A Method for the Rapid Automated Assessment of Olfactory Function

Steven L. Youngentob

Department of Neuroscience and Physiology, SUNY Upstate Medical University, 750 East Adams St., Syracuse, NY 13210, USA

Correspondence to be sent to: Dr Steven L. Youngentob, Department of Neuroscience and Physiology, SUNY Upstate Medical University, Institute for Human Performance, 750 East Adams Street, Syracuse, NY 13210. USA. e-mail: youngens@upstate.edu

Abstract

We have developed a strategy for the rapid high-throughput screening of odor responsivity in genetically altered mice (in fact, any experimentally altered animal). Specifically, the report presents the development and validation of a fully automated procedure based on the evaluation of an animal's stimulus-induced reflexive breathing response (i.e. sniffing behavior) to both air and odorant stimuli. The method requires no training of the animal to be screened and the outcome of the evaluation yields an operationally defined measure. Briefly, using whole-body plethysmography, the procedure determines the numerical values for a set of 14 respiratory measures in response to the presentation of air and a well-above-threshold concentration of the odorant propanol. These measures of stimulus-induced sniffing are incorporated into a model that defines a single univariate measure of response behavior, or 'Sniffing Index', for each screened animal. The approach significantly discriminated between the reflexive sniffing response of a control group of mice and that of an experimentally defined manipulated group for which, a priori, we expected to observe a robust altered breathing response to odorant stimulation (i.e. non-odor-aversion-conditioned versus odor-aversion-conditioned C57BL/6J mice). Further, the procedure was able to significantly discriminate between a mutant phenotype with documented alterations in physiologic and behavioral function (namely, the OMP-null mutant), and their background strain. In addition, applying epidemiologic screening principles to the observed data, we established an operational procedure for the evaluation of unknown animals.

Key words: high-throughput screening, odor-guided behavior, olfactory function, random mutagenesis, sniffing, transgenic mice

Introduction

Despite the progress in our understanding of the organization of the olfactory system anatomically, molecularly and physiologically, olfactory-guided behavior remains poorly understood mechanistically. Standard approaches, such as anatomical and pharmacological experimental manipulations (Doty and Risser, 1989; Youngentob *et al.*, 1997; Apfelbach *et al.*, 1998; Lu and Slotnick, 1998; Greiner *et al.*, 1999, 2001; Yee and Wysocki, 2002), aside, the application of modern molecular genetic approaches to the field of olfaction has greatly accelerated our understanding of the workings of this system. Indeed, a variety of new and important insights on function and behavior have come from the specific experimental manipulation of well-characterized genes (e.g. Brunet *et al.*, 1996; Buiakova *et al.*, 1996; Mombaerts *et al.*, 1996; Cau *et al.*, 1997; Bulfone *et al.*, 1998; Youngentob and Margolis, 1999; Wong *et al.*, 2000; Youngentob *et al.*, 2001a; Luo *et al.*, 2002; Cutforth *et al.*, 2003; Youngentob *et al.*, 2004).

At present, the number of targeted mutations (i.e. knock-out, knock-ins, etc.) is rapidly increasing. Nonetheless, the number of uncharacterized genes is still rather large and an alternate approach that is currently being pursued by the Trans-NIH supported mouse mutagenesis centers (see: the Neurogenomics Project at Northwestern University @http://genomics.northwestern.edu/neuro/, the Neuroscience Mutagenesis Facility at The Jackson Laboratory @ http://nmf.jax.org/ and the Neuromutagenesis Project of the Tennessee Mouse Genome Consortium @ http://www.tnmouse.org/neuromutagenesis/index.html) is to identify genes based upon the phenotypes they confer. That is, the centers are systematically mutagenizing mice and evaluating the progeny of these animals for individuals with altered phenotypes (see review by O'Brien and Frankel, 2004). Thus, the need to behaviorally screen mice (prior to an investment in the rigorous and time-consuming methods typically associated with psychophysical evaluations of olfaction) resulting from the manipulation of well characterized genes, as well as those from the now functioning mouse mutagenesis centers, continues to grow at a rapid pace.

Given these conceptually different, yet parallel, avenues for the production of genetically altered animals, one task required of those that would account for olfactory-guided behavior becomes defining the consequences of genetic manipulation in terms of system function. To this end, we have developed a strategy for the rapid evaluation of genetically altered mice (indeed, any experimentally altered animal). In this report, we present a fully automated screening procedure based on the evaluation of an animal's stimulus-induced reflexive breathing response (i.e. sniffing behavior) to both air and odorant stimuli. The procedure requires no training of the animal to be screened and the outcome of the evaluation yields an operational measure of olfactory performance.

Materials and methods

The SUNY Upstate Medical University's Committee for the Humane Use of Animals approved all methods applied to these studies.

Sniff recording apparatus and testing chamber

A continuous-flow RM-80 Respiration Frequency and Volume Monitoring System (Columbus Instruments, Columbus, OH) was used to monitor changes in respiration generated by the presentation of air or odorant stimuli. The RM-80 measured changes in respiration by detecting minute changes created by an unrestrained animal's breathing inside a testing chamber (Figure 1). The changes in pressure inside the chamber reflected the depth of respiration while the cyclic variations of the pressure signal over time revealed respiration frequency. This basic technique is called whole-body plethysmography. For the present application of the technique, the RM-80 system was ideal. That is, it was highly sensitive (i.e. it can monitor a 20 g mouse in a 300 ml chamber) and, more importantly, it could be

employed with a chamber through which a constant flow of air could pass. As outlined below, odorant stimuli were presented and exhausted from the monitoring chamber with the same level of stimulus control that is standard with our more rigorous psychophysical testing paradigms (e.g. Youngentob et al., 1997, 2001a). The output of the RM-80 system, in turn, was monitored and analyzed by the same computer (see below) that also controlled, on a trial-by-trial basis, the timing and data collection contingencies required both for testing and for the generation and delivery of the odorant stimuli.

The testing chamber (internal volume = 300 ml) consisted of a Plexiglas cylinder with a Teflon conical input and output designed to permit both rapid onset and clean out of the stimulus. A perforated disk at the input served as a stimulus-diffusing screen. Air from the chamber was continuously exhausted through the conical output at the rear of the chamber as a consequence of the positive flow generated by the olfactometer. The testing chamber was paired with a reference chamber of identical design and internal volume. Both chambers were housed inside a black Plexiglas enclosure (dimensions: $56 \times 42 \times 32$ cm) with a translucent lid (Figure 1).

