible estimates of the values, enrichment ratios in our case, that represent the objects (genes) assigned to categories (cell types), and thereby yield more accurate probability estimates.

In situ hybridization data for 335 mRNAs that labeled distinguishable cell types and for which we had both enrichment ratio values were used to fit the MDA model. Our calculations for the probability of the cell type categories were based on the assumption that, a priori, all four categories were equally likely. Ten-fold cross-validation was used to determine the degrees of freedom for the mixture distributions.

Functional bioinformatics analyses of genes specific to immature and mature OSNs, as determined by the enrichment ratio criteria, were performed by using NIH DAVID2008 with $\alpha = 0.01$ as previously described (Huang da et al., 2009a,b; Sammeta et al., 2007). Functional bioinformatics analyses of gene expression patterns as determined by MDA probability was done by two-tailed t-tests of the hypothesis that Gene Ontology terms associated with the genes would be independent of the Genomics of olfactory sensory neurons cell type probabilities. False discovery rate correction for multiple testing was used, reported herein as q (Storey and Tibshirani, 2003) Significance was set at $q < 0.05$.

RESULTS

Identifying immature and mature olfactory sensory neurons

In order to characterize gene expression in immature olfactory sensory neurons, we first needed to generate samples enriched in these neurons. To accomplish this goal we utilized TgN1-2G transgenic mice in which enhanced GFP is under the control of a 7.5-kb fragment of the Neurog1 (Ngn1; neurogenin-1) promoter (Nakada et al., 2004). Expression of the native Neurog1 gene in the olfactory epithelium is restricted to a transient population of immediate neuronal precursor basal cells (Calof and Chikaraishi, 1989; DeHamer et al., 1994). The transience of this last basal cell stage and the perdurance of GFP (LeSauter et al., 2003) gave reason to test for GFP labeling of immature OSNs in the olfactory epithelia of TgN1-2G mice. Even if the TgN1-2G transgene is faithful to the expression pattern of the endogenous gene, rapid transition from immediate neuronal precursor basal cells into immature OSNs could result in fluorescent labeling primarily of immature OSNs. Consistent with this expectation, we found that the GFP fluorescence pattern consisted of only a few basal cells but many cells with fluorescent dendrites and axons—cells located at the depth of the immature OSN layer of the olfactory epithelium (Fig. 1C). This GFP fluorescence

overlapped well with immunoreactivity for the immature OSN-specific marker Gap43, but not for the mature OSN-specific marker, Omp, confirming that the highly fluorescent cells were nearly all immature OSNs (Fig. 1C).

The olfactory epithelium is pseudostratified, with cell body layers found at characteristic positions (Fig. 1A). Although the boundaries between these cell body layers have some irregularity, depth in the olfactory epithelium is a consistent and reliable indicator of cell type as long as two conditions are met: 1) expression must be detected in a majority of cell bodies in a cell layer to rule out the possibility that infiltrating cells or rare cell types (e.g., macrophages or olfactory microvillar cells) produced the labeling; and 2) conditions that might disrupt the cell body layers of the epithelium must be avoided (e.g., certain lesions or genetic mutations). To demonstrate the ability of epithelial depth to identify cell types, we calculated the mean position of the peak in situ hybridization signal for a panel of mRNAs that specifically label globose basal cells, nascent OSNs, immature OSNs, mature OSNs, and sustentacular cells (Supplemental Table 2) and plotted these data for comparison (Fig. 1B). As expected, this measure produced distinct distributions for globose basal cells, immature OSNs, mature OSNs, and sustentacular cells; nascent OSNs were positioned at the basal end of the distribution of immature OSNs.

Armed with confirmation that strong GFP fluorescence in TgN1-2G mice identifies immature OSNs, we isolated three FACS fractions from age P5–P8 olfactory epithelium (Fig. 1D,E). At this age, immature OSNs are near their peak frequency relative to other cell types within the epithelium. The fluorescence patterns in tissue sections predicted that the highly GFP-positive fraction (GFP⁺⁺⁺) would be enriched in immature OSNs, the weakly GFP-positive fraction (GFP⁺) enriched in mature OSNs, and the GFP-negative fraction (GFP[–]) should contain all cell types but be depleted of immature OSNs. Messenger RNA abundance in these fractions was quantified by using the Affymetrix M430v2.0 GeneChip (Gene Expression Omnibus accession number GSE33368). Probe sets with signals above background identified 10,731 unique genes in the GFP⁺⁺⁺ sample, 11,744 genes in the GFP⁺ sample and 12,088 genes in the GFP[–] sample.

Atlas of gene expression in olfactory sensory neurons

In order to identify mRNAs enriched in specific cell types without having to physically map the expression of every gene, for each probe set we calculated enrichment ratios of the signal intensities among the three FACS samples, and then compared these enrichment ratios against physical map data generated by in situ hybridization. This identified ranges of enrichment ratios that define expression in specific cell types. The most useful enrichment ratio in the TgN1-2G data was the GFP⁺⁺⁺ signal divided by the GFP⁺ signal (hereafter termed the High/Weak ratio). We predicted that the High/Weak ratio directly correlates with the degree of specific expression in immature OSNs. To further increase discriminatory power, we also used the previously published Omp-GFP^{+/-} enrichment ratio data generated from FACS isolation of mature OSNs (Sammata et al., 2007) to produce enrichment ratio criteria sets that predict expression in three distinct OSN cell type categories (Table 2). To illustrate the discriminatory ability of this approach, consider the distinct patterns formed by the High/Weak and Omp-GFP^{+/-} enrichment ratios, respectively, of the four best known

markers of cell types in the OSN cell lineage: *Omp* at 0.5 and 44.4 for mature OSNs, *Gap43* at 1.9 and 0.6 for immature OSNs, *Ascl1* (*Mash1*) at 0.2 and 0.1 for early neural progenitors, and *Neurog1* at 1.1 and 0.1 for late neural progenitors.

Given the ability of our enrichment ratio criteria to distinguish known gene expression patterns, we then tested the enrichment ratio criteria set for immature OSNs by *in situ* hybridization. Of the eight predicted immature OSN mRNAs that were randomly selected, all were specifically detected in the immature OSN layer of the olfactory epithelium ($n = 8$, $P = 0.0039$). Similarly, *in situ* hybridization was used to confirm the criteria used to identify mature OSN-specific transcripts. In total, 29 mature OSN predicted genes were tested and all displayed mature OSN expression patterns. Confirmation of the reliability of the enrichment ratio criteria sets then allowed their use to predict genes expressed by mature, but not immature, OSNs and vice versa. In all, 847 immature OSN-specific genes, 691 mature OSN-specific genes, and 6,761 genes shared between the two OSN populations were identified (Supplemental Tables 3–5). Because these predictions included mRNAs present in our *in situ* hybridization database, we were able to assess the accuracy of the enrichment ratio criteria (Table 3). The numbers in Table 3 could be overly optimistic because the data used to establish the enrichment ratio criteria were included. We see no evidence of this from the data, however; among the smaller set of mRNAs that were not involved in the establishment of the criteria and thus serve only for validation, there were zero misclassifications. The accuracy of prediction was limited largely by the close relationship between immature and mature OSNs. All of the failed specificity predictions involved genes expressed in both immature and mature OSNs. This predictable result is consistent with the view that the sharp distinction between mature and immature OSNs defined by *Gap43* and *Omp* expression is somewhat artificial, arguing for analysis methods that allow graded representations of the transition from immaturity to maturity (see mixture discriminant analysis, described below).

Immature-OSN specific mRNAs: transcription, RNA processing, differentiation, and axonogenesis

Of the ~10,000 genes expressed by OSNs, our enrichment ratio criteria identified 847 whose expression is likely to be restricted to immature OSNs. Overall, we generated new *in situ* hybridization data for 31 of these mRNAs; all were expressed exclusively or primarily by immature OSNs. Figure 2 shows examples of these data.

To test the prediction that these 847 mRNAs should encode proteins involved in biological processes important for neural differentiation, we identified functional annotations statistically overrepresented among the Gene Ontology categories associated with these genes. We detected 31 overrepresented categories that can be organized into five groups by their functional similarities (Table 4).

Neuronal differentiation requires major changes in gene expression programs, so, not surprisingly, many mRNAs encoding proteins that perform these functions were specific to immature OSNs. For example, we identified numerous chromatin remodeling genes and transcription factor genes among the immature OSN-specific genes (Supplemental Tables 6, 7). Immature OSNs specifically express two SWI/SNF-related, matrix-associated, actin-

dependent regulators of chromatin (Smarcd2 and Smarce1), two actin-like 6 genes (Actl6a and Actl6b) that antagonize chromatin-mediated transcriptional repression, and two SWI/SNF cofactors (Rb and Rcor1). Some members of the classical (class I and II) histone deacetylases (Hdac2, -5, and -10), four histone methyltransferases (Setdb2, Suv420h2, Smchd1, and Whsc1), and one component of Polycomb group repressor complex 1, Cbx2, were also specific to immature OSNs. All of these chromatin modifiers are parts of larger protein complexes whose other components were found to be expressed more broadly in the olfactory epithelium, typically in both mature and immature OSNs, but also in basal cells in some instances (Sammata et al., 2007; Shetty et al., 2005).

Chromatin remodeling in immature OSNs should be directed to specific sites in the genome by transcription factors and control transcription factor activity by placing chromatin marks that regulate transcription factor access and activity. Of the 847 immature OSN-specific genes, 91 were transcription factors (Supplemental Table 7). A few, like Dlx5, Emx2, Fezf1, and Klf7, are already known to be important for OSN differentiation because without them, OSN axons fail to innervate the olfactory bulb (Bishop et al., 2003; Laub et al., 2005; Levi et al., 2003; Long et al., 2003; McIntyre et al., 2008; Shimizu and Hibi, 2009; Watanabe et al., 2009). Whether the other transcription factors we identified are similarly critical for immature OSNs is as yet unknown.

A major phenotypic change during neuronal differentiation is the elaboration of neurites. Given the specificity of OSN connections to the olfactory bulb, the guidance of OSN axons is particularly critical. The components of axonogenesis that were specific to immature OSNs were related to the control of cytoskeletal dynamics (Ablim1, Evl, Dclk1, Kif5c) or axon guidance and axon fasciculation (Dpysl5, Nrp1, Ephb2, Gap43, Nrcam, Tubb3) (Supplemental Table 8). These data agree with our previous finding that the downstream signaling elements that connect axon guidance cue receptors to cytoskeletal dynamics are often expressed in immature OSNs but are lacking in mature OSNs (McIntyre et al., 2010).

Extending neurites, especially axons that course long distances, also requires a significant investment in the production of the lipid components of the plasma membrane. This agrees with the overrepresentation of sterol biosynthesis in immature OSNs, largely due to expression of genes encoding enzymes involved in the synthesis of cholesterol (Supplemental Table 9). Our data predict that several of these genes are expressed by immature, but not mature, OSNs (Hmgcr, Pmvk, Mvd, Fdps, Nsdhl, Idi1, Sc4mol, Cyp51, and Dhcr7). This was consistent with our finding that Srebf2, a sterol-sensitive basic helix–loop–helix transcription factor that is critical for the expression of cholesterol synthesis genes, was specific to immature OSNs (Shimano, 2002; Tarr and Edwards, 2008). Immature OSNs also express other genes that have roles in the regulation of lipid membrane components. This included two lipid receptors, Osbpl8 and Ldlr, plus the associated Ldlrap1.

Perhaps more surprising than these other functions was the overrepresentation of RNA processing in immature OSNs. These genes (Supplemental Table 10) encoded ribonucleases, proteins involved in pre-mRNA processing (Hnrnpa1, Hnrnpk, Lsm4, Lsm5, Phrf1, Prpf19, and C1d), one protein involved in histone RNA processing (Lsm10), RNA transport proteins

(Ncbp2, Pnn, Sfrs3, and Sfrs7), and mRNA splicing or transport factors (Rbm9, Bat1a, Magohb, Nhp211, Snrnp35, Snrnp40, Snrpb2, and Thoc1), all processes that have never been studied in OSNs. They also included genes involved in the regulation of tRNAs (Qrt1, Qrt1, Elac2, Trmt61a, Trmt1, and Trub2). These data argue that the differentiation of OSNs requires a set of RNA processing functions that are not needed in mature OSNs.

Nascent OSN-specific mRNAs

We recently demonstrated that cells in transition from globose basal cells to immature OSNs could be distinguished by their expression of *Cxcr4* and *Dbn1*, two axon initiation genes (McIntyre et al., 2010). *Cxcr4* expression overlaps poorly with expression of globose basal cell markers and moderately with *Gap43*, the canonical marker of immature OSNs. About half of these *Cxcr4*⁺ cells, which we term nascent OSNs, are not identified by canonical markers of basal cells or immature OSNs. Nascent OSNs lie at the interface of the basal cell layer and the immature OSN layer (Fig. 1B). Given that TgN1-2G transgene expression should begin in immediate neuronal precursor globose basal cells that give rise directly to nascent OSNs, we reasoned that those mRNAs with the largest High/Weak ratios might be specific to nascent OSNs. However, this prediction was not supported by statistical testing of in situ hybridization data for mRNAs with the largest High/Weak ratios ($n = 6$, $P = 0.2344$), suggesting that the GFP⁺⁺⁺ sample was broadly representative of immature OSNs and not biased toward nascent OSNs. Nevertheless, we identified two mRNAs with expression patterns restricted to the nascent OSN sublayer (Fig. 3). These were *Col9a2* (collagen, type IX, alpha 2) and *Gng2* (guanine nucleotide binding protein (G protein), gamma 2).

