HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

Nat Neurosci. 2015 October ; 18(10): 1446–1454. doi:10.1038/nn.4104.

Molecular profiling of activated olfactory neurons identifies odorant receptors for odors *in vivo*

Yue Jiang[1,2,3], **Naihua Natalie Gong**[1], **Xiaoyang Serene Hu**[1], **Mengjue Jessica Ni**[1], **Radhika Pasi**[1], and **Hiroaki Matsunami**[1,4]

[1]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

[2]University Program in Genetics and Genomics, Duke University Medical Center, Durham, NC 27710, USA

[3]Department of Statistical Science, Duke University, Durham, NC 27708, USA

[4]Department of Neurobiology, Duke Institute for Brain Sciences, Duke University Medical Center, Durham, NC 27710, USA

Abstract

The mammalian olfactory system uses a large family of odorant receptors to detect and discriminate amongst a myriad of volatile odor molecules. Understanding odor coding requires comprehensive mapping between odorant receptors and corresponding odors. Here we present high–throughput *in vivo* identification of odorant receptor repertoires responding to odorants, using phosphorylated ribosome immunoprecipitation of mRNA from olfactory epithelium of odor–stimulated mice followed by RNA–Seq. This approach screens the endogenously expressed odorant receptors against an odor in one set of experiments, using awake and freely behaving mice. In combination with validations in a heterologous system, we identify sets of odorant receptors for two odorants, acetophenone and 2,5–dihydro–2,4,5–trimethylthiazoline (TMT), encompassing 69 odorant receptor–odorant pairs. We also identified shared amino acid residues specific to the acetophenone or TMT receptors, and developed models to predict receptor activation by acetophenone. This study provides a means to understand the combinatorial coding of odors *in vivo*.

Introduction

In mammals, olfactory sensation starts with the detection of odor ligands predominantly by odorant receptors (ORs), a large family of seven transmembrane G protein–coupled receptors (GPCRs) [1, 2]. ORs are individually expressed in olfactory sensory neurons (OSNs)

Correspondence to: Hiroaki Matsunami.

Author contributions

H.M. supervised all experiments and data analysis. N.N.G., Y.J., X.S.H. and H.M. performed immunohistochemistry experiments. Y.J. and H.M. performed pS6–IP. Y.J. performed bioinformatics and statistical analysis. M.J.N. cloned some of the OR constructs. Y.J. and R.P. performed *in vitro* luciferase assays. Y.J., N.N.G. and H.M. wrote the paper.

located in the olfactory epithelium (OE). The mouse genome encodes over 1000 intact ORs [2]. Odor recognition follows a combinatorial coding scheme, where one OR can be activated by a set of odorants, and one odorant can activate a combination of ORs [3, 4]. Through such combinatorial coding, mammals can detect and discriminate a large number of olfactory stimuli.

Deciphering the coding of olfactory information requires the comprehensive identification of ORs that respond to a given odorant [5]. Various *in vivo*, *ex vivo*, and *in vitro* methods, such as virus–mediated OR overexpression [6, 7], calcium imaging of dissociated OSNs combined with single–cell RT–PCR [3, 7], the use of fluorescently–labeled transgenic mice [8], and *in vitro* OR expression [9] have been used to match ORs with their cognate ligands. Notably, molecular receptive range analyses of a few ORs have revealed diverse odor tuning properties among the tested ORs [10, 11].

Currently, large–scale identification of active ligands for mammalian ORs relies on *in vitro* heterologous cell systems [4, 9, 12]. Though many studies, including those of M71, M72, I7, OR–EG, MOR23, and SR1 in mice, and OR7D4, OR11H7, OR5A1, OR2J3, and OR10G4 in humans, have shown that *in vitro* responses predict OSN activation in mice and odor perception in humans [12–16], the lack of high–throughput *in vivo* mapping methods makes it difficult to estimate the correspondence between *in vitro* and *in vivo* results across a large number of ORs activated by a given odor.