Sniffing characteristics

In a previous study from our laboratory (Youngentob et al., 1987), we demonstrated that of 52 direct and derived measures used to characterize sniffing behavior during the performance of an odorant detection task, only a subset of these variables was reliably altered between both air and odorant trials and between different odorants. More importantly, it was determined that a single univariate response measure incorporating all the characteristics of the subset, along with their corresponding weights, was better at describing how sniffing varied with odorant stimuli than any individual
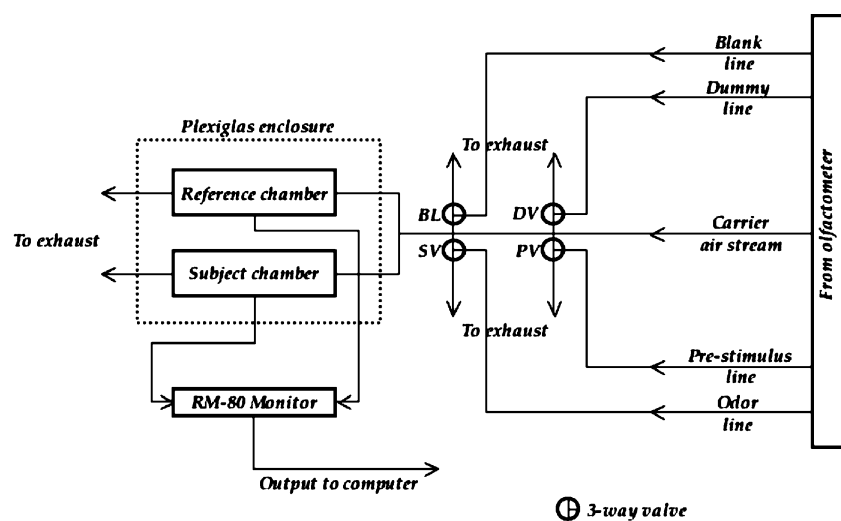

Figure 1 A schematic diagram of the testing apparatus and stimulus delivery system. BL: blank valve, DV: dummy valve, SV: odorant stimulus valve, PV: pre-stimulus valve. See text for details.

measure. Thus, based on our previous experience with this type of physiological assessment, we measured the following 14 basic parameters of reflexive sniffing: respiratory frequency, volume of an inspiratory (expiratory) sniff, number of inspiratory (expiratory) sniffs, duration of an inspiratory (expiratory) sniff, average flow rate of an inspiratory (expiratory) sniff, total inspiratory (expiratory) volume, peak flow rate of an inspiratory (expiratory) sniff and total apneic time. Further, we developed a univariate response measure incorporating the linear combination of response characteristics that maximized the discrimination between control and experimental animals (see below).

Control of testing procedure and delivery of stimulus

A computer controlled the contingencies required for testing, as well as the generation of the odorant stimuli. Odorant stimuli were generated using our standard flow-dilution olfactometer according to previously established methods and procedures (e.g. Youngentob *et al.*, 1997; Youngentob and Margolis, 1999). Briefly, computer-driven electronic mass flow controllers (EMFCs) (Teledyne Hastings Raydist Co., Hampton, VA) controlled volume flow rates through the olfactometer. A customized Balston air and nitrogen generation system (Balston Co., Lexington, MA) provided filtered, deodorized and de-humidified air to the diluent lines of the olfactometer, the carrier air and three additional air lines, designated 'pre-stimulus', 'dummy' and 'blank' (Figure 1). Saturated vapor was produced by sparging nitrogen through an odorant-filled 250 ml gas washing bottle immersed in a constant temperature bath (Neslab Inc., Newington, NH) maintained at 10°C. Odorant saturated nitrogen, in turn, was diluted to produce the desired final concentration by varying the volume flow rate ratio of odorized to diluent air within a single dilution stage of the olfactometer. The individual dilution stage was also connected to a separate, computer-activated, three-way Teflon solenoid valve (Mace Co., South El Monte, CA) to which it provided odorized air over a specified range of odorant concentrations. The odorant stimulus valve plus the pre-stimulus, dummy and blank valves, converged into a common Teflon line and the carrier airstream. The output of the common line was connected to both the behavioral test chamber and the reference chamber.

The final flow output through any one of the odorant stimulus valves, and the pre-stimulus, dummy and blank valves was a constant 400 cc/min, while the carrier airstream was maintained at 3600 cc/min at all times. During the inter-trial interval, flow through the dummy valve was normally open (see Figure 2) to the common line and carrier airstream. During the pre-stimulus interval, the dummy valve was closed simultaneously with the activation (i.e. opening) of the pre-stimulus valve. During the presentation of a stimulus (either air or odorant), the dummy valve remained closed and the pre-stimulus valve was closed simultaneously with the activation (i.e. opening) of the stimulus valve. Thus, the total output flow
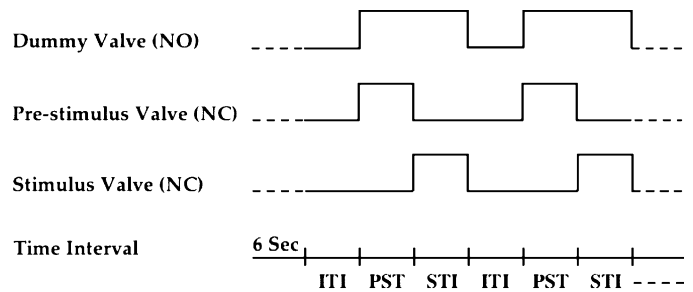
Figure 2 Valve status and sequence of activation during testing. The upward deflection in the timing sequence for each valve indicates a change in activation status (and thus a shift of flow through the valve) per 6 s period. For example, the dummy valve, which is normally open (NO) to the common manifold, is diverted to exhaust during the pre-stimulus and stimulus presentation intervals. Two complete trial cycles are illustrated. NO = normally open; NC = normally closed; ITI = inter-trial interval; PST = pre-stimulus interval; STI = stimulus interval.

through the common Teflon line was a constant 4.0 l/min. The 4.0 l/min flow was, in turn, split to provide a total airflow of 2.0 l/min to both the subject and reference chambers. It should be noted that the switching between valves was balanced so that no pressure or flow cues were available to the animal or added signal noise to the sniff-monitoring system. As with all our odorant delivery systems, stimulus input to the testing and reference chambers was designed to minimize dead space and maximize rapid stimulus transition.

Odorant

The current rapid assessment procedure was designed from the perspective that its purpose was to rapidly identify whether a particular perturbation (genetic or otherwise) produced a robust generalized effect on olfactory function. Therefore, for the present procedure we used the single odorant propanol at a concentration of 5% of vapor saturation at 10°C. Propanol has the advantage of both rapid onset and clean-out in the testing chamber, as well in the olfactometer, itself. Moreover, propanol has previously been shown to give a robust behavioral response in several mice species (Youngentob and Margolis, 1999; Youngentob *et al.*, 2001b) as well as other experimental animals such as rats (Youngentob *et al.*, 1997). Finally, threshold testing for this odorant represents the only example where targeted gene deletion resulted in a loss in behavioral odorant sensitivity without anosmia (i.e. the OMP-null mutant; Youngentob and Margolis, 1999). Thus, as outlined below, we validated the efficacy of our screening procedure to identify the OMP-null mutant as 'interesting' relative to our previous ability to do so, using more rigorous psychophysical methods.