Mature OSN-specific mRNAs: ion channels, synaptic transmission, and cilia

The High/Weak enrichment ratio proved invaluable in making distinctions between mRNAs specific to mature OSNs and those shared with immature OSNs, allowing us to identify 691 genes expressed specifically by mature OSNs. Representative in situ hybridization data for nine of these genes are shown in Figure 4. Functional bioinformatics analysis of these 691 genes identified 20 overrepresented Gene Ontology categories. These 20 categories represented three related functional groups (Table 5). The most prominent grouping was ion transport, which was primarily composed of calcium, potassium, and sodium channel subunits, but also included the *Trpc1* channel and the Na⁽⁺⁾/K⁽⁺⁾ ATPase regulator, *Fxyd2* (Supplemental Table 11). The cell–cell signaling group of categories stemmed from genes (Supplemental Table 12) that encode proteins involved in synaptic vesicle function (*Dlgap3*, *Myo5a*, *Myo6*, *Tmod2*, and *Slc17a6*), neurotransmitter release (*Snap25*, *Sv2b*, *Sv2c*, *Syn2*, *Syp*, *Syt1*, *Syt4*, and *Unc13c*), or presynaptic inhibition (*Drd2*). Lastly, the cellular projections group was composed mostly of genes (Supplemental Table 13) that encode cilia/flagella proteins, consistent with our previous analysis of cilia-related genes in mature OSNs (McClintock et al., 2008).

Genes expressed in both immature and mature OSNs

Most genes expressed by OSNs are shared by both immature and mature OSNs, and although some show noticeably different intensities in the mature and immature layers when

detected by in situ hybridization, most appear fairly equivalent. We were able to identify 6,761 genes that we predict with confidence are expressed in both immature and mature OSNs. In situ hybridizations for five representative shared OSN mRNAs (*Becn1*, *Cpe*, *Palm*, *Panx1*, and *Unc45a*) are shown in Figure 5, along with *Ncam1*, which is often used to identify all OSNs. Functional bioinformatics suggests that these shared genes tend to be associated with broad functional categories such as metabolism and cell organization.

Mixture discriminant analysis classification

As mentioned above, the analysis by enrichment ratio criteria was particularly useful in identifying genes specifically expressed in mature or immature OSNs, but such analysis had difficulty representing expression patterns that did not mimic the canonical *Omp–Gap43* distinction of OSN maturity, such as mRNAs whose abundance changed more gradually or changed sharply at a different point during OSN differentiation. To model gene expression in a way that more realistically represents phenotypic transitions during OSN differentiation, we employed a classification model, mixture discriminant analysis (MDA), to estimate probabilities of expression in each cell type category for all mRNAs detected. These estimates include probabilities for specific expression in single cell type categories ($P_{(sp)}$) and inclusive probabilities of being expressed in a given cell type category ($P_{(in)}$). The categories, including a category for mRNAs expressed in both immature and mature OSNs (shared), and examples of the variations in probability distributions obtained are listed in Table 6. This table uses as examples only genes whose functions have never been investigated to illustrate the ability of this approach to provide information about genes that have long been ignored.

When we simply assigned each gene to the cell type category in which its $P_{(sp)}$ was the highest, the MDA $P_{(sp)}$ predictions were able to reproduce nearly all of the cell type distributions determined by in situ hybridization. To illustrate this correspondence in a manner that reflects distinctions made possible by the enrichment ratio data, we plotted the in situ hybridization and MDA prediction data on enrichment ratio axes (Fig. 6A,B). Quantification of this comparison revealed that $P_{(sp)}$ accurately predicted the expression patterns for 289 of the 335 (86%) genes for which we had in situ hybridization data. Just as with enrichment ratio criteria sets, the MDA $P_{(sp)}$ classification errors were mostly at the boundaries between cell types (Fig. 6C). Classification errors were typically genes whose expression graded across a transition between cell types in the OSN lineage such that they had $P_{(sp)}$ probabilities near 0.5 in two cell type categories, such as *Dpysl2* and *Gprc5c* (Fig. 6D,E). Misclassifications were distributed among cell type categories (Table 7), although the larger number of shared mRNA misclassifications of mature and immature OSN mRNAs suggested that the MDA probabilities were slightly conservative in predicting specificity in mature and immature OSNs. Only 12 of the misclassifications were mRNAs whose in situ hybridization patterns diverged radically from the MDA predictions, meaning that they had MDA probabilities that were distant (0.8 or higher) from the 0.5 boundary used in this analysis. An example is *Ch11*, which was enriched in immature OSNs but given a high probability of expression in the other cell type category by the MDA (Fig. 6F). Whether these mRNAs represent poor performance of the FACS, the microarray probe sets, the in situ hybridizations, or the MDA is as yet unknown. For the vast majority of genes the

MDA probability distribution patterns accurately represent differences between cell type categories in the amounts of mRNAs contained in these cells.

Identifying genes with specific expression in a cell type is useful, but in terms of understanding the functional capabilities of cell types, identifying all genes expressed in each cell type is just as important. The MDA method was therefore used to calculate a second set of probabilities ($P_{(in)}$) designed to gauge the mere expression in mature OSNs, immature OSNs, and the broad category of all other cells. This analysis identified 9,300 genes with a $P_{(in)} = 0.5$ in either or both mature and immature OSNs, plus 1,664 genes with a $P_{(in)} = 0.5$ in the other population (Fig. 7A). The majority of mRNAs we detected arose from genes that had high probabilities of expression in mature and/or immature OSNs (Fig. 7B). Less abundant were genes with intermediate probabilities of expression in mature and/or immature OSNs ($P_{(in)}$ values between 0.5 and 0.75, for example). When plotted on enrichment ratio axes, these genes concentrated at the boundaries of the enrichment ratio criteria for prediction of expression in immature and mature OSNs and were therefore difficult to distinguish (Fig. 7C). However, in plots using $P_{(in)}$ axes, these genes were dispersed, depicting how gene expression sometimes grades across the conventional distinctions between cell types in the olfactory epithelium (Fig. 7B,D). For example, the majority of these genes had higher probabilities of expression in immature OSNs than in mature OSNs (lower right versus upper left quadrants in Fig. 7B). This bias was expected given the direct line-age relationship between basal progenitor cells and immature OSNs, which argues that a substantial number of mRNAs would be shared by basal cells and immature OSNs, thereby generating a population of genes that had intermediate $P_{(in)}$ immature OSN values but low $P_{(in)}$ mature OSN values. For the same reason, a biased distribution should also be apparent among genes expressed primarily ($P_{(in)} > 0.5$) in the other cell type category, which should contain genes expressed by the basal cell progenitors of OSNs. Indeed, these genes (blue circles in Fig. 7B,D) tended to have higher probabilities of expression in immature OSNs than in mature OSNs.

Biological processes active in OSNs

The functional bioinformatics of genes specific to cell types (Tables 4, 5) were remarkably informative about the biology of immature and mature OSNs, but necessarily incomplete due to the limited set of genes. Better analyses were made possible by the $P_{(in)}$ gene lists for mature and immature OSNs, and the $P_{(in)}$ list of genes shared by all OSNs. As expected, this approach gave a more complete picture of the properties that emerge from the functional relationships among the gene products expressed by OSNs. Restricting the analysis to Gene Ontology annotations, we detected 25, 90, and 19 biological process terms overrepresented in mature OSNs, immature OSNs, and all OSNs (shared genes), respectively. For mature OSNs, we identified not only the mature OSN-specific transport of ions necessary to support electrical excitability and synaptic function but also even larger numbers of genes that support the transport of proteins within these morphologically complex cells (Table 8). The high metabolic activity of neurons was represented by the electron transport chain and metabolic/biosynthetic process categories. The unfolded protein response, also known as the ER stress response, that is activated by chemical stress in OSNs was apparent in the ER overload response category (Sammata et al., 2010b). The specific function of OSNs,

detecting odors, was now apparent in the perception of smell category. A final step in OSN differentiation, the elaboration of cilia, was represented in the striatum development and photoreceptor cell maintenance categories because these biological processes depend critically on cilia transport proteins and cilia function (Adams et al., 2007; Davis et al., 2007).

In immature OSNs we detected overrepresentation of even broader groups of biological processes involved in gene expression control, RNA processing and mRNA splicing, the development of neurons, the extension of neurites, and the synthesis of lipids than was apparent in the analysis of genes specifically expressed in immature OSNs (Table 9; Supplemental Fig. 2; cf. Table 4). Intriguing new categories also arose. Immature OSNs showed the same ER overload response as mature OSNs, consistent with *in situ* hybridization data for unfolded protein response transcripts (Sammeta and McClintock, 2010). Immature OSNs had evidence of a high capacity for protein translation and transport, consistent with expression of large amounts of numerous proteins that populate the neurites being extended by these cells. Finally, immature OSNs showed evidence of apoptosis via death domain receptors, consistent with evidence that apoptosis of immature OSNs is a continuous process that might serve to balance an overproduction of OSNs (Holcomb et al., 1995).

Analysis of the $P_{(in)}$ list of genes shared by all OSNs was used to identify biological processes shared by mature and immature OSNs (Table 10; Supplemental Fig. 2). As expected from the results for mature and immature OSNs, this analysis identified protein transport, ER overload response, and electron transport. The data also predict that OSNs have a significant capacity for protein catabolism.

Analysis of cellular component and molecular function Gene Ontology terms was also done for each of the three OSN cell type categories. As expected, these terms were mostly functions or compartments involved in the overrepresented biological processes described above (Supplemental Figs. 3, 4). However, a few terms appeared to be distinct from these processes or are worthy of mention for other reasons. The number of genes that encode zinc binding proteins expressed in OSNs was very large, at 1,151 genes. In addition, immature OSNs show overrepresentation of genes related to the Barr body and the chromocenter, nuclear complexes involved in heterochromatin formation, which may be of significance for the development of strong repression of many genes in a fully differentiated neuron (Chadwick and Willard, 2003; Magklara et al., 2011).

DISCUSSION

Atlases of gene expression in mammalian tissues have typically been massive efforts that focused on genes. They have not produced comprehensive, statistically validated lists of genes expressed in individual cell types. Taking an approach centered on cell types rather than genes, we have invented a strategy that appears capable of reliably predicting the expression of nearly all genes in each cell type in a tissue. This strategy estimates error and uses it to generate probabilities of expression in the targeted cell type rather than assume that cell type purification and RNA detection methods are perfect. We determined probabilities

of expression in mature OSNs and immature OSNs for thousands of genes and calculated that each OSN expresses 7,000–10,000 genes. These probabilities showed graded patterns that reflected the degree of relatedness among mature OSNs, immature OSNs, and the residual sample of all other cells in the olfactory epithelium (which includes the OSN progenitor cells that give rise to the immature OSNs). In all, 1,538 genes that distinguish mature and immature OSNs from each other were identified. The functional relationships between these genes identified biological processes specific to mature and immature OSNs, whereas bioinformatic analysis of the $P_{(in)}$ lists of genes expressed in these cells identified even broader sets of biological processes active in OSNs. By simultaneously identifying these processes and the genes that supply their component parts, these data facilitate moving directly to mechanistic studies. This only provides information for about half of the genes expressed in OSNs, however, because functional bioinformatics analyses are restricted to annotated genes and therefore provide little information on the many mammalian genes that have never been investigated beyond what can be gleaned from their sequences. However, by identifying the understudied genes expressed in mature and immature OSNs, we have generated de facto hypotheses about their functions, predicting that many of them contribute to the biological processes we found to be overrepresented in OSNs. To facilitate further study of mechanisms in OSNs, a file detailing the enrichment of each mRNA in mature and immature OSNs, and the probabilities of expression of each gene in these cells, is available (Supplemental database).

Each OSN expresses 7,000–10,000 genes

We consistently detect the mRNAs from 13,000–15,000 distinct genes in olfactory epithelium samples, and have estimated that ~10,000 of them are expressed in the OSNs (McIntyre et al., 2010; Sammeta et al., 2007). By enriching for both mature and immature OSNs, we have distinguished which genes were expressed only by mature OSNs, which were expressed only by immature OSNs, which were shared by both mature and immature OSNs, and which were unlikely to be expressed in OSNs (Fig. 8). We identified 8,299 genes whose enrichment ratios predicted that they were expressed by OSNs. This quantity includes only those mRNAs detected by both this study and our previous work on mature OSNs (Sammeta et al., 2007). Another 1,586 mRNAs were detected in only one of our studies but in each case had an enrichment ratio consistent with expression in OSNs. Most of these genes are expressed in OSNs, we believe.