The S6 ribosomal subunit is phosphorylated following neuronal activation [17]. This phosphorylation is comparable to induction of immediate early gene expression, such as *c–Fos* and *Egr–1*, which is widely used to mark active neurons [15, 18, 19]. However, unlike immediate early gene expression, phosphorylated S6 has the advantage of being physically associated with mRNA species expressed in the activated neurons. Thus, phospho–S6 immunoprecipitation (pS6–IP) followed by mRNA profiling with RNA–Seq has been developed as a method to identify mRNAs expressed in activated neurons [17]. This approach has been successfully applied in brain regions such as the hypothalamus in identifying markers of neurons that respond to feeding, starvation and high salt [17].

Here we demonstrate that S6 phosphorylation occurs in the mouse OE following odor stimulation and is a reliable marker for OSN activation. pS6–IP reveals enrichment in subsets of OR mRNAs in mice stimulated with odorants. Combining the use of high–throughput *in vivo* mapping followed by *in vitro* validation, we identify diverse sets of ORs responding to acetophenone and TMT.

Results

Odor exposure leads to S6 phosphorylation in the OE

In the OE, each OSN chooses to express one OR allele out of over 1000 possible OR genes [3, 20]. Thus, we reasoned that pS6–IP could be applied in the olfactory system to map odor–activated ORs, as ORs associated with activated OSNs are likely responding to the odor. However, it is unknown whether ribosome phosphorylation occurs in the OSNs activated by odor exposure, and if so, whether pS6–associated ORs are responding to the

tested odor. To determine whether S6 phosphorylation occurs when ORs are activated in the OSNs, we first tested whether odor stimulation leads to S6 phosphorylation in the OE. We presented each of the tested mice with a stimulation cassette enclosing a piece of filter paper spotted with 10 μL undiluted (100%) or 1% acetophenone (odor), or distilled water (control) in a clean disposable cage. One hour later, the animals were sacrificed and coronal sections from the OE were stained with anti–pS6 antibody (Fig. 1a). While the background pS6 signal was low in the control OE, a subpopulation of OSNs in the stimulated OE displayed strong staining for pS6 (fraction of OSNs showing positive pS6 staining following stimulation with 100% acetophenone: 16% ± 3%, n = 3 images; 1% acetophenone: 6% ± 2%, n = 3 images). Double staining with an antibody against a known acetophenone receptor, Olfr160, also known as M72 (the related acetophenone receptor M71 is a pseudogene in C57BL/6 strain), revealed colocalization of Olfr160 and pS6 signals, suggesting that M72–expressing OSNs show S6 phosphorylation in response to acetophenone stimulation. To determine the specificity of pS6 induction in response to odor exposure, we stimulated the mice with other known Olfr160 agonists [21] along with control odors (Supplementary Fig. S1a), and quantified the proportion of Olfr160–expressing OSNs that show pS6 staining. As expected, odor stimulation with Olfr160 agonists (methyl salicylate, methyl benzoate, acetophenone) led to S6 phosphorylation in a large proportion of Olfr160–expressing OSNs, while in animals stimulated by control odors (heptanoic acid, 2,5–dihydro–2,4,5–trimethylthiazoline (TMT), (+)–carvone), pS6 signals showed little overlap with Olfr160 signals (Fig. 1b and Supplementary Fig. S1b). To further evaluate the specificity and sensitivity of the pS6 method, we quantified pS6 induction using staining intensity for five known OR–odorant pairs, including Olfr690 (MOR31–2) – isovaleric acid [22], Olfr961 (MOR224–5) – eugenol [22], Olfr2 (I7) – heptanal [6, 8, 23] and Olfr1440 (MOR215–1) – muscone [15, 24] in addition to Olfr160 (M72) – acetophenone and control pairs of ORs with non–activating odorants. Three ORs were dorsally expressed (Olfr160, Olfr690, and Olfr961) and two were ventrally expressed (Olfr2 and Olfr1440). We stimulated individual mice with one of the odorants (10 μL, undiluted) and performed in situ hybridization with each of the five OR probes followed by pS6 immunostaining (Fig. 1c,d and Supplementary Fig. S2a). When we quantified pS6 signal intensities in individual OSNs labeled with each of the OR probes, significant pS6 induction was observed with previously reported OR–odorant pairs compared to control OR–odorant combinations ($p < 0.0001$, one–way ANOVA and Dunnett's post–hoc test). When we tested the ORs with two different concentrations (100% and 1%) of cognate odorants, significant pS6 induction was observed ($p < 0.001$, one–way ANOVA and Dunnett's post–hoc test), with the exception of Olfr961 with 1% eugenol (Fig. 1e, Supplementary Fig. S3 and Supplementary table 1). Time–course analysis of pS6 signals after acetophenone exposure indicated that pS6 signals started by 30 min and reached a plateau at one hour (Supplementary Fig. S2b). These results suggest that pS6 is a reliable marker for OSN activation following odor stimulation.