General experimental protocol

Prior to the initiation of data collection in a testing session, each animal received a brief habituation period in the apparatus that consisted of 40 air-only trials. Stimuli were presented to the test and reference chambers according to the

following scheme (Figure 2): a trial cycle started with a 6 s inter-trial interval period; at the end of the fixed interval a 6 s pre-stimulus period was monitored and the appropriate stimulus delivered for an additional 6 s. Each of the forgoing 6 s transition times were employed with the restriction that, beginning with the end of the inter-trial interval, each subsequent period was initiated on the detection of the beginning of an inspiration by the computer.

Stimuli were randomly presented in 10 pseudo blocks of 20 presentations (i.e. 10 air and 10 odorant) and testing continued for a total of 100 air and 100 odorant stimulus presentations. The entire automated testing procedure for an individual untrained animal (i.e. habituation and testing trials combined) required ~72 min to complete.

For each stimulus presentation, the sniffing patterns produced in response to both air and odorant were analyzed by computer, which determined the numerical values for each of the 14 characteristics of reflexive sniffing.

Results

Phase 1

In this and all subsequent phases of development we standardized the age of the animals at seven weeks. Consequently, for this study, 40 seven-week-old C57BL/6J mice (20 males and 20 females) were randomly allocated to either the control (10 males and 10 females) or experimental (10 males and 10 females) groups.

Phase 1 was designed to evaluate whether, and to what degree, we could define a linear combination of respiratory parameters that would discriminate between the reflexive sniffing response of a control group of mice and that of an experimentally defined manipulated group for which, a priori, we expected to observe a robust altered breathing response to odorant stimulation. To accomplish this, we applied a modified version of the odor aversion-conditioning paradigm described by Wysocki *et al.* (1977) and Pourtier and Sicard (1990) to the experimental animals. As previously demonstrated by these studies, this procedure induces a clear concentration-dependent avoidance of aromatized solutions.

All mice were individually housed in standard micro-isolator cages supplied by SUNY Upstate Medical University's Department of Laboratory Animal Resources. For the odor aversion-conditioning procedure, each experimental mouse was provided access to drinking water for 15 min per day, over seven consecutive days, in a micro-isolator cage identical to its home cage. This water restriction procedure was done in order to ensure rapid fluid consumption during aversion conditioning on the eighth day. Water was available through a stainless steel sipper tube centered in an odor sampling port (diameter: 12 mm) that was mounted into the front of the cage. The sniffing port, in turn, was connected to the output of the same olfactometer used for behavioral testing. On each of two consecutive days of conditioning, an aqueous solution of 5% (v/v) propanol flavored drinking water (CS) was presented to each experimental mouse for 15 min. In addition, 5% of vapor saturation propanol odor (@ 10°C) was simultaneously presented to the sniffing port at a flow of 2.0 l/min (identical to the flow used in later testing). Immediately following each of the two exposures, the mice were injected i.p. with 0.3 M LiCl (lithium chloride) at a dose of 0.02 ml/g body weight (UCS). It should be noted that the two consecutive days of conditioning strategy permitted an evaluation of the efficacy of the procedure for each individual animal. In this regard, without exception, each conditioned mouse approached the sipping tube on the second day and upon detecting the propanol odor retreated to the rear of the cage. Consequently, to prevent dehydration on the second conditioning day each animal was given subcutaneous injections of sterile water following recovery from LiCl-induced illness.

Twenty-four hours following the second aversion-conditioning session, each experimental mouse and a sex- and age-matched control animal had their stimulus-induced respiratory response to air and odor stimuli monitored. As previously noted, the initial objective of Phase 1 was to establish a linear combination of the 14 respiratory measures that best distinguished between the reflexive sniffing response of control and experimental mice. For this purpose, the data set consisted of the 14 measures in response to each of 10 air and 10 odorant presentations in each of the 10 pseudo-blocks of trials for each of the 10 male and 10 female aversion-conditioned and control animals. Consequently, for each of the 14 measures, 39 mice (note: one aversion-conditioned mouse died following the second LiCl injection), each receiving 10 blocks of stimuli, provided $39 \times 10 = 390$ averages of 10 air responses and 390 averages of 10 odorant responses. In order to take advantage of the anticipated redundancy in the 14 measures [i.e. the possibility that some characteristics may vary together or reflect different measures of the same underlying change in sniff response (Youngentob *et al.*, 1987)], the principal components of the standardized measures were determined. In this respect, the goal of the principal components analysis was to summarize our multivariate data set (i.e. 14 response variables), as accurately as possible, using a fewer number of uncorrelated variables (Morrison, 1967). Using least squares multiple regression analysis of the principal components, we then determined that only the second principal component met the predictive error criterion for variable selection ($F \geq 2.0$) originally described by Mozell *et al.* (1984). Indeed, the second principal component not only exceeded our standard $F \geq 2$ criterion, but rather, significantly differed between control and odor aversion-conditioned animals [$F(1,779) = 341.1$, $P = 0$]. More importantly, the analysis established the set of 14 coefficients, which when applied in a linear combination of the 14 respiratory measures of stimulus-induced sniffing, established a single numerical value that was determined for each trial.

With the coefficients defined, we developed a regression model that yielded a single value defining the 'Sniffing Index'

for each individual animal. As illustrated in Figure 3, the model segregated the two groups of mice into distinctly different population distributions. The model, which significantly discriminated between control and aversion-conditioned animals [$F(3,35) = 27.3$, $P = 0$], incorporated the mean numerical value calculated for all trials in a testing session, the difference between the mean value calculated for the air and odor trials, a block trend indicator (i.e. a measure of change in response over the 10 blocks of trials), and the appropriately corresponding regression coefficients for each term plus a constant [it should be noted that, in developing the model we could find no evidence to warrant the inclusion of sex as a variable (component $T = 0.9721$; $P = 0.33$)].