MDA probabilities may better estimate the total number of genes expressed by OSNs than enrichment ratio criteria. $P_{(in)}$ values > 0.5 identify 9,300 genes that we predict are expressed in OSNs, with 7,045 of those in mature OSNs and 8,573 in immature OSNs. The MDA probabilities formed patterns that better represented genes whose expression crosses or otherwise differs from the conventional boundaries between immature and mature OSNs than enrichment ratio criteria sets. MDA probability patterns emphasized the pivotal position of immature OSNs in the OSN cell lineage by identifying mRNAs that peak in abundance in immature OSNs but are also expressed in mature OSNs or perhaps in basal cell progenitors of OSNs (Fig. 8B,C). Just as with enrichment ratio data, however, the MDA probabilities underestimate the number of genes expressed in OSNs because they also did not include the 1,586 genes that had only one enrichment ratio value, but nevertheless a

value consistent with OSN expression. Given that MDA probabilities > 0.5 had an error rate of just 4% for prediction of genes expressed in OSNs versus genes not expressed in OSNs (a 14% error rate applies to discriminating genes expressed in mature versus immature OSNs), our findings continue to support the interpretation that each mature and immature OSN expresses at least 7,000 genes and perhaps as many as 10,000 genes.

Another measure of whether the predictions made by these probability distributions are accurate is to assess whether they agree with mRNA abundance changes 5–7 days after removal of the olfactory bulbs when mature OSNs are nearly absent and immature OSN production is increased (Shetty et al., 2005). Of those mRNAs detected in both this study and after bullectomy, we found that 93% of the mRNAs that decreased had $P_{(in)} > 0.5$ in mature OSNs. Conversely, 82% of the mRNAs that increased after bullectomy had $P_{(in)} > 0.5$ in immature OSNs. This measure therefore also supports the ability of our approach to accurately distinguish mRNAs expressed differentially between mature and immature OSNs.

Enrichment ratio criteria identified at least 2,677 genes, and perhaps as many as 4,402 genes, that we predict are not expressed in OSNs. This would seem to imply that the other cell types in the olfactory epithelium, such as basal cells and sustentacular cells, express many fewer genes than OSNs. Such an interpretation would be premature, however. A significant number of genes expressed by OSNs are probably also expressed at lower levels in these neighboring cells. For example, high energy consumption in OSNs may require high expression of enzymes related to energy production, as evidenced by the overrepresentation of electron transport chain enzyme mRNAs in OSNs, yet these genes are expressed in all cells. Similarly, unfolded protein response genes, which are abundant mRNAs in OSNs, are almost certainly also expressed at much lower levels in neighboring cells (Sammeta and McClintock, 2010). Only by separately analyzing these other cell types will we be able to settle this question.

Gene expression in the olfactory epithelium shows several types of variation that do not conform to simple expectations. A significant number of mRNAs show patterns that grade across the conventional distinctions between successive cell types in the OSN cell lineage. Genes that have different transcriptional onsets or offsets than canonical cell type markers are one source of variation. In the sense that nascent OSNs could be considered a subset of immature OSNs, their mRNAs are examples of this. Another consistent pattern of spatial variation was observed with *Umod11* (previously known as *n8*), an mRNA specific to mature OSNs. It is expressed in a graded pattern of high apical to low basal in the mature OSN layer and it is much less abundant in a mutant mouse strain whose mature OSNs have reduced survival (Sammeta et al., 2010a; Yu et al., 2005). We hypothesize that *Umod11* is not controlled simply by apical–basal position but by some other factor, perhaps related to OSN age. More common than these consistent patterns of variation, however, were the apparently random differences in signal intensity we often observed between neighboring cells. Even for mRNAs whose expression was clearly limited to either mature or immature OSNs such spatial heterogeneity was noticeable (e.g., Figs. 2, 4). One type of variation that can appear in a spatially random pattern across OSNs is activity dependence (Bennett et al., 2010; Serizawa et al., 2006). However, we believe that this is also rare because our

microarray studies of naris occlusion to block odor stimulation detected less than 20 differentially abundant mRNAs (A.M. Fischl and T.S. McClintock, unpublished data).

Perhaps a more compelling explanation comes from the recent realization that transcription of a gene occurs in bursts that are not coordinated between cells (Suter et al., 2011). When viewed as a snapshot in time, the resulting temporal variation in mRNA abundance should be revealed as differences in signal intensity across a population of cells. Variation in the abundance of an mRNA across a population of cells may therefore be the norm.

Molecular evidence of mature OSN functions

Many genes expressed specifically by mature OSNs fell into three functional groups. First, many genes encoding proteins involved in the elaboration and maintenance of cilia were not expressed until after OSNs become mature, consistent with anatomical evidence (Cuschieri and Bannister, 1975). Cilia are the sites of odor detection and signal transduction, so mutations in cilia genes often cause loss of OSN cilia, reduction in size of cilia, and defects in transport of cilia proteins that lead to anosmia or hyposmia (Jenkins et al., 2009; Tadenev et al., 2011). Second, maturation of OSNs is accompanied by expression of several genes involved in presynaptic functions, arguing that the maturation of synapses in the olfactory bulb is another characteristic feature of the final steps in OSN maturation (Kim and Greer, 2000). These genes encoded synaptic vesicle proteins and neurotransmitter release proteins, consistent with a previous analysis of a subset of these genes (Marcucci et al., 2009). Third, the communication of odor detection in cilia to the distant sites of synaptic transmission in the olfactory bulb requires the ion channel and transport functions that support electrical activity. Many of these ion channel and ion transport genes are also not expressed until OSNs mature. When seeking genetic causes of peripheral anosmia or hyposmia, the critical physiological functions performed by the products of these mature OSN-specific genes make them likely candidates.

Mature OSNs also share with immature OSNs high expression of genes that support one general characteristic of neurons: high energy consumption, consistent with the strong expression of genes encoding proteins involved in the production of ATP by OSNs. More unusual is the strong expression of ER stress genes in all OSNs, a pattern of gene expression related to the exposure of OSNs to damaging agents, especially chemicals (Sammata and McClintock, 2010). These ER stress genes are not the only evidence that OSNs experience stress. The overrepresentation of DNA repair (immature OSNs) and ubiquitin-dependent catabolic terms (all OSNs), which are associated with cells under oxidative stress (Barzilai et al., 2008; Gasch et al., 2000), might also be due in part to stress experienced by OSNs. Both developmental stages also express >1,100 genes encoding proteins that bind zinc or are predicted to bind zinc. Although the unusual sensitivity of OSNs to zinc might lie within the functions of these gene products (Lim et al., 2009; Viswaprakash et al., 2010) it is very possible that elevated zinc in the olfactory epithelium affects proteins that do not normally bind zinc.

Molecular evidence of immature OSN functions

The analysis of immature OSN gene expression emphasizes the importance of the control of gene expression as these cells transition from neurally fated basal cells into mature neurons. We identified 31 epigenetic modification genes and 128 direct transcriptional regulators that were specific to immature OSNs. Expanding this to all genes expressed in immature OSNs, we detected 193 chromatin remodeling/modification genes and 766 genes encoding other types of known or predicted transcriptional regulators. The epigenetic genes include a variety of chromatin modification-related functions including histone methyltransferases, histone demethylases, histone acetylases, histone variants, polycomb genes, and SWI/ SNF complex genes. Several chromatin remodeling mRNAs were previously tested by *in situ* hybridization and they typically showed high expression in immature OSNs, grading off into lower expression in mature OSNs (Sammata et al., 2007). The transcriptional regulators expressed in both mature and immature OSNs were primarily transcription factors, a fertile field for studying the mechanisms of gene expression regulation that drive neural differentiation.

The most functionally significant gene expression change that happens in immature OSNs is the choice of which single odorant receptor gene allele to express. Evidence implicating the transcription factors *Emx2* and *Lhx2* cannot explain any mechanism for the precise expression of only one allele of one odorant receptor gene in each OSN (Hirota and Mombaerts, 2004; McIntyre et al., 2008). One hypothesis does specify part of a potential mechanism, proposing the existence of universal epigenetic repression of odorant receptors followed by removal of this repression at a single locus (Magklara et al., 2011; McClintock, 2010). Repressive methylation marks on histones H3 and H4 (H3K9me3 and H4K20me3) have recently been described at odorant receptor genes (Magklara et al., 2011) in both mature and immature OSNs. The identity of the methyltransferases responsible for adding these repressive marks, the precise timing of their addition, and whether other repressive marks are present on the OR clusters remain unclear. However, some methyltransferases found within our OSN gene lists are good candidates for the agents that catalyze addition of these methyl groups. In fact, the immature OSN-specific *Suv420h2* and three methyltransferases (*Suv420h1*, *Ash1l*, and *Nsd1*) shared by all OSNs are known to generate H4K20me3 marks. Generation of H3K9me3 *in vitro* has been demonstrated to occur in a stepwise manner, whereby monomethylation of H3K9 is catalyzed by the HP1a-CAF1-SetDB1 complex, priming this locus for trimethylation by *Suv39h1* (Loyola et al., 2009). Our data predict that all four of these genes are expressed by both immature and mature OSNs. Even though these data begin to suggest that H3K9me3 and H4K20me3 are added to odorant receptor genes in immature OSNs, we cannot rule out expression of these methyltransferases in basal cells; only work with purified basal cells will be sufficient to determine when in the OSN cell lineage these marks are added.

While immature OSNs are making decisions about which genes to express, they are also greatly increasing their capacity to make the proteins encoding these genes by enhancing expression of genes encoding proteins that transport and splice mRNAs, regulate tRNAs, and translate the mature mRNAs into proteins. Some of this increased capacity may be general, and much of it is shared with mature OSNs, but some of it may instead have more

specific functions. For example, Rbm9 (Rbfox2) is a splicing enhancer of neuronally regulated exons, and Lsm10 may participate in the changes in histone gene expression we detect in immature OSNs (Godfrey et al., 2009; Underwood et al., 2005).

Immature OSNs appear to manufacture cholesterol. Cholesterol is a major component of the plasma membrane, it is enriched in membrane compartments that regulate distributions of membrane proteins, and it is important for neurite outgrowth (Hayashi et al., 2004; Ko et al., 2005). Cholesterol production is energetically expensive, however. Perhaps as a result, many neurons in the brain use cholesterol uptake (in the form of apoE particles derived from astrocytes) as their primary source of membrane cholesterol after birth (Boyles et al., 1989; Lopes-Cardozo et al., 1986; Saito et al., 1987). It has been estimated that astrocytes generate roughly three times more cholesterol than neurons (Saito et al., 1987), but OSNs are not located near a pool of astrocyte-synthesized cholesterol. Our data argue that immature OSNs instead rely heavily on internal metabolism for their cholesterol needs. In particular, they express Srebf2, which regulates genes involved in cholesterol biosynthesis and is commonly expressed in cholesterol-generating astrocytes in the brain (Shimano, 2002; Tarr and Edwards, 2008). However, both immature and mature OSNs express genes encoding proteins necessary for cholesterol uptake, such as the Ldlr. We hypothesize that immature OSNs synthesize cholesterol to meet the need imposed by growing neurites, but once mature, cholesterol uptake is sufficient for OSNs.

Database of OSN gene expression

A summary of the enrichment data for every mRNA detected above a conservative threshold and the calculated probabilities of expression of each of these genes in mature OSNs, immature OSNs, and the residual sample of all other cells in the tissue is available as a supplement to this article (Supplemental database). Illustrative examples of how enrichment ratios and MDA probabilities can be used to discriminate expression patterns are shown in Table 6. The database file can be manipulated and sorted to select different degrees of risk in the prediction of genes expressed by each of the cell type categories. To facilitate assessment of the functional implications of gene expression predictions, the database also includes the functional annotation terms associated with each gene, allowing extraction of lists of genes associated with any Gene Ontology biological process, function, or cellular compartment.

In general, the MDA probabilities in the database express statistical confidence in the expression patterns of the mRNAs we detected. This confidence calculation should perform poorly only for unusual mRNAs, such as mRNAs that are expressed in the targeted cell type (OSNs, in this case) but are many-fold more abundant in the residual cell sample. This appears to be rare, according to our in situ hybridization data, perhaps because the cell types specific to the residual cell sample make only a small contribution to the High/Weak enrichment ratio. Other sources of errors are mRNAs not detected due to poor probe set performance or low abundance of the mRNA species in the bulk sample. The odorant receptors, each strongly expressed only in a small fraction of OSNs, are examples of low-abundance mRNAs. The GeneChip microarray types used to generate the enrichment ratios described herein detect odorant receptor mRNAs poorly, and so the OSN gene expression

database includes only 20 of them, most identified as mature OSN-specific. While odorant receptor genes begin to be expressed in immature OSNs, their mRNAs are probably of low abundance in immature OSNs compared with mature OSNs, which single-cell reverse transcriptase-polymerase chain reaction experiments have consistently found to be one of the most abundant mRNAs in individual OSNs (Iwema and Schwob, 2003; Malnic et al., 1999; Mizrahi et al., 2004). Lastly, the use of the Neurog1 promoter to drive GFP expression to allow FACS for immature OSNs raises the possibility that some mRNAs enriched in the GFP+++ fraction might be specific to neurally fated globose basal cells rather than immature OSNs. However, this is unlikely because enrichment ratios of several mRNAs specific to globose basal cells were used to define the enrichment ratio criteria specifically to prevent this problem.