Formal evaluation of the above data clearly demonstrated the utility of the method to rapidly evaluate two defined populations of animals in an experimental setting. However, one thrust of these studies was to develop a screening procedure by which a single 'unknown' animal could be evaluated against a reference population. To examine initially the diagnostic efficacy of the approach, we examined the distribution of Sniffing Index values calculated for the individual animals in the control and experimental groups (Figure 3). As an example of an initial screening criterion, an upper and lower 5% cutoff value in the control (i.e. reference) population would produce an estimated 10% false positive rate or 90% specificity (that is, the proportion of true negative animals which are, in fact, indicated as negative). Consequently, setting the 5% upper and lower cutoff point for the 20 controls, we evaluated the results for the odor aversion-conditioned mice. As can be seen in Figure 3, only one aversion-conditioned animal
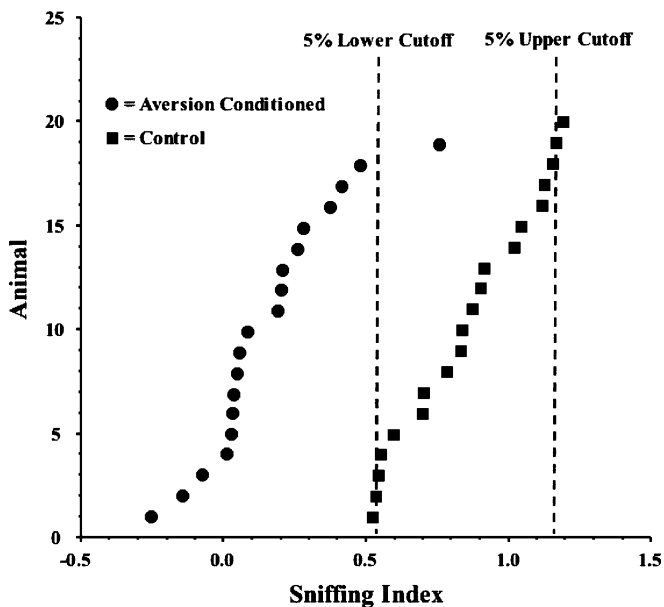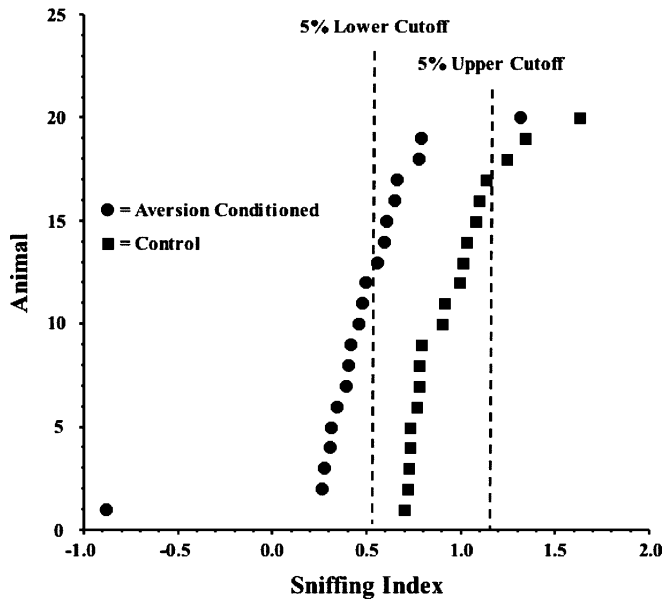
was incorrectly classified. In short, within the context of the population of mice upon which the model was built the approach achieved 95% sensitivity to detect an experimental animal. Thus, it appeared from these data that we could establish a stimulus-induced response model, which could discriminate a group of 'perturbed' animals from controls and had potential utility as a screening procedure.

Phase 2

To test the efficacy of the model developed in Phase 1, we simulated an actual screening procedure. That is, using the existing model for calculating an individual animal's Sniffing Index, we evaluated additional animals that contributed no prior information to the development of the model. To accomplish this, forty additional, seven-week-old mice (20 males and 20 females) were randomly allocated to the control (10 males and 10 females) or odor aversion-conditioned experimental (10 males and 10 females) groups. Male and female experimental mice were odor aversion-conditioned, as described above. Twenty-four h after the second aversion conditioning session each experimental animal and a sex- and age-matched control had their stimulus-induced respiratory response to air and odorant stimuli monitored, as described above, and a single Sniffing Index value calculated for the animal.

Setting the 5% upper and lower cutoff point in Sniffing Index values based on the original 20 control animals in Phase 1, the calculated Sniffing Index value for each new animal (both control and experimental) was evaluated to determine the efficacy of our initial criterion in classifying screened animals. As illustrated in Figure 4, the criterion based strictly on the distribution of the control population from Phase 1 yielded a 35% false negative rate or 65% sensitivity. That is, 7 of 20 aversion-conditioned animals were classified as yielding a Sniffing Index consistent with the original control population. In addition, the criterion yielded a false alarm rate of 15 or 85% specificity. In other words, 3 of the 20 control animals were incorrectly classified as 'different' from the original control population. In total, the initial criterion (i.e. Sniffing Index cutoff values) from Phase 1 correctly identified 75% of the mice. Nonetheless, formal analysis of the two populations of mice again demonstrated a highly significant ability of the response model to discriminate between the two populations of mice ($T = 4.78$, df = 38, $P = 0$).

Although the application of the initial screening criterion demonstrated the general utility of the approach, it was critical for the proposed high-throughput application (or for that matter any random screening approach), that a selected criterion be based on statistical grounds that would allow for adjustment based on consideration (and experimenter tolerance) of the cost of Type I and Type II errors. To accomplish this, the individual Sniffing Index values determined for each animal in Phase 2 (Figure 4) were first transformed into absolute z-score values based on the control population of Phase 1. The z-score transform was a convenient way of

Figure 3 The distributions of Sniffing Index values are illustrated for 39 C57BL/6J mice (odor aversion-conditioned: 10 males and 9 females; control: 10 males and 10 females). The two vertical dashed lines represent the 5% lower and upper cutoff values, respectively, based on the distribution of control animal Sniffing Index values.
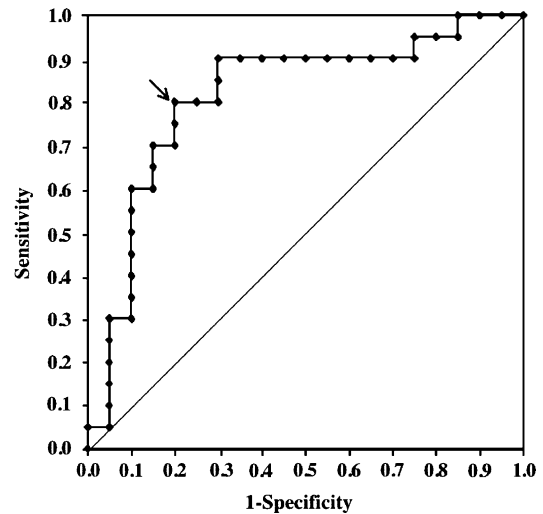
Figure 4 The distributions of Sniffing Index values are illustrated for 40 C57BL/6J mice (odor aversion-conditioned: 10 males and 10 females; control: 10 males and 10 females). The two vertical dashed lines represent the 5% lower and upper cutoff values, respectively, determined from the control distribution in Phase 1.

Figure 5 ROC based on the analysis of z-transformed Sniffing Index values in Phase 2 (see text for details). Sensitivity is the probability that a 'positive' case (i.e. in this phase an odor aversion-conditioned mouse) is correctly classified and is plotted on the y-axis of the ROC (note: 1 – sensitivity is the false negative rate). 1 – specificity is the false positive rate and is plotted on the x-axis. The arrow indicates coordinates of 20% false alarm rate (i.e. 80% specificity) and 80% sensitivity. The identity line represents the null-hypothesis that the true area under the ROC = 0.5 (i.e. chance line).

referencing our sampling distribution of 'unknown' animals to the parent or control distribution of Phase 1. Using standard decision theory of categorization principles (i.e. signal detection theory) (Egan, 1975), these data, in turn, were applied to an ROC (Receiver Operator Curve) analysis (SPSS version 10.1). This approach served two purposes. First, it defined the relationship between a particular z-score criterion value (data not shown) and its associated sensitivity and false positive rate (i.e. 1 – specificity). In other words, it shows the tradeoff between sensitivity and specificity. For example, as illustrated in Figure 5 (see arrow), a criterion value was established such that a false positive rate of 20% (i.e. 80% specificity) could be achieved in association with 80% sensitivity (i.e. 20% false negative). Second, the procedure was a useful way to evaluate the performance of our classification scheme (i.e. the use of one variable by which mice were classified into one of two categories). The area under the ROC in Figure 5 was instructive in this regard in that it equaled the probability of a correct decision. That is, the closer the curve comes to the y-axis and top border of the ROC space the more accurate the test. By contrast, the closer the curve comes to the identity line the less accurate the test. For the current set of test animals in Phase 2 the analysis yielded a highly significant ROC (area or d' = 0.813; P = 0.001).
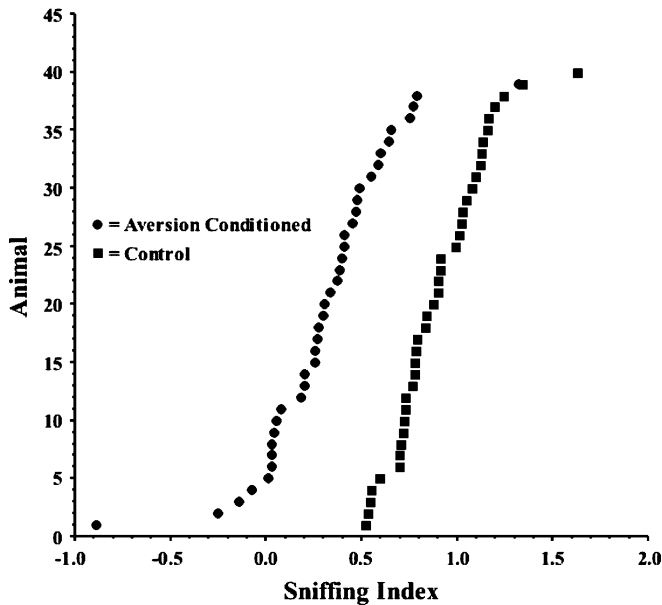
Phase 3

As noted above, the screening criterion based on the ROC approach was designed to permit investigator adjustments

based on the tolerance for Type I and Type II errors. Further, as with any screening procedure, the evaluation criterion was designed to be updated and refined based on experience. Therefore, we combined the data sets from Phases 1 and 2 in order to refine the approach based on an expanded control population (i.e. 40 mice: 20 males and 20 females). Not surprisingly, as illustrated in Figure 6, there was still a clear difference in the reflexive response to the presentation of odorant stimuli between the control and aversion-conditioned populations. Indeed, based on the expanded data set, formal analysis of Sniffing Index values demonstrated highly significant differences between the two groups [$F(1,77)$ = 74.87, P = 0].

To test whether our approach improved with experience (i.e. expansion of the reference population) we evaluated a third set of animals. Forty additional, seven-week-old mice (20 males and 20 females) were again randomly allocated to the control and experimental groups. Twenty-four h after the second aversion-conditioning session of the experimental group each animal (both experimental and control) was tested, as above, and a single Sniffing Index value calculated. The Sniffing Index value of each tested animal was transformed into an absolute z-score value based on the combined control populations of Phases 1 and 2 (Figure 6). The z-score values for the new set of animals were then applied to an ROC analysis. As illustrated by the ROC in Figure 7, expanding the control population improved the detectability of the approach such that the area under the ROC, or d', increased from 0.813 to 0.945. Moreover, it should be
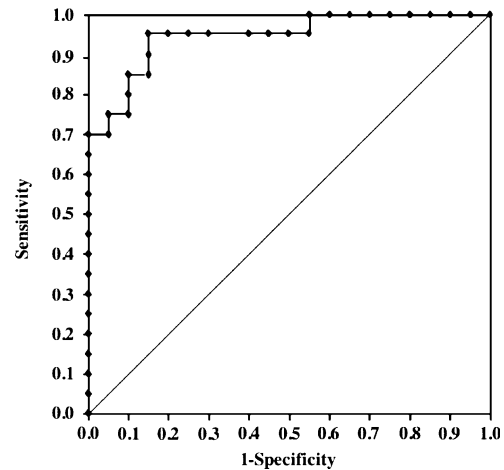
Figure 6   Combined distributions of Sniffing Index values for the odor aversion-conditioned and control animals of Phases 1 and 2 (odor aversion-conditioned: 20 males and 19 females; control: 20 males and 20 females).

Figure 7   ROC based on the analysis of z-transformed Sniffing Index values in Phase 3 (see text for details). The identity line represents the null-hypothesis that the true area under the ROC = 0.5 (i.e. chance line).

emphasized that the new ROC analysis established new criterion z-score values by which future mice could be evaluated with improved levels of specificity and sensitivity. For example, examination of the curve in Figure 7 illustrates that levels of 10 and 15% false alarm rates (i.e. 90 and 85% specificity) were associated with z-score criteria providing 85 and 95% sensitivity, respectively. Thus, these data emphasize both the efficacy of the general approach and the proposition that the application of the method as a screening technique can improve with experience.

Phase 4

The foregoing phases made use of a manipulation that was intended, a priori, to yield a robust change in breathing response to odorant stimulation. By contrast, genetic manipulations, whether targeted or random, will likely vary in the magnitude and mode (i.e. absolute sensitivity or quality perception) of effect on olfactory function. To test directly the efficacy of our stimulus-induced sniffing procedure to distinguish a mutant phenotype from its background strain we made use of the OMP-null mouse (Buiakova *et al.*, 1996). Although OMP-null mutants appear ostensibly 'normal' with respect to the acquisition of a number of complex assessments [i.e. an air versus odor discrimination task (Youngentob and Margolis, 1999), a five-odorant identification task (Youngentob *et al.*, 2001a)] and natural behaviors (Buiakova *et al.*, 1996), these animals represent a model in which a specific gene deletion results in an alteration in physiological (Buiakova *et al.*, 1996; Ivic *et al.* 2000; Youngentob *et al.*, 2003) and behavioral (Youngentob and Margolis,

1999; Youngentob *et al.*, 2001a, 2004) function without resulting in anosmia. With specific regard to behavioral function, rigorous psychophysical procedures were required previously to identify specific sensitivity and quality perception defects in a seemingly normal appearing animal. Consequently, to be effective our procedure should be capable of identifying an altered phenotype such as that observed in the OMP-null mouse.

Twenty seven-week-old OMP-null mice on the 129/Sv+tyr+p+stl J background (10 males and 10 females) and 20 seven-week-old 129/Sv+tyr+p+stl J controls (10 males and 10 females) participated in the study. Following the testing procedures outlined above, each OMP-null mouse and a sex- and age-matched control animal had their stimulus-induced reflexive respiratory response to air and odor stimuli monitored. It should be emphasized that no aversion conditioning took place in this phase.