A final caveat is that the absence of any gene from this database cannot be interpreted to mean that the gene is not expressed in OSNs because its absence may have been due to poor performance of a probe set or the absence of a few genes from the microarrays used. For this same reason, low microarray signal intensities can be interpreted to mean low abundance of an mRNA only in relative terms and only when a simultaneously processed sample shows high signal intensities for this mRNA.

Even with these caveats, these data provide an unparalleled view of gene expression in OSNs. The mature OSN data alone (Sammeta et al., 2007) provided sufficient information to contribute to the discovery of *Emx2*'s role in OR gene expression, to the identification of specific markers of nascent OSNs, sustentacular cells, and respiratory epithelial cells, to evidence that OSNs resist chemical stress via the unfolded protein response, to analysis of olfactory cilia, to work on the transcription factor *Uncx*, and to the identification of chloride channels, transporters, and tumor necrosis factor- α receptors in OSNs (Hengl et al., 2010; Lane et al., 2010; Mayer et al., 2009; McClintock et al., 2008; McIntyre et al., 2008, 2010; Nickell et al., 2007; Sammeta et al., 2010b; Sammeta and McClintock, 2010; Stephan et al., 2009). With these new data distinguishing the gene expression patterns of mature and immature OSNs, the study of molecular mechanisms in these neurons should be even further enhanced.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank N. Sammeta, J. McIntyre, T.T. Yu, S. Bose, and D. Hardin for generating images used in Supplemental Figure 1.

Grant sponsor: National Institutes of Health; Grant number: DC002736 (to T.Mc.).

LITERATURE CITED

- Adams NA, Awadein A, Toma HS. The retinal ciliopathies. *Ophthalmic Genet.* 2007; 28:113–125. [PubMed: 17896309]
- Agresti A, Coull B. Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat.* 1998; 52:119–126.

- Baker H, Grillo M, Margolis FL. Biochemical and immunocytochemical characterization of olfactory marker protein in the rodent central nervous system. *J Comp Neurol.* 1989; 285:246–261. [PubMed: 2760264]
- Barzilai A, Biton S, Shiloh Y. The role of the DNA damage response in neuronal development, organization and maintenance. *DNA Repair (Amst).* 2008; 7:1010–1027. [PubMed: 18458000]
- Bennett MK, Kulaga HM, Reed RR. Odor-evoked gene regulation and visualization in olfactory receptor neurons. *Mol Cell Neurosci.* 2010; 43:353–362. [PubMed: 20080187]
- Bishop KM, Garel S, Nakagawa Y, Rubenstein JL, O’Leary DD. Emx1 and Emx2 cooperate to regulate cortical size, lamination, neuronal differentiation, development of cortical efferents, and thalamocortical pathfinding. *J Comp Neurol.* 2003; 457:345–360. [PubMed: 12561075]
- Boyles JK, Zoellner CD, Anderson LJ, Kosik LM, Pitas RE, Weisgraber KH, Hui DY, Mahley RW, Gebicke-Haerter PJ, Ignatius MJ, et al. A role for apolipoprotein E, apolipoprotein A-I, and low density lipoprotein receptors in cholesterol transport during regeneration and remyelination of the rat sciatic nerve. *J Clin Invest.* 1989; 83:1015–1031. [PubMed: 2493483]
- Breipohl W, Laugwitz HJ, Bornfeld N. Topological relations between the dendrites of olfactory sensory cells and sustentacular cells in different vertebrates. An ultrastructural study. *J Anat.* 1974; 117:89–94. [PubMed: 4844653]
- Calof AL, Chikaraishi DM. Analysis of neurogenesis in a mammalian neuroepithelium: proliferation and differentiation of an olfactory neuron precursor in vitro. *Neuron.* 1989; 3:115–127. [PubMed: 2482777]
- Cau E, Casarosa S, Guillemot F. Mash1 and Ngn1 control distinct steps of determination and differentiation in the olfactory sensory neuron lineage. *Development.* 2002; 129:1871–1880. [PubMed: 11934853]
- Cau E, Gradwohl G, Casarosa S, Kageyama R, Guillemot F. Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium. *Development.* 2000; 127:2323–2332. [PubMed: 10804175]
- Chadwick BP, Willard HF. Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum Mol Genet.* 2003; 12:2167–2178. [PubMed: 12915472]
- Cheng LE, Reed RR. Zfp423/OAZ participates in a developmental switch during olfactory neurogenesis. *Neuron.* 2007; 54:547–557. [PubMed: 17521568]
- Cuschieri A, Bannister LH. The development of the olfactory mucosa in the mouse: electron microscopy. *J Anat.* 1975; 119:471–498. [PubMed: 1141050]
- Dahl AR, Hadley WM, Hahn FF, Benson JM, McClellan RO. Cytochrome P-450-dependent monooxygenases in olfactory epithelium of dogs: possible role in tumorigenicity. *Science.* 1982; 216:57–59. [PubMed: 7063870]
- Dahl AR, Bond JA, Petridou-Fischer J, Sabourin PJ, Whaley SJ. Effects of the respiratory tract on inhaled materials. *Toxicol Appl Pharmacol.* 1988; 93:484–492. [PubMed: 3285522]
- Davis RE, Swiderski RE, Rahmouni K, Nishimura DY, Mullins RF, Agassandian K, Philp AR, Searby CC, Andrews MP, Thompson S, Berry CJ, Thedens DR, Yang B, Weiss RM, Cassell MD, Stone EM, Sheffield VC. A knockin mouse model of the Bardet-Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. *Proc Natl Acad Sci U S A.* 2007; 104:19422–19427. [PubMed: 18032602]
- DeHamer MK, Guevara JL, Hannon K, Olwin BB, Calof AL. Genesis of olfactory receptor neurons in vitro: regulation of progenitor cell divisions by fibroblast growth factors. *Neuron.* 1994; 13:1083–1097. [PubMed: 7946347]
- Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, Lin-Marq N, Koch M, Bilio M, Cantiello I, Verde R, De Masi C, Bianchi SA, Cicchini J, Perroud E, Mehmeti S, Dagand E, Schrinner S, Nurnberger A, Schmidt K, Metz K, Zwingmann C, Brieske N, Springer C, Hernandez AM, Herzog S, Grabbe F, Sieverding C, Fischer B, Schrader K, Brockmeyer M, Dettmer S, Helbig C, Alunni V, Battaini MA, Mura C, Henrichsen CN, Garcia-Lopez R, Echevarria D, Puelles E, Garcia-Calero E, Kruse S, Uhr M, Kauck C, Feng G, Milyaev N, Ong CK, Kumar L, Lam M, Semple CA, Gyenesei A, Mundlos S, Radelof U, Lehrach H, Sarmientos P, Reymond A, Davidson DR, Dolle P, Antonarakis SE, Yaspo ML, Martinez S,

- Baldock RA, Eichele G, Ballabio A. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 2011; 9:e1000582. [PubMed: 21267068]
- Dougherty JD, Schmidt EF, Nakajima M, Heintz N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* 2010; 38:4218–4230. [PubMed: 20308160]
- Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell.* 2008; 135:749–762. [PubMed: 19013282]
- Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugenics.* 1936; 7:179–188.
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell.* 2000; 11:4241–4257. [PubMed: 11102521]
- Godfrey AC, White AE, Tatomer DC, Marzluff WF, Duronio RJ. The Drosophila U7 snRNP proteins Lsm10 and Lsm11 are required for histone pre-mRNA processing and play an essential role in development. *RNA.* 2009; 15:1661–1672. [PubMed: 19620235]
- Gokoffski KK, Wu HH, Beites CL, Kim J, Kim EJ, Matzuk MM, Johnson JE, Lander AD, Calof AL. Activin and GDF11 collaborate in feedback control of neuroepithelial stem cell proliferation and fate. *Development.* 2011; 138:4131–4142. [PubMed: 21852401]
- Graziadei PP, Metcalf JF. Autoradiographic and ultra-structural observations on the frog's olfactory mucosa. *Z Zellforsch Mikrosk Anat.* 1971; 116:305–318. [PubMed: 4931733]
- Graziadei PP, Levine RR, Graziadei GA. Regeneration of olfactory axons and synapse formation in the forebrain after bulbectomy in neonatal mice. *Proc Natl Acad Sci U S A.* 1978; 75:5230–5234. [PubMed: 283428]
- Guo Z, Packard A, Krolewski RC, Harris MT, Manglapus GL, Schwob JE. Expression of pax6 and sox2 in adult olfactory epithelium. *J Comp Neurol.* 2010; 518:4395–4418. [PubMed: 20852734]
- Hastie, Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Trans Pattern Analysis and Machine Intelligence.* 1996; 18:607–616.
- Hayashi H, Campenot RB, Vance DE, Vance JE. Glial lipoproteins stimulate axon growth of central nervous system neurons in compartmented cultures. *J Biol Chem.* 2004; 279:14009–14015. [PubMed: 14709547]
- Hengl T, Kaneko H, Dauner K, Vocke K, Frings S, Mohrlen F. Molecular components of signal amplification in olfactory sensory cilia. *Proc Natl Acad Sci U S A.* 2010; 107:6052–6057. [PubMed: 20231443]
- Hinds JW, Hinds PL, McNelly NA. An autoradiographic study of the mouse olfactory epithelium: evidence for long-lived receptors. *Anat Rec.* 1984; 210:375–383. [PubMed: 6542328]
- Hirata T, Nakazawa M, Yoshihara S, Miyachi H, Kitamura K, Yoshihara Y, Hibi M. Zinc-finger gene Fez in the olfactory sensory neurons regulates development of the olfactory bulb non-cell-autonomously. *Development.* 2006; 133:1433–1443. [PubMed: 16540508]
- Hirota J, Mombaerts P. The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proc Natl Acad Sci U S A.* 2004; 101:8751–8755. [PubMed: 15173589]
- Holcomb JD, Mumm JS, Calof AL. Apoptosis in the neuronal lineage of the mouse olfactory epithelium: regulation in vivo and in vitro. *Dev Biol.* 1995; 172:307–323. [PubMed: 7589810]
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009a; 37:1–13. [PubMed: 19033363]
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009b; 4:44–57. [PubMed: 19131956]
- Inaki K, Nishimura S, Nakashiba T, Itohara S, Yoshihara Y. Laminar organization of the developing lateral olfactory tract revealed by differential expression of cell recognition molecules. *J Comp Neurol.* 2004; 479:243–256. [PubMed: 15457507]
- Iwema CL, Schwob JE. Odorant receptor expression as a function of neuronal maturity in the adult rodent olfactory system. *J Comp Neurol.* 2003; 459:209–222. [PubMed: 12655505]

- Jenkins PM, McEwen DP, Martens JR. Olfactory cilia: linking sensory cilia function and human disease. *Chem Senses*. 2009; 34:451–464. [PubMed: 19406873]
- Kajimura D, Dragomir C, Ramirez F, Laub F. Identification of genes regulated by transcription factor KLF7 in differentiating olfactory sensory neurons. *Gene*. 2007; 388:34–42. [PubMed: 17123745]
- Kim H, Greer CA. The emergence of compartmental organization in olfactory bulb glomeruli during postnatal development. *J Comp Neurol*. 2000; 422:297–311. [PubMed: 10842233]
- Ko M, Zou K, Minagawa H, Yu W, Gong JS, Yanagisawa K, Michikawa M. Cholesterol-mediated neurite out-growth is differently regulated between cortical and hippocampal neurons. *J Biol Chem*. 2005; 280:42759–42765. [PubMed: 16267051]
- Kolterud A, Alenius M, Carlsson L, Bohm S. The Lim homeobox gene *Lhx2* is required for olfactory sensory neuron identity. *Development*. 2004; 131:5319–5326. [PubMed: 15456728]
- Koo JH, Gill S, Pannell LK, Menco BP, Margolis JW, Margolis FL. The interaction of Bex and OMP reveals a dimer of OMP with a short half-life. *J Neurochem*. 2004; 90:102–116. [PubMed: 15198671]
- Lane AP, Turner J, May L, Reed R. A genetic model of chronic rhinosinusitis-associated olfactory inflammation reveals reversible functional impairment and dramatic neuroepithelial reorganization. *J Neurosci*. 2010; 30:2324–2329. [PubMed: 20147558]
- Laub F, Lei L, Sumiyoshi H, Kajimura D, Dragomir C, Smaldone S, Puche AC, Petros TJ, Mason C, Parada LF, Ramirez F. Transcription factor KLF7 is important for neuronal morphogenesis in selected regions of the nervous system. *Mol Cell Biol*. 2005; 25:5699–5711. [PubMed: 15964824]
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. 2007; 445:168–176. [PubMed: 17151600]
- LeSauter J, Yan L, Vishnubhotla B, Quintero JE, Kuhlman SJ, McMahon DG, Silver R. A short half-life GFP mouse model for analysis of suprachiasmatic nucleus organization. *Brain Res*. 2003; 964:279–287. [PubMed: 12576188]
- Levi G, Puche AC, Mantero S, Barbieri O, Trombino S, Paleari L, Egeo A, Merlo GR. The *Dlx5* homeodomain gene is essential for olfactory development and connectivity in the mouse. *Mol Cell Neurosci*. 2003; 22:530–543. [PubMed: 12727448]
- Lim JH, Davis GE, Wang Z, Li V, Wu Y, Rue TC, Storm DR. Zicam-induced damage to mouse and human nasal tissue. *PLoS One*. 2009; 4:e7647. [PubMed: 19876403]
- Long JE, Garel S, Depew MJ, Tobet S, Rubenstein JL. *DLX5* regulates development of peripheral and central components of the olfactory system. *J Neurosci*. 2003; 23:568–578. [PubMed: 12533617]
- Lopes-Cardozo M, Larsson OM, Schousboe A. Acetoacetate and glucose as lipid precursors and energy substrates in primary cultures of astrocytes and neurons from mouse cerebral cortex. *J Neurochem*. 1986; 46:773–778. [PubMed: 3081684]
- Loyola A, Tagami H, Bonaldi T, Roche D, Quivy JP, Imhof A, Nakatani Y, Dent SY, Almouzni G. The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep*. 2009; 10:769–775. [PubMed: 19498464]
- Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C, Barnea G, Kellis M, Lomvardas S. An epigenetic signature for monoallelic olfactory receptor expression. *Cell*. 2011; 145:555–570. [PubMed: 21529909]