Following the basic analytic model building procedures outlined in Phase 1, we established a Sniffing Index based only upon the response of the reference population, namely, the 129/Sv+tyr+p+stl J background animals. Briefly, we first defined the linear combination of respiratory parameters that discriminated between the reflexive sniffing response to air and odorant trials in the control group of mice. This analysis determined that two principal components met the predictive error criterion for variable selection (factor 2 – component $T = -2.13$, $P = 0.03$; factor 3 – component $T = -3.20$, $P = 0.001$) in the regression analysis. Furthermore, the analysis again established the set of 14 coefficients for each appropriate factor which when applied in a linear combination of the 14 respiratory measures of stimulus-induced sniffing, established a single numerical value for that factor for each individual trial. These data, in turn, were used to develop a single value defining the Sniffing Index for each individual animal. To accomplish this, we used a stepwise

multiple regression analysis to evaluate the following derived variables: mean numerical value calculated for all trials in a testing session for each factor; the difference between the mean value calculated for the air and odor trials for each factor; and a block trend indicator for each factor for inclusion in the model. The results of this analysis demonstrated significant predictive effects of the difference between the mean value calculated for the air and odor trials for factors 2 and 3 to warrant their inclusion in the final Sniffing Index model. Consequently, the final model incorporated, the two forgoing variables, their appropriately corresponding regression coefficients and a constant.

As illustrated in Figure 8, the procedure segregated the OMP-null and background control animals. Indeed, the two groups of mice were, on average, significantly different from each other in terms of their stimulus-induced reflexive response to odorant stimuli [mean ± SEM: OMP-null = 0.365 ± 0.043, control = 0.634 ± 0.057, $T(38) = 3.74$, $P = 0$]. As a result, these data demonstrated that our monitoring approach could discriminate a group of genetically altered animals from their background strain. Further, by comparison to the distributions for the aversion-conditioned and control animals illustrated in Figure 6, these data also emphasize the ability of our testing procedure to detect a relatively subtle shift in reflexive odor responsivity. That is, there was a greater degree of overlap in the populations in Figure 8 relative to that seen in Figures 3 and 6.

With regard to the above, the degree of overlap in the OMP-null and control populations also highlights that the efficacy of the approach for evaluating individual animals of unknown status is very much dependent on the interrela-

Figure 8 Distributions of Sniffing Index values for 20, OMP-null (10 males and 10 females) and 20, 129/Sv+tyr+p+stl J background strain (10 males and 10 females) mice.

tionship between the robustness of the phenotype, the size of the reference population and the tolerance for Type I and II errors.

Discussion

Normal olfactory sampling involves the drawing of air through the nasal cavity and, given quiet breathing as a reference level, animals have been shown to change a number of sniff parameters in response to a change in the odorous environment. This change can occur either reflexively (i.e. a response to a perceived change in the olfactory ambience induces a change in respiration) (Teichner, 1966; Alberts and May, 1980a,b) or actively (i.e. purposeful active sampling while exploring the odorant environment) (Welker, 1964; Marshall and Moulton, 1977; Youngentob *et al.*, 1987). As such, the measurement of this form of behavior provided an excellent opportunity for a high-throughput quick screen directed towards monitoring the attentiveness and, potentially, the responsivity to odorant stimuli. Indeed, the notion that some parameters of stimulus sampling (i.e. sniffing) can be used as an indirect measure of odorant responsivity was not new. Alberts and May (1980a,b) used reflexive odor-induced polypnea to examine the development of odorant sensitivity in neonatal rats. Further, based on the detailed observations of Youngentob *et al.* (1987), both Doty *et al.* (1988) and Slotnick (1990) have used the gross measure of sampling duration during a psychophysical procedure (as measured by the interruption of a photo cell) as a relatively sensitive measure of an animal's ability to detect or discriminate odors. Consequently, the present set of studies took advantage of the reflexive aspect of this behavior and developed a precision screening technique that utilized the monitoring of stimulus-induced sniffing in response to odor as a method for the automated quantification of 'relative' olfactory function. In so doing, it extends upon the previous body of work and further confirmed the utility of monitoring this form of odor-guided behavior as an index of sensory function.

Using conditioned odor-aversion as a means to create a robust alteration in reflexive sniffing, Phase 1 unambiguously demonstrated we could establish a linear combination of respiratory parameters that significantly discriminated between control and experimental animals (Figure 3). Moreover, based on the populations of mice upon which the model was built, the approach achieved a high degree of separation between control and experimental animals. That is, there was very little overlap in the two populations. Consequently, using a 5% lower and upper cutoff value in the distribution of control animal Sniffing Index values (or 90% specificity), the approach achieved 95% sensitivity to detect the experimental mice.

However, one objective of the approach was to develop a rapid screening procedure for identifying potentially interesting chemosensory phenotypes in individual animals of unknown genetic/phenotypic status. Thus, the goals of Phase 2 were directed toward: (i) examining whether, and to what
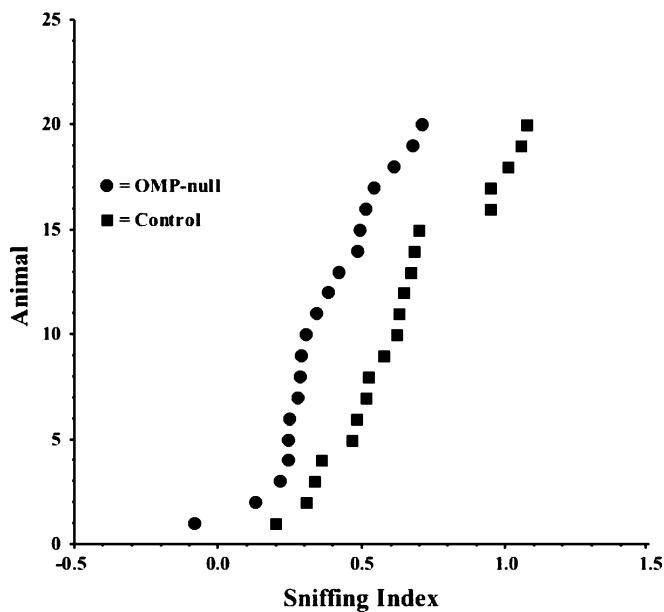
degree, the proposed screening procedure and Sniffing Index model developed in Phase 1 could differentiate between an independent set of control and experimental animals not contributing to the development of the model; and (ii) evaluating the sensitivity and specificity of the screening procedure and evaluation criteria to correctly identify individual animals. As qualitatively illustrated in Figure 4, and confirmed by formal analysis, the approach had clear utility in terms of providing an easily applied screen for the evaluation of defined populations of animals. That is, when the data were evaluated on the basis of control and experimental group means, there was a distinct separation in the distributions of Sniffing Index values that was highly significant. Likewise, when the data sets for Phases 1 and 2 were combined (Figure 6), the group differences were also highly statistically significant. Thus, the first goal of Phase 2 was fulfilled.