- Malnic B, Hirono J, Sato T, Buck LB. Combinatorial receptor codes for odors. *Cell*. 1999; 96:713–723. [PubMed: 10089886]
- Marcucci F, Zou DJ, Firestein S. Sequential onset of presynaptic molecules during olfactory sensory neuron maturation. *J Comp Neurol*. 2009; 516:187–198. [PubMed: 19598283]
- Mayer U, Kuller A, Daiber PC, Neudorf I, Warnken U, Schnolzer M, Frings S, Mohrlen F. The proteome of rat olfactory sensory cilia. *Proteomics*. 2009; 9:322–334. [PubMed: 19086097]
- McClintock TS. Achieving singularity in mammalian odorant receptor gene choice. *Chem Senses*. 2010; 35:447–457. [PubMed: 20460312]
- McClintock TS, Glasser CE, Bose SC, Bergman DA. Tissue expression patterns identify mouse cilia genes. *Physiol Genomics*. 2008; 32:198–206. [PubMed: 17971504]
- McIntyre JC, Bose SC, Stromberg AJ, McClintock TS. Emx2 stimulates odorant receptor gene expression. *Chem Senses*. 2008; 33:825–837. [PubMed: 18854508]
- McIntyre JC, Titlow WB, McClintock TS. Axon growth and guidance genes identify nascent, immature, and mature olfactory sensory neurons. *J Neurosci Res*. 2010; 88:3243–3256. [PubMed: 20882566]
- Miller AM, Treloar HB, Greer CA. Composition of the migratory mass during development of the olfactory nerve. *J Comp Neurol*. 2010; 518:4825–4841. [PubMed: 21031554]
- Mizrahi A, Matsunami H, Katz LC. An imaging-based approach to identify ligands for olfactory receptors. *Neuropharmacology*. 2004; 47:661–668. [PubMed: 15458837]
- Mombaerts P. Axonal wiring in the mouse olfactory system. *Annu Rev Cell Dev Biol*. 2006; 22:713–737. [PubMed: 17029582]
- Monath TP, Cropp CB, Harrison AK. Mode of entry of a neurotropic arbovirus into the central nervous system. Reinvestigation of an old controversy. *Lab Invest*. 1983; 48:399–410. [PubMed: 6300550]
- Nakada Y, Parab P, Simmons A, Omer-Abdalla A, Johnson JE. Separable enhancer sequences regulate the expression of the neural bHLH transcription factor neurogenin 1. *Dev Biol*. 2004; 271:479–487. [PubMed: 15223348]
- Nickell WT, Kleene NK, Kleene SJ. Mechanisms of neuronal chloride accumulation in intact mouse olfactory epithelium. *J Physiol*. 2007; 583:1005–1020. [PubMed: 17656441]
- Rafols JA, Getchell TV. Morphological relations between the receptor neurons, sustentacular cells and Schwann cells in the olfactory mucosa of the salamander. *Anat Rec*. 1983; 206:87–101. [PubMed: 6881554]
- Saito M, Benson EP, Rosenberg A. Metabolism of cholesterol and triacylglycerol in cultured chick neuronal cells, glial cells, and fibroblasts: accumulation of esterified cholesterol in serum-free culture. *J Neurosci Res*. 1987; 18:319–325. [PubMed: 3694714]
- Sammata N, McClintock TS. Chemical stress induces the unfolded protein response in olfactory sensory neurons. *J Comp Neurol*. 2010; 518:1825–1836. [PubMed: 20235094]
- Sammata N, Yu TT, Bose SC, McClintock TS. Mouse olfactory sensory neurons express 10,000 genes. *J Comp Neurol*. 2007; 502:1138–1156. [PubMed: 17444493]
- Sammata N, Hardin DL, McClintock TS. Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium. *Mol Cell Neurosci*. 2010a; 45:398–407. [PubMed: 20692344]
- Sammata N, Hardin DL, McClintock TS. Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium. *Mol Cell Neurosci*. 2010b; 45:398–407. [PubMed: 20692344]
- Schwob JE. Neural regeneration and the peripheral olfactory system. *Anat Rec*. 2002; 269:33–49. [PubMed: 11891623]
- Serizawa S, Miyamichi K, Takeuchi H, Yamagishi Y, Suzuki M, Sakano H. A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting. *Cell*. 2006; 127:1057–1069. [PubMed: 17129788]
- Shetty RS, Bose SC, Nickell MD, McIntyre JC, Hardin DH, Harris AM, McClintock TS. Transcriptional changes during neuronal death and replacement in the olfactory epithelium. *Mol Cell Neurosci*. 2005; 30:583–600. [PubMed: 16456926]

- Shimano H. Sterol regulatory element-binding protein family as global regulators of lipid synthetic genes in energy metabolism. *Vitamins and hormones*. 2002; 65:167–194. [PubMed: 12481547]
- Shimizu T, Hibi M. Formation and patterning of the forebrain and olfactory system by zinc-finger genes *Fezf1* and *Fezf2*. *Dev Growth Differ*. 2009; 51:221–231. [PubMed: 19222525]
- Song HJ, Stevens CF, Gage FH. Neural stem cells from adult hippocampus develop essential properties of functional CNS neurons. *Nat Neurosci*. 2002; 5:438–445. [PubMed: 11953752]
- Stephan AB, Shum EY, Hirsh S, Cygnar KD, Reisert J, Zhao H. ANO2 is the ciliary calcium-activated chloride channel that may mediate olfactory amplification. *Proc Natl Acad Sci U S A*. 2009; 106:11776–11781. [PubMed: 19561302]
- Storey J, Tibshirani R. Statistical significance for genome-wide studies. *Proc Natl Acad Sci U S A*. 2003; 100:9440–9445. [PubMed: 12883005]
- Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011; 332:472–474. [PubMed: 21415320]
- Tadenev AL, Kulaga HM, May-Simera HL, Kelley MW, Katsanis N, Reed RR. Loss of Bardet-Biedl syndrome protein-8 (BBS8) perturbs olfactory function, protein localization, and axon targeting. *Proc Natl Acad Sci U S A*. 2011; 108:10320–10325. [PubMed: 21646512]
- Tarr PT, Edwards PA. ABCG1 and ABCG4 are coexpressed in neurons and astrocytes of the CNS and regulate cholesterol homeostasis through SREBP-2. *J Lipid Res*. 2008; 49:169–182. [PubMed: 17916878]
- Theriault FM, Nuthall HN, Dong Z, Lo R, Barnabe-Heider F, Miller FD, Stifani S. Role for *Runx1* in the proliferation and neuronal differentiation of selected progenitor cells in the mammalian nervous system. *J Neurosci*. 2005; 25:2050–2061. [PubMed: 15728845]
- Tucker ES, Lehtinen MK, Maynard T, Zirlinger M, Dulac C, Rawson N, Pevny L, Lamantia AS. Proliferative and transcriptional identity of distinct classes of neural precursors in the mammalian olfactory epithelium. *Development*. 2010; 137:2471–2481. [PubMed: 20573694]
- Underwood JG, Boutz PL, Dougherty JD, Stoilov P, Black DL. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol*. 2005; 25:10005–10016. [PubMed: 16260614]
- Viswaprakash N, Josephson EM, Dennis JC, Yilma S, Morrison EE, Vodyanoy VJ. Odorant response kinetics from cultured mouse olfactory epithelium at different ages in vitro. *Cells Tissues Organs*. 2010; 192:361–373. [PubMed: 20664250]
- Wang SS, Lewcock JW, Feinstein P, Mombaerts P, Reed RR. Genetic disruptions of *O/E2* and *O/E3* genes reveal involvement in olfactory receptor neuron projection. *Development*. 2004; 131:1377–1388. [PubMed: 14993187]
- Wang SS, Tsai RY, Reed RR. The characterization of the *Olf-1/EBF*-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. *J Neurosci*. 1997; 17:4149–4158. [PubMed: 9151732]
- Watanabe Y, Inoue K, Okuyama-Yamamoto A, Nakai N, Nakatani J, Nibu K, Sato N, Iiboshi Y, Yusa K, Kondoh G, Takeda J, Terashima T, Takumi T. *Fezf1* is required for penetration of the basal lamina by olfactory axons to promote olfactory development. *J Comp Neurol*. 2009; 515:565–584. [PubMed: 19479999]
- Yu TT, McIntyre JC, Bose SC, Hardin D, Owen MC, McClintock TS. Differentially expressed transcripts from phenotypically identified olfactory sensory neurons. *J Comp Neurol*. 2005; 483:251–262. [PubMed: 15682396]

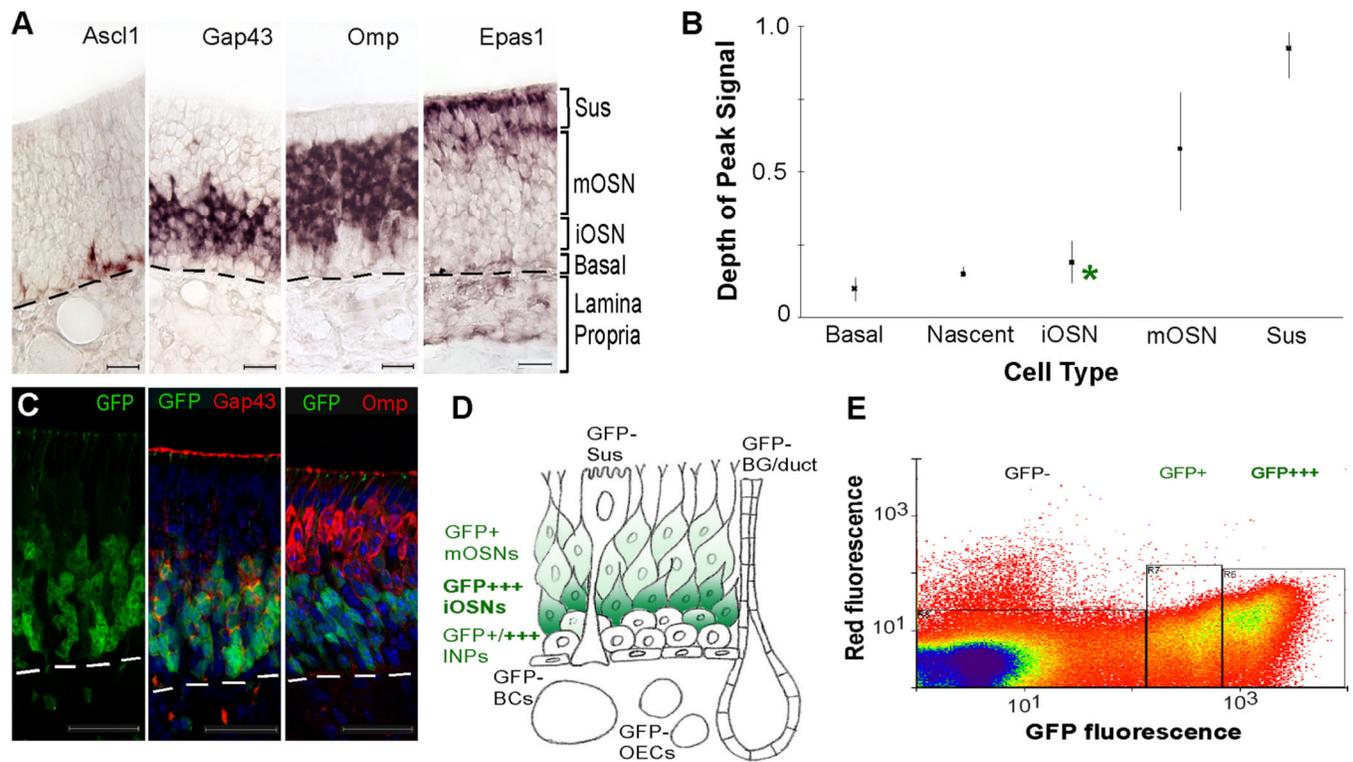


Figure 1.

Isolation of immature olfactory sensory neurons (OSNs). **A:** In situ hybridization for mRNAs that specifically label distinct cell types within the olfactory epithelium (age P22). Ascl1 (achaete-scute complex homolog 1) labels a subset of globose basal cells; Gap43 (growth-associated protein 43), immature OSNs; Omp (olfactory marker protein), mature OSNs; Epas1 (endothelial PAS domain protein 1), sustentacular cells. **B:** Quantification of depth in the olfactory epithelium of in situ hybridization signals per cell type, showing means and ranges. *, depth of the peak GFP fluorescence in TgN1-2G mice at age P21–P24. **C:** Confocal images show that strong GFP fluorescence in TgN1-2G mice overlaps with immunohistochemistry for Gap43; apically located weak GFP fluorescence corresponds to the mature OSN layer (age P7). **D:** Schematic representation of fluorescence patterns in the olfactory epithelia of TgN1-2G mice. **E:** Fluorescence-activated cell sorting (FACS) trace demonstrating the three fractions obtained from the olfactory epithelia of TgN1-2G mice. Basal, globose basal cell; BC, basal cells; BG, Bowman's gland; nascent, nascent immature olfactory sensory neuron; INP, Neurog1⁺ immediate neuronal progenitor cell; iOSN, immature olfactory sensory neuron; mOSN, mature olfactory sensory neuron; OECs, olfactory ensheathing cells; Sus, sustentacular cell. Red fluorescence, propidium iodide staining of DNA to detect dead cells. Scale bar = 50 μ m in A and C. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

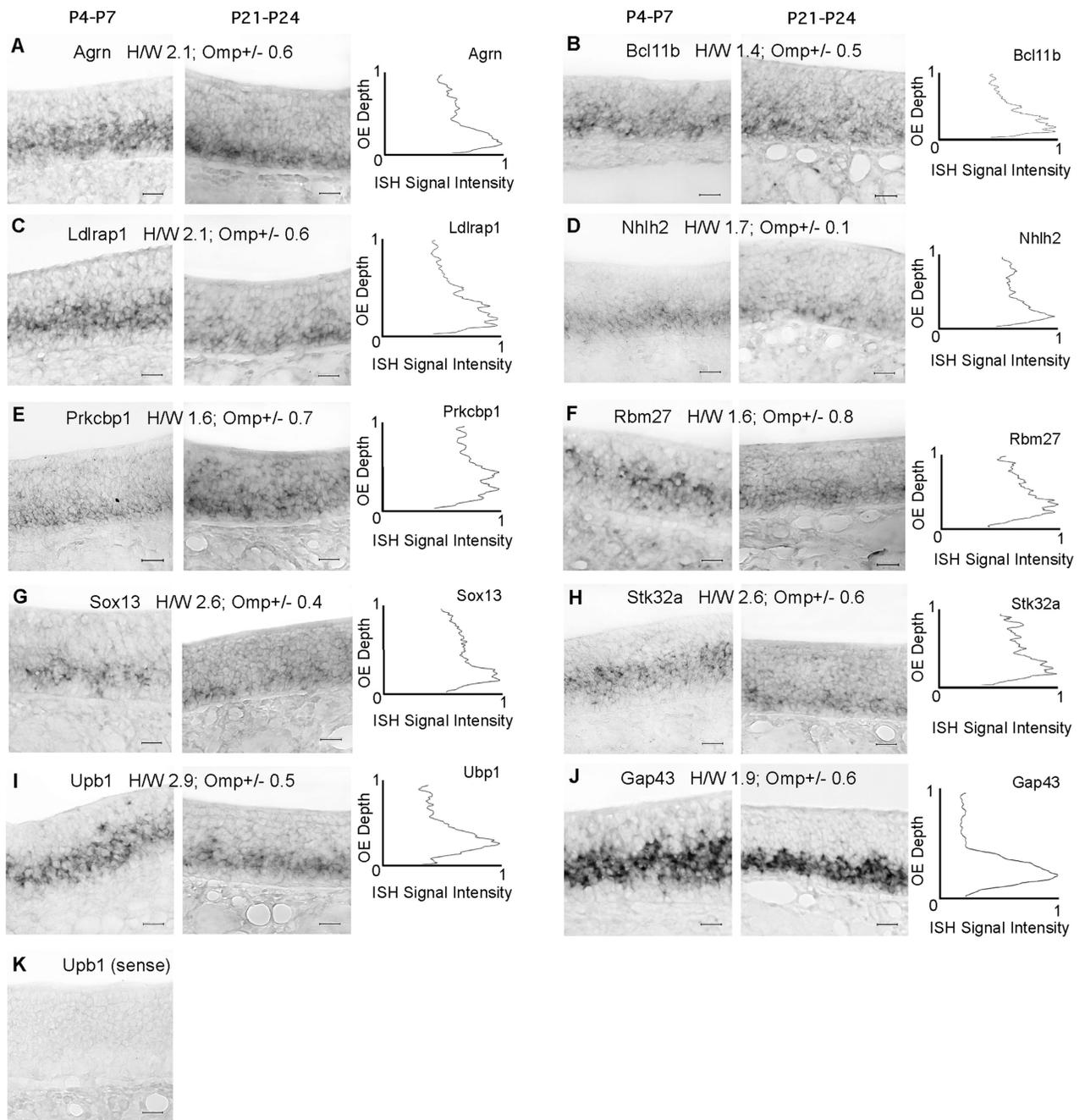


Figure 2.

In situ hybridization (ISH) for mRNAs specific to immature OSNs. Each mRNA is represented by three panels: two ages, P4–P7 (left) and P21–P24 (middle), plus a profile of signal intensity across the depth of the olfactory epithelium (right) at P21–P24. Enrichment ratio values from experiments targeting immature OSNs (H/W) and mature OSNs (Omp+/-) are shown. **A:** Agrn (agrin). **B:** Bcl11b (B-cell leukemia/lymphoma 11B). **C:** Ldlrap1 (low-density lipoprotein receptor adaptor protein 1). **D:** Nhlh2 (nescient helix loop helix 2). **E:** Prkcbp1 (protein kinase C binding protein 1). **F:** Rbm27 (RNA binding motif protein 27). **G:**

Sox13 (SRY-box containing gene 13). **H**: Stk32a (serine/threonine kinase 32A). **I**: Upb1 (ureidopropionase, beta). **J**: Gap43 (growth-associated protein 43), canonical marker of immature OSNs. **K**: Upb1 sense control. OE, olfactory epithelium. Scale bar = 25 μm in A–J.

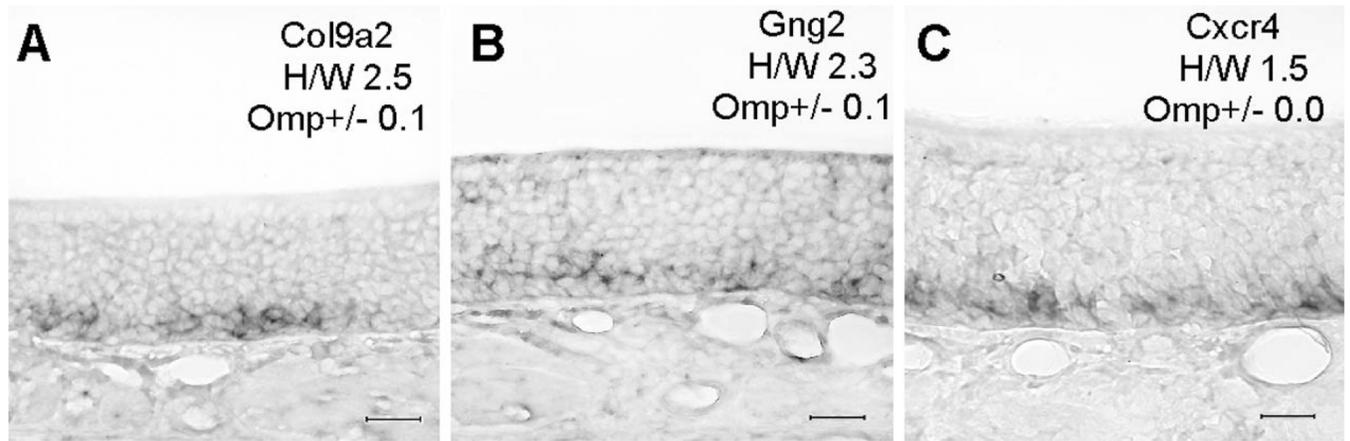


Figure 3.

In situ hybridization for putative nascent OSN mRNAs (age P21–P24). **A:** Col9a2 (collagen, type IX, alpha 2). **B:** Gng2 (guanine nucleotide binding protein [G protein], gamma 2). **C:** Cxcr4 (chemokine [C-X-C motif] receptor 4), canonical marker of nascent immature OSNs. Enrichment ratio values from experiments targeting immature OSNs (H/W) and mature OSNs (Omp+/-) are shown. ISH, in situ hybridization; Scale bar = 25 μ m in A–C.

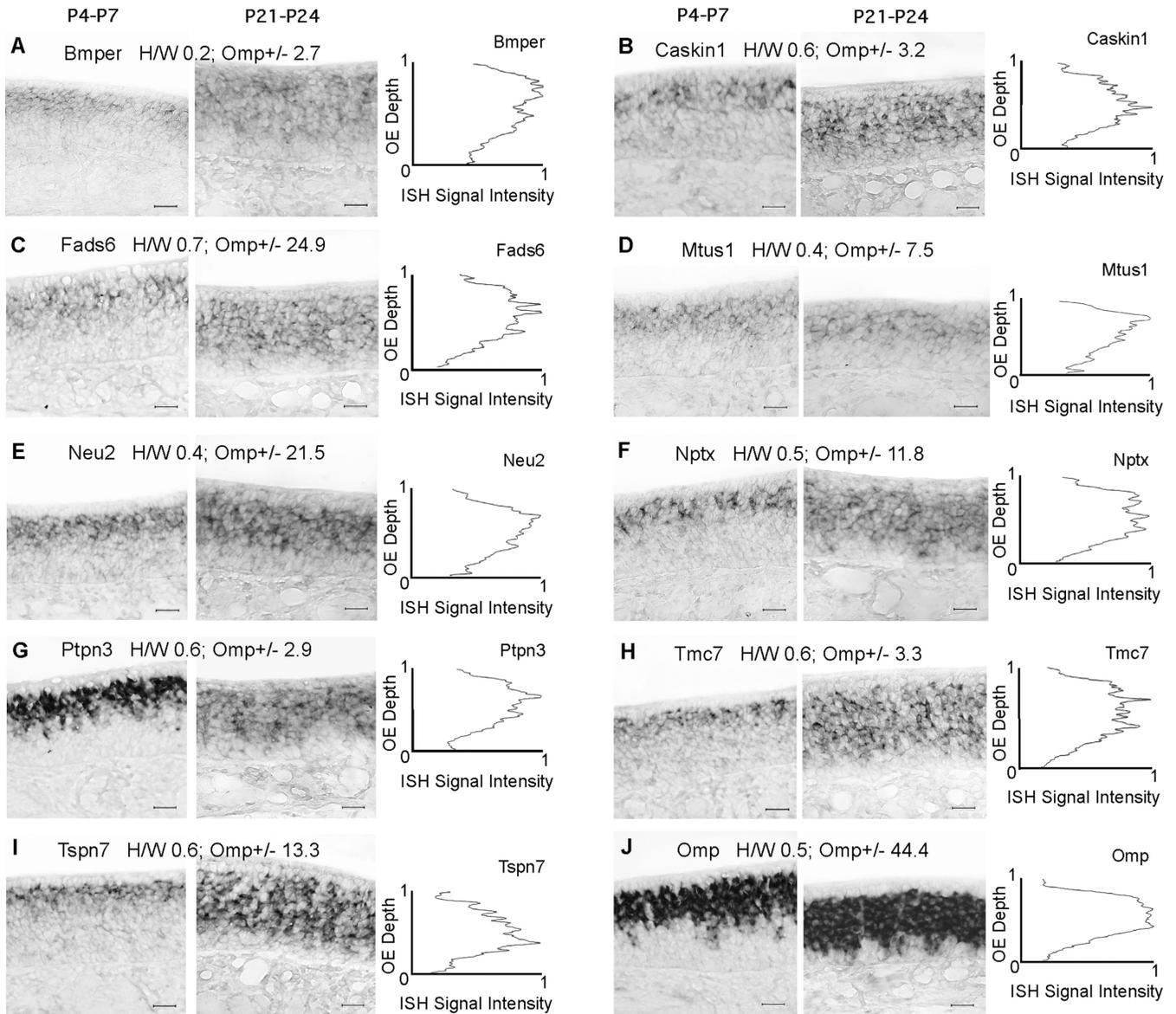
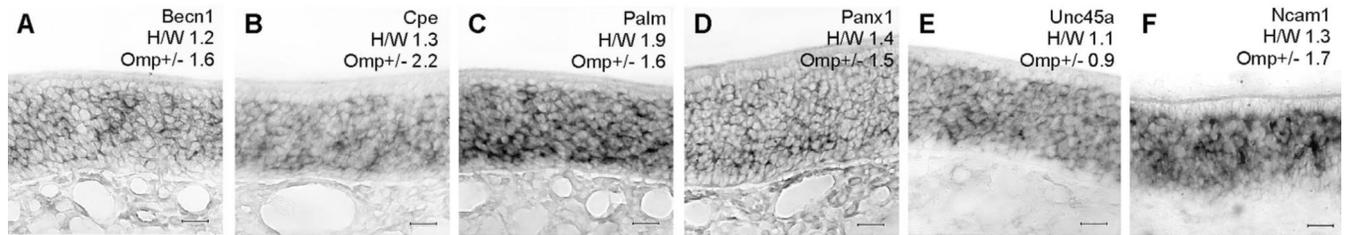


Figure 4.