The effectiveness of our approach to distinguish between two populations of animals was further verified in Phase 4. As illustrated in Figure 8, OMP-null animals were, on average, significantly different than their background controls. Recall that previous studies have demonstrated mice lacking the gene for OMP have a number of physiologic (Buiakova *et al.*, 1996; Ivic *et al.*, 2000; Youngentob *et al.*, 2003) and behavioral (Youngentob and Margolis, 1999; Youngentob *et al.*, 2001a, 2004) defects. Consequently, our procedure was capable of detecting the impact of a robust experimental manipulation (Figures 3 and 6) as well as a relatively subtler shift in reflexive odor responsivity that resulted from genetic manipulation. Nonetheless, as highlighted below, the degree of overlap in the OMP-null and background distributions had important implications in terms of the procedures utility as an individual animal screening method.

By contrast to the above results, evaluation of the independent set of individual animals allocated to Phase 2 (both control and odor aversion-conditioned) based strictly on the previously established screening criterion developed in Phase 1 (i.e. 5% upper and lower cutoff values in the reference population of Phase 1) achieved less than the desired 90% specificity. That is, when we applied our initial criterion to the new data set the procedure correctly identified 75% of the mice with a sensitivity and specificity of 65 and 85%, respectively. With regard to this later result, our initial screening criterion (based upon the background animal data of Phase 1) was chosen to produce an estimated 10% false positive rate without regard to its impact on sensitivity. This initial approach was applied in order to estimate the true sensitivity to detect our experimental animals. Ideally, we would have liked there to be no overlap in the populations (i.e. 100% sensitivity). However, the results of Phase 2 (Figures 4 and 6) demonstrated a degree of overlap in the control and aversion-conditioned animals that yielded less then the ideal. Consequently, it was clear from these data that a selected criterion be empirically determined, modifiable with experience and based on statistical grounds that would allow for adjustment based on consideration (and tolerance) of the cost of Type I and Type II errors (i.e. any alteration in screening criterion must be weighed against the cost of including too many animals for further analysis or missing animals that may be genetically altered).

To accomplish the above, we applied a standard ROC analysis to the absolute *z*-score transformed Sniffing Index data of Phases 2 and 3 (based on the appropriate control/reference population, respectively). These analyses, in turn, generated functions (Figures 5 and 7) that illustrated the relationship between the sensitivity to detect an experimental animal in each data set and the false-alarm rate. More importantly, the results demonstrated that the relationship between hit-rate and false-alarm rate depended on the cutoff value (i.e. absolute *z*-score value) and that the cutoff value could be determined based on the decision goals of the investigator. Thus, for example (see Figures 5 and 7), depending on the selected cutoff value, higher levels of sensitivity could be achieved at the expense of a higher false alarm probability and vice versa. Finally, the results illustrated in Figures 5 and 7 highlight the intention that the screening procedure can improve with experience. That is, increasing the reference population upon which the absolute *z*-score transform was based improved the area under the ROC or d', where $d' = 0$ represented the chance or unity line.

In considering the above, it must be emphasized that the discriminability between two events (i.e. true-negative and true-positive) is ultimately determined solely by the degree of overlap in the two conditional distributions (note: in a screening setting, one is known, i.e. the reference population, and the other is not, i.e. the randomly screened animal). Thus, as a practical matter, the success rate of identifying an animal as 'interesting' must be considered against the success rate in identifying animals which can be further shown, indeed, to have a deficit. In order to prioritize animals for further evaluation (in consideration of the cost of making Type I errors), we propose a logistic regression model to establish a predictive probability statement that an individual animal is indeed a mutant phenotype (i.e. genetically altered). In other words, even if an animal's observed *z*-score-transformed Sniffing Index satisfies a defined criterion, the more the observed value deviates from the population mean of the reference animals, the greater will be the predictive value of the observation (i.e. the higher the probability that the animal is genetically altered). The logistic regression model for this evaluation is as follows:

$$\ln(P_r/[1 - P_r]) = \ln(P_p/[1 - P_p]) - \ln(n_+/n_-) + \beta_0 + \beta_1 X$$

where, X is the Sniffing Index value (standardized to the mean and SD of the background log-scale distribution) for an individual animal, β_0 is a constant, β_1 is the coefficient of X, n_+ is the number of mutant animals in the reference sample from which β_1 is estimated, n_- is the number of congenic controls in the reference sample from which β_1 is estimated, P_p is the prior probability that a manipulated mouse

is a mutant (i.e. the success rate of, for example, a mutagenesis process to create an olfactory mutant), $1 - P_p$ is the prior probability that a manipulated animal is not a mutant (i.e. failure rate of, for example, a mutagenesis process to create an olfactory mutant) and P_r is the predictive probability that, given an observed Sniffing Index value, an individual mouse has a genetic defect.

The value of the above approach lies not only in its ability to establish a probability statement for an individual animal based upon an observed Sniffing Index value, but, more importantly, in its ability to be refined with experience. In terms of an actual high-throughput screening procedure, there will initially exist little, if any, information about the true success rate in creating identifiable olfactory mutants. However, given a postulated success rate, the logistic regression model will provide a predictive probability for each unknown animal screened. As feedback information becomes available, then updated prior probabilities (P_p), based upon the observed success rate (i.e. rate of identification of genetic mutants among positively screened animals), will permit refinement of the entire screening evaluation. Consequently, our screening criterion is designed to be updated and refined based upon experience.

In summary, it is evident from the forgoing discussion that the strength of our approach varied with its application. That is, whether it was being applied to: (i) the evaluation of two defined populations of animals; or (ii) a random genetic/phenotypic screening procedure. Regarding the former application, given an appropriate sample size in each population, our method of monitoring stimulus-induced sniffing provided a highly reliable approach for the rapid evaluation of a particular experimental/genetic manipulation's impact on sensory function. By contrast, the procedures specificity and sensitivity as a screening method was very much a function of the robustness of the observed phenotype and the tolerance for making observational errors. This observation was not unexpected given the prospect that, a priori, it would be difficult to predict the impact of a given random mutation on olfactory system function. Nonetheless, employing epidemiologic screening procedures that are: modifiable with experience; experimenter adjustable based on detection tolerance; and permit the prioritization of animals for further evaluation, the utility of the method as a screening procedure has been maximized.

Acknowledgement

References

Alberts, J.R. and **May, B.** (1980a) *Ontogeny of olfaction: development of the rats' sensitivity to urine and amyl acetate*. Physiol. Behav., 24, 965–970.

Alberts, J.R. and **May, B.** (1980b) *Development of nasal respiration and sniffing in the rat*. Physiol. Behav., 24, 957–963.

Apfelbach, R., Weiler, E., Asselbergs, W.F., Polak, E.H. and **Slotnick, B.M.** (1998) *Selective and reversible reduction of odor sensitivity in the rat by concanavalin A*. Physiol. Behav., 65, 513–516.

Brunet, L.J., Gold, G.H. and **Ngai, J.** (1996) *General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel*. Neuron, 17, 681–693.