In situ hybridization (ISH) for mRNAs specific to mature OSNs. Each mRNA is represented by three panels: two ages, P4–P7 (left) and P21–P24 (middle), plus a profile of signal intensity across the depth of the olfactory epithelium (right) at P21–P24. Enrichment ratio values from experiments targeting immature OSNs (H/W) and mature OSNs (Omp+/-) are shown. **A:** Bmper (BMP-binding endothelial regulator). **B:** Caskin1 (CASK-interacting protein 1). **C:** Fads6 (fatty acid desaturase domain family, member 6). **D:** Mtus1 (mitochondrial tumor suppressor 1). **E:** Neu2 (neuraminidase 2). **F:** Nptx2 (neuronal pentraxin 2). **G:** Ptpn3 (protein tyrosine phosphatase, nonreceptor type 3). **H:** Tmc7 (transmembrane channel-like gene family 7). **I:** Tspan7 (tetraspanin 7). **J:** Omp (olfactory marker protein), canonical marker of mature OSNs. OE, olfactory epithelium. Scale bar = 25 μ m in A–J.

**Figure 5.**

Examples of in situ hybridization for mRNAs shared by mature and immature OSNs (age P21–P24). Enrichment ratio values from experiments targeting immature OSNs (H/W) and mature OSNs (Omp+/-) are shown. **A:** *Becn1* (beclin 1, autophagy related). **B:** *Cpe* (carboxypeptidase E). **C:** *Palm* (paralemmin). **D:** *Panx1* (pannexin 1). **E:** *Unc45a* (unc-45 homolog A [*C. elegans*]). **F:** *Ncam1* (neural cell adhesion molecule 1), pan OSN marker. Scale bar = 25 μ m in A–F.

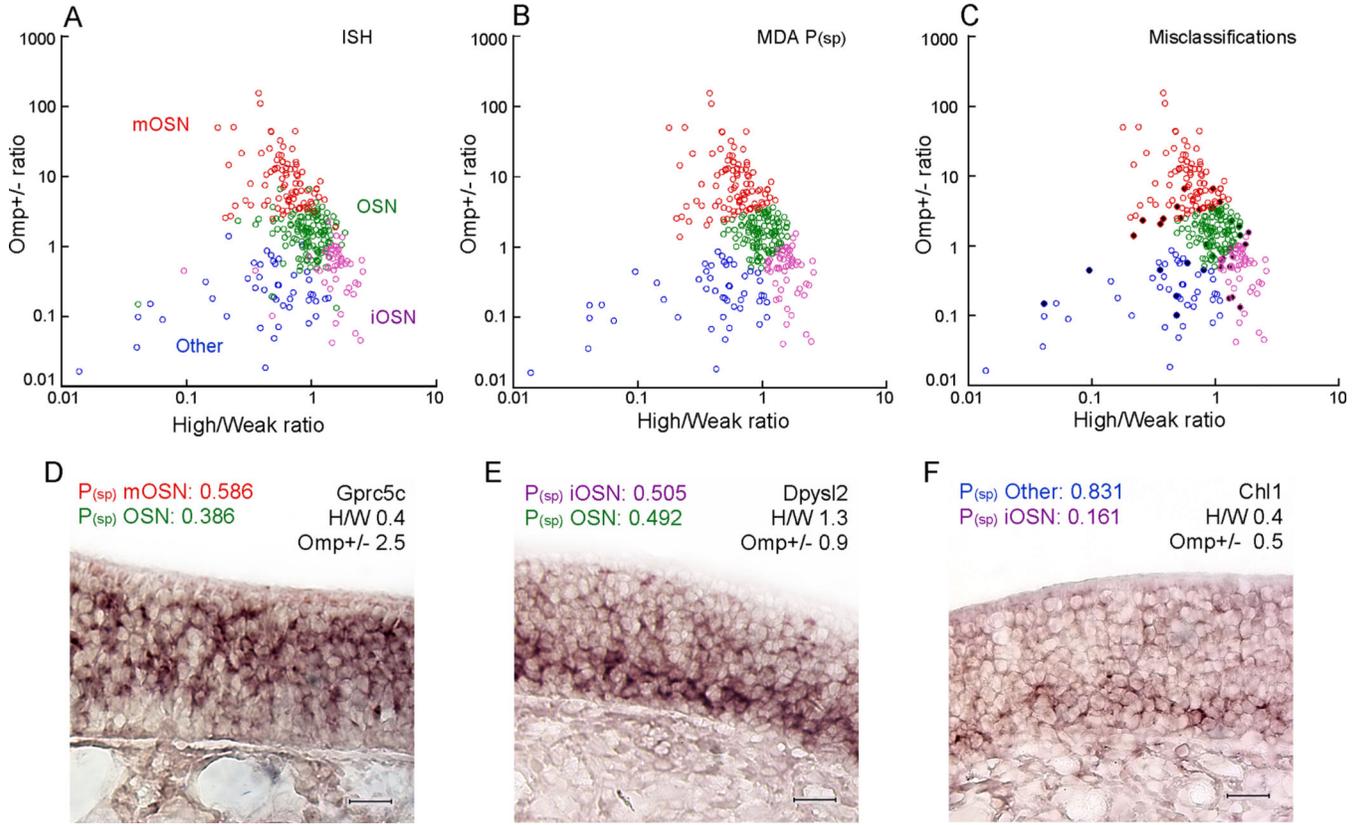


Figure 6. Mixture discriminant analysis (MDA) prediction of specific expression in cell type categories ($P_{(sp)}$) matched well with in situ hybridization data. **A:** In situ hybridization data plotted on enrichment ratio axes. Red, mature OSN; magenta, immature OSN; green, shared OSN; blue, not specific to OSNs. **B:** MDA cell type predictions plotted on enrichment ratio axes. **C:** Distribution of misclassifications (black diamonds) made by MDA $P_{(sp)}$ probabilities as identified by in situ hybridization data. **D–F:** In situ hybridization for representative misclassified mRNAs. Gprc5c, Gprotein– coupled receptor, family C, group 5, member C; Dpysl2, dihydropyrimidinase-like 2; Chl1, cell adhesion molecule with homology to L1CAM. Scale bar = 25 μ m in **D–F**. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

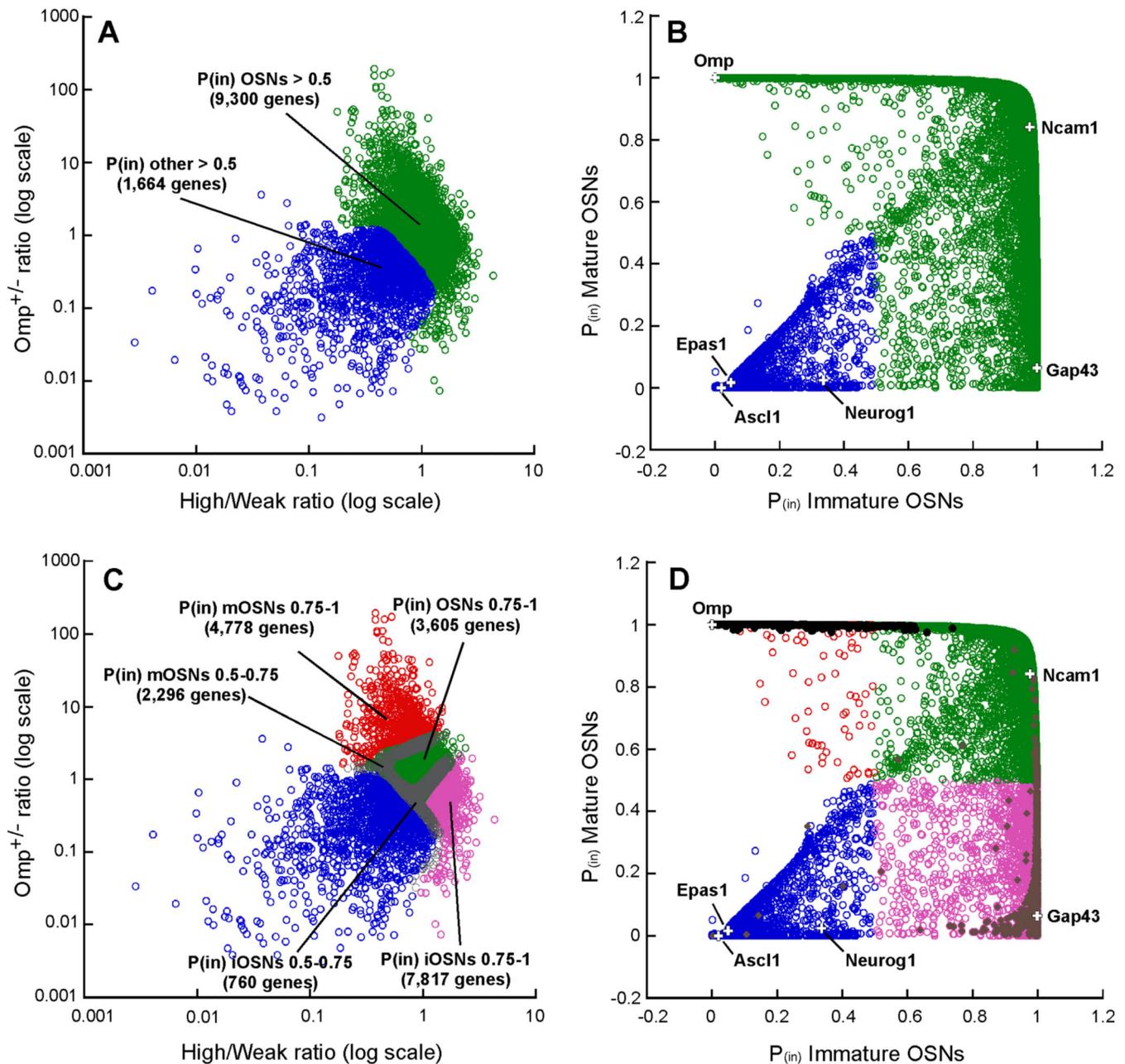


Figure 7.

Mixture discriminant analysis (MDA) $P_{(in)}$ probability distributions reveal expression patterns that grade across conventional distinctions between cell types. **A:** MDA $P_{(in)}$ predictions for all 10,976 mRNAs detected on both microarrays cluster together when plotted on enrichment ratio axes; 9,300 mRNAs whose cell type categories probabilities were highest for OSNs (green), compared with 1,664 mRNAs whose cell type categories probabilities were highest for other olfactory epithelium cell types (blue). **B:** Plotting the $P_{(in)}$ data on $P_{(in)}$ immature and $P_{(in)}$ mature axes placed the mRNAs with a high probability of expression in OSNs against the outer limits while emphasizing the distinctions between

mRNAs that had intermediate probabilities representing shared expression between cell types. **C:** MDA $P_{(in)}$ values distributed according to their probability ranges when plotted on enrichment ratio axes. Red, mature OSN genes with $P_{(in)}$ 0.75–1; magenta, immature OSN genes with $P_{(in)}$ 0.75–1; green, shared OSN genes with $P_{(in)}$ 0.75–1; gray, OSN genes with $P_{(in)}$ 0.5–0.749; blue, mRNAs whose values were highest in the other cell type category. **D:** Cell type predictions determined by $P_{(in)} > 0.5$ compared against enrichment ratio criteria predictions of specificity in mature OSNs (black) and immature OSNs (reddish brown); other colors as in C. +, the position of several known cell type markers. Red, mature OSN; magenta, immature OSN; green, shared OSN; blue, mRNAs whose values were highest in the other cell type category. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

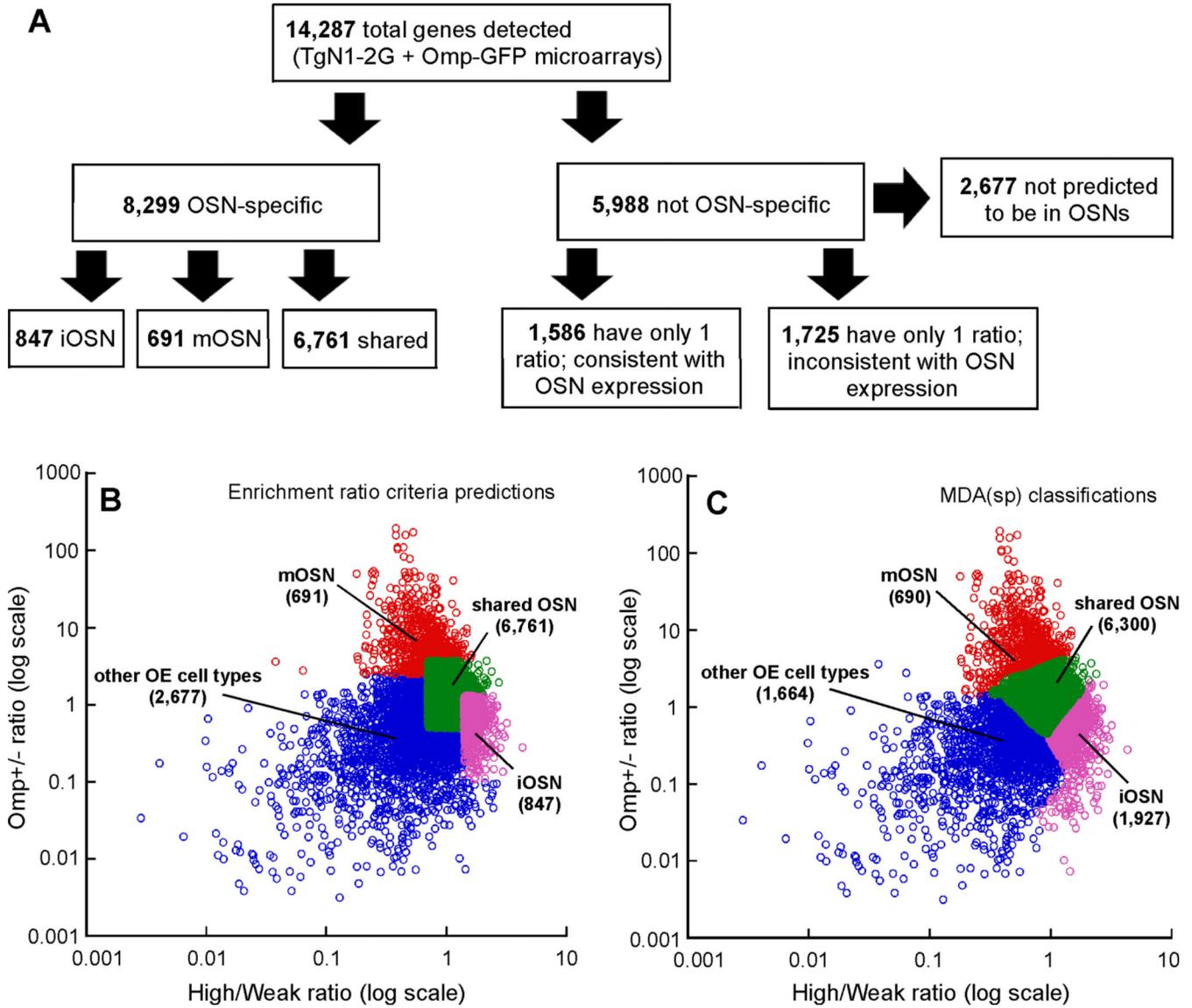


Figure 8. Summary of gene expression patterns in the olfactory epithelium and cell type predictions for all mRNAs detected. **A:** Cell type identifications based on enrichment ratio classifications. **B:** Enrichment ratio criteria prediction distributions. **C:** MDA $P_{(sp)}$ prediction distributions plotted on enrichment ratio axes. OE, olfactory epithelium. Red, mature OSN; magenta, immature OSN; green, shared OSN; blue, other cell types. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE 1

Primary Antibodies

Antiserum	Immunogen	Source (cat. no.)	Dilution	Specificity
Rabbit, polyclonal anti-growth-associated protein 43 (GAP43)	Recombinant rat GAP43 (complete sequence)	Millipore, Bedford, MA (AB5220)	1:200	Western blot using this antibody and mouse brain lysate shows an expected band of ~45 kDa; immunoreactivity matches mRNA expression pattern (Song et al., 2002; Inaki et al., 2004; McIntyre et al., 2010).
Goat, polyclonal anti-olfactory marker protein (OMP)	Purified natural rat OMP	Wako, Richmond, VA (544-10001)	1:1,000	Western blot using this antibody and mouse brain shows a single ~19-kDa band; immunoreactivity matches mRNA expression pattern (Baker et al., 1989; Koo et al., 2004; Miller et al., 2010).

TABLE 2

Enrichment Ratio Criteria Used to Define Olfactory Sensory Neuron (OSN) Cell Type Categories

OSN fraction	TgN1-2G High/Weak ratio		Omp-GFP+/- ratio (OMP+/- ratio)
Immature OSN	>1.4	and	0.1–1.3
Mature OSN	Any	and	>3.7
	<0.7	and	2.5–3.7
Shared OSN	>1.4	and	>1.3
	0.7–1.4	and	0.5–2.5
	>0.7	and	2.5–3.7

TABLE 3

Accuracy of Enrichment Ratio Criteria for Olfactory Sensory Neuron (OSN) Specificity

Cell type	No. of genes	No. tested	No. correct	% Correct (95% CI)
Immature OSN	847	51	32	63 (49–75)
Mature OSN	691	85	79	93 (85–97)
Shared OSN	6,761	135	109	81 (73–87)

TABLE 4

Overrepresented Gene Ontology Biological Process Categories Among Transcripts Specific to Immature Olfactory Sensory Neurons¹

Gene ontology term	No. of genes
Nucleic acid metabolism/processing (11)	44
RNA metabolism/processing (10)	42
RNA splicing	22
Gene expression (9)	146
Regulation of transcription (4)	128
Chromatin organization/modification (5)	31
Neuron differentiation	24
Axonogenesis (4)	18
Sterol biosynthesis/metabolism (6)	13

¹ Related categories were combined and organized under the broadest overrepresented category in the Gene Ontology hierarchy. Parentheses indicate the number of categories combined into one term. Genes may appear in multiple related categories.

TABLE 5

Overrepresented Gene Ontology Biological Process Categories Among Transcripts Specific to Mature Olfactory Sensory Neurons¹

Gene ontology term	No. of genes
Ion transport (9)	50
Ion transport	50
Cation transport (6)	47
Metal ion transport	44
Transmembrane transport	34
Cell–cell signaling (9)	20
Cell–cell signaling	20
Synaptic transmission (3)	19
Regulation of synaptic plasticity	7
Transmission of nerve impulse	19
Secretion (3)	17
Cell projection organization (2)	27
Cell projection organization	27
Microtubule-based process	14

¹ Related categories were combined and organized under the broadest overrepresented category in the Gene Ontology hierarchy. Parentheses indicate the number of categories combined into one term. Genes may appear in multiple categories.

TABLE 6

Examples of MDA Probability Patterns Found in the Olfactory Sensory Neuron (OSN) Gene Database¹

Gene symbol	High/Weak ratio	Omp +/- ratio	mOSN P _(sp)	iOSN P _(sp)	Shared P _(sp)	Other P _(in)	mOSN P _(in)	iOSN P _(in)
D930002L09Rik	0.4	103.9	1.000	0.000	0.000	0.000	1.000	0.000
6430500D05Rik	0.2	49.9	1.000	0.000	0.000	0.000	1.000	0.000
D430019H16Rik	2.4	0.2	0.000	0.991	0.004	0.005	0.004	0.995
2810408A11Rik	2.2	0.3	0.000	0.998	0.007	0.005	0.007	0.995
2210403K04Rik	0.8	1.3	0.031	0.032	0.920	0.017	0.951	0.952
4933439F18Rik	0.8	1.3	0.028	0.035	0.920	0.017	0.948	0.955
5430402E10Rik	0.0	0.1	0.000	0.000	0.000	1.000	0.000	0.000
5430413K10Rik	0.0	0.0	0.000	0.000	0.000	1.000	0.000	0.000
2700094K13Rik	1.0	0.4	0.000	0.429	0.277	0.294	0.277	0.706
Zfp945	0.9	0.5	0.000	0.248	0.424	0.327	0.425	0.673

¹ Only a subset of the columns of information contained in the OSN gene database are shown. Mixture discriminant analysis (MDA) probability columns are ordered by relevance to cell type specificity and maturation of OSNs. Bold, highest probabilities for each gene.

TABLE 7

Confusion Matrix for mRNAs Classified Into Cell Type Categories According to Their Largest Mixture Discriminant Analysis (MDA) Probabilities Compared With In Situ Hybridization Data

MDA classification	In situ hybridization			
	Other	Immature	Shared	Mature
Other	38	4	3	0
Immature	1	42	5	0
Shared	4	12	133	12
Mature	1	0	4	76

TABLE 8Significantly Overrepresented Gene Ontology Biological Process Categories in Mature Olfactory Sensory Neurons¹

Gene ontology term	No. of genes
Sensory perception of smell	37
Photoreceptor cell maintenance	9
Striatum development	7
Transport (14)	1,398
Protein transport (4)	520
Intracellular protein transport	151
Golgi organization	18
ER to Golgi vesicle-mediated transport	33
Transmembrane transport	262
Vesicle-mediated transport (6)	175
Exocytosis	76
Neurotransmitter transport	17
Calcium ion-dependent exocytosis	17
Synaptic vesicle exocytosis	7
L-glutamate transport	6
Sodium ion transport	63
Proton transport	42
Electron transport chain (3)	95
Respiratory electron transport chain	8
Mitochondrial electron transport, ubiquinol to cytochrome c	8
ER overload response	6
Protein amino acid deacetylation	9
Metabolic/biosynthetic processes (3)	23

¹ Related categories were combined and organized under the broadest category. Parentheses indicate the number of associated categories combined into one term. Genes may appear in multiple categories.

TABLE 9

Significantly Overrepresented Gene Ontology Biological Process Categories in Immature Olfactory Sensory Neurons¹

Gene ontology term	No. of genes
Nucleic acid metabolism/processing (16)	481
RNA elongation from RNA polymerase II promoter	8
RNA metabolism/processing (10)	419
RNA splicing (3)	217
Nucleic acid transport (2)	62
Gene expression (13)	1,662
Transcription (4)	1,565
Transcription initiation (2)	42
Regulation of transcription	1,453
Chromatin organization/modification (8)	193
Chromatin organization	9
Chromatin modification (6)	193
Chromatin silencing by rDNA	5
Gene silencing by RNA	19
DNA repair (4)	207
Ribosome biogenesis/assembly (3)	57
Ribosome biogenesis (3)	57
Ribosomal small subunit biogenesis	13
Ribosomal small subunit assembly	15
Translation (7)	442
Protein metabolism/processing (23)	233
Lipid membrane components biosynthesis/metabolism (2)	23
Isoprenoid biosynthetic process	15
Phosphatidylinositol metabolic process	8
Neuronal processes (2)	15
Positive regulation of dendrite morphogenesis	9
Regulation of actin cytoskeleton organization	6
Development (3)	28
Mesoderm development	15
Anterior/posterior axis specification	6
Establishment of cell polarity	7
ER overload response	6
Induction of apoptosis via death domain receptors	5
Transport (7)	520
Intracellular protein transport (6)	520
Golgi organization	18
Golgi vesicle transport	7
Intra-Golgi vesicle-mediated transport	7

Gene ontology term	No. of genes
Retrograde vesicle-mediated transport, Golgi to ER	8
Peroxisome organization	15
Vesicle-mediated transport	175
Electron transport chain (5)	95
Electron transport chain	95
ADP biosynthetic process	7
Mitochondrial electron transport (3)	21
Iron-sulfur cluster assembly	8
Mitochondrial respiratory chain complex I complex assembly	7
Mitochondrial electron transport, ubiquinol to cytochrome c	6

¹ Related categories were combined and organized under the broadest category. Parentheses indicate the number of associated categories combined into one term. Genes may appear in multiple categories.

TABLE 10

Significantly Overrepresented Gene Ontology Biological Process Categories for Genes Shared by All Olfactory Sensory Neurons¹

Gene ontology term	No. of genes
Transport (9)	1,398
Protein transport (6)	520
Intracellular protein transport (4)	151
Golgi organization	18
ER to Golgi vesicle-mediated transport	33
Retrograde vesicle-mediated transport, Golgi to ER	8
Peroxisome organization	15
Vesicle-mediated transport	175
Proton transport	42
Protein catabolic process (5)	233
Ubiquitin-dependent protein catabolic process	152
Protein ubiquitination (2)	96
Mo-molybdopterin cofactor biosynthetic process	6
Protein amino acid N-linked glycosylation via asparagine	6
Electron transport chain (3)	95
ATP synthesis coupled proton transport (2)	32
Mitochondrial electron transport, ubiquinol to cytochrome c	8
ER overload response	6

¹ Related categories were combined and organized under the broadest category. Parentheses indicate the number of associated categories combined into one term. Genes may appear in multiple categories.