Buiakova, O.I., Baker, H., Scott, J.W., Farbman, A.I., Kream, R., Grillo, M., Franzen, L., Richman, M., Davis, L.M., Abbondanzo, S., Stewart, C.L. and **Margolis, F.L.** (1996) *Olfactory marker protein (OMP) gene deletion causes altered physiological activity of olfactory sensory neurons*. Proc. Natl Acad. Sci. USA, 93, 9858–9863.

Bulfone, A., Wang, F., Hevner, R., Anderson, S., Cutforth, T., Chen, S., Meneses, J., Pedersen, R., Axel, R. and **Rubenstein, J.L.** (1998) *An olfactory sensory map develops in the absence of normal projection neurons or GABAergic interneurons*. Neuron, 21, 1273–1282.

Cau, E., Gradwohl, G., Fode, C. and **Guillemot, F.** (1997) *Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors*. Development, 124, 1611–1621.

Cutforth, T., Moring, L., Mendelsohn, M., Nemes, A., Shah, N.M., Kim, M.M., Frisen, J. and **Axel, R.** (2003) *Axonal ephrin-as and odorant receptors: coordinate determination of the olfactory sensory map*. Cell, 114, 311–322.

Doty, R.L. and **Risser, J.M.** (1989) *Influence of the D-2 dopamine receptor agonist quinpirole on the odor detection performance of rats before and after spiperone administration*. Psychopharmacology (Berl), 98, 310–315.

Doty, R.L., Ferguson-Segall, M., Lucki, I. and **Kreider, M.** (1988) *Effects of intrabulbar injections of 6-hydroxydopamine on ethyl acetate odor detection in castrate and non-castrate male rats*. Brain Res., 444, 95–103.

Egan, J.P. (1975) Signal Detection Theory and Analysis. Academic Press, New York.

Greiner, R.S., Moriguchi, T., Hutton, A., Slotnick, B.M. and **Salem, N., Jr** (1999) *Rats with low levels of brain docosahexaenoic acid show impaired performance in olfactory-based and spatial learning tasks*. Lipids, 34, S239–43.

Greiner, R.S., Moriguchi, T., Slotnick, B.M., Hutton, A. and **Salem, N.** (2001) *Olfactory discrimination deficits in n-3 fatty acid-deficient rats*. Physiol. Behav., 72, 379–385.

Ivic, L., Pyrski, M.M., Margolis, J.W., Richards, L.J., Firestein, S. and **Margolis, F.L.** (2000) *Adenoviral vector-mediated rescue of the OMP-null phenotype* in vivo. Nat. Neurosci., 3, 1113–1120.

Lu, X.-M. and **Slotnick, B.M.** (1998) *Olfaction in rats with extensive lesions of the olfactory bulbs: implications for odor coding*. Neuroscience, 84, 849–866.

Luo, A.H., Cannon, E.H., Wekesa, K.S., Lyman, R.F., Vandenbergh, J.G. and **Anholt, R.R.** (2002) *Impaired olfactory behavior in mice deficient in the alpha subunit of G(o)*. Brain Res., 941, 62–71.

Marshall, D. and **Moulton, D.** (1977) *Quantification of nasal air flow patterns in dogs performing an odor detection task*. In LeMangen, J. and MacLeod, P. (ed.), Olfaction and Taste VI. IRL Press, Oxford, pp. 170.

Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J. and **Axel, R.** (1996) *Visualizing an olfactory sensory map*. Cell, 87, 675–686.

Morrison, D.F. (1967) Multivariate Statistical Methods. McGraw-Hill Book Co., New York.

Mozell, M.M., Sheehe, P.R., Swieck, S.W.,Jr, Kurtz, D.B. and **Hornung, D.E.** (1984) *A parametric study of the stimulation variables affecting*

the magnitude of the olfactory nerve response. J. Gen. Physiol., 83, 233–267.

O'Brien, TP. and **Frankel, WN** (2004) *Moving forward with chemical mutagenesis in the mouse.* J. Physiol. (Lond.), 554, 13–21.

Pourtier, L. and **Sicard, G.** (1990) *Comparison of the sensitivity of C57BL/6J and AKR/J mice to airborne molecules of isovaleric acid and amyl acetate.* Behav. Genet., 20, 499–509.

Slotnick, B. (1990) *Olfactory perception.* In Berkeley, M.A. and Stebbins, W.C. (ed.), Comparative Perception. Vol. 1. Basic Mechanisms. Wiley & Sons, New York, pp. 155–214.

Teichner, W.H. (1966) *A method for studying olfaction in the unrestrained rat.* J. Psychol., 63, 291–297.

Welker, W.I. (1964) *Analysis of sniffing of the albino rat.* Behaviour, 22, 223–244.

Wong, S.T., Trinh, K., Hacker, B., Chan, G.C., Lowe, G., Gaggar, A., Xia, Z., Gold, G.H. and **Storm, D.R.** (2000) *Disruption of the type III adenylyl cyclase gene leads to peripheral and behavioral anosmia in transgenic mice.* Neuron, 27, 487–497.

Wysocki, C.J., Whitney, G. and **Tucker, D.** (1977) *Specific anosmia in the laboratory mouse.* Behav. Genet., 7, 171–188.

Yee, K. and **Wysocki, C.** (2002) *Differential responses to odorant analogs after recovery from nerve transection.* Physiol. Behav., 76, 661–667.

Youngentob, S.L. and **Margolis, F.L.** (1999) *OMP gene deletion causes an elevation in behavioral threshold sensitivity.* Neuroreport, 10, 15–19.

Youngentob, S.L., Mozell, M.M., Sheehe, P.R. and **Hornung, D.E.** (1987) *A quantitative analysis of sniffing strategies in rats performing odor detection tasks.* Physiol. Behav., 41, 59–69.

Youngentob, S.L., Schwob, J.E., Sheehe, P.R. and **Youngentob, L.M.** (1997) *Odorant threshold following methyl bromide-induced lesions of the olfactory epithelium.* Physiol. Behav., 62, 1241–1252.

Youngentob, S.L., Margolis, F.L. and **Youngentob, L.M.** (2001a) *OMP gene deletion results in an alteration in odorant quality perception.* Behav. Neurosci., 115, 626–631.

Youngentob, S.L., Schwob, J.E., Saha, S., Manglapus, G. and **Jubelt, B.** (2001b) *Functional consequences following infection of the olfactory system by intranasal infusion of the olfactory bulb line variant (OBLV) of mouse hepatitis strain JHM.* Chem. Senses, 26, 953–963.

Youngentob, S.L., Kent, P.F. and **Margolis, F.L.** (2003) *OMP gene deletion results in an alteration in odorant-induced mucosal activity patterns.* J. Neurophysiol., 90, 3864–3873.

Youngentob, S.L., Pyrski, M.M., Margolis, F.L. (2004) *Adenoviral vector-mediated rescue of the OMP-null behavioral phenotype: Enhancement of odorant threshold sensitivity.* Behav. Neurosci., 118, 636–642.

Accepted December 30, 2004