pS6–IP enriches OR mRNAs following odor stimulation

To more comprehensively identify the ORs expressed in the pS6–positive OSNs, we performed pS6–IP followed by RNA–Seq (Fig. 2a). Three pairs of mice, each pair constituted of littermates of the same gender and age, were used for each pS6–IP RNA–Seq experiment. Within each pair, mice were exposed to either acetophenone or no odor control.

One hour later, we dissected the OE, homogenized the tissue, immunoprecipitated pS6 along with associated RNAs, and analyzed the purified RNA. Comparison of the samples from mice stimulated with 10 μL of acetophenone versus control revealed higher level of the immediate early genes *c–Fos* and *Egr1* in the stimulated samples, suggesting pS6–IP enriches mRNA species in activated OSNs (Fig. 2b). We measured expression levels for ORs by counting reads that mapped within coding exons of OR transcripts. Although the vast majority of genes were expressed at comparable levels in the two groups, a subset of ORs was enriched in the stimulated group. Differential expression analysis identified a large set of ORs, including Olfr160 (M72), that were significantly enriched following stimulation with 10 μL of undiluted (100%) acetophenone (47 and 75 ORs with $p < 0.001$ and $p < 0.05$, respectively, FDR adjusted, Fig. 2c, d). Stimulation with 10 μL of 1% acetophenone (Fig. 2e, f) enriched 9 and 25 ORs with $p < 0.001$ and $p < 0.05$, respectively (FDR adjusted). Most of these ORs were also enriched following 100% acetophenone stimulation (9 out of 9 for ORs enriched with $p < 0.001$, 14 out of 25 for ORs enriched with $p < 0.05$, Fig. 2g), suggesting that the enrichment process is reproducible and is consistent with previous findings that high odor concentrations activate additional ORs [3, 25].

Enriched ORs tend to respond to acetophenone *in vitro*

To validate that the *in vivo* receptor mapping identified acetophenone–activated ORs, we leveraged the previously established cAMP–mediated luciferase reporter gene assay in heterologous cells [26, 27] to measure the response of these ORs to acetophenone (Fig. 3 a). To systematically test whether the enriched ORs were more likely to respond to acetophenone, we tested 71 ORs enriched by 100% acetophenone (out of 75 ORs with $p < 0.05$ after FDR correction) and 449 control ORs that were not enriched (Fig. 3b). We independently expressed these ORs in Hana 3A cells, which are a HEK293T–derived cell line that supports the robust expression of various transiently expressed ORs [9, 26], and stimulated the cells with 3μM, 30μM, or 300μM of acetophenone as well as no–odor controls. We quantified the degree of OR activation by determining the relative fold–increase of luciferase activity as compared to control stimulation. As expected, acetophenone–induced *in vitro* activation of enriched ORs was higher than that of control ORs ($p = 0.005$ at 30μM, $p = 7 \times 10^{-21}$ at 300μM, Wilcoxon rank–sum test, Fig. 3c). Similarly, the ORs that were activated by acetophenone *in vitro* were enriched in the pS6–IP analysis as a group ($p = 1 \times 10^{-14}$, Wilcoxon rank–sum test, Fig. 3d. Cut–off for defining positive response to acetophenone *in vitro* is fold of luciferase induction 2.33 fold at 300μM; see methods for the determination of the cut–off, which was set to exclude 99% of negative receptors). Out of the 75 ORs enriched in 100% acetophenone experiment ($p < 0.05$), 71 ORs were tested *in vitro* and 69% (49 ORs) responded. In contrast, out of 1047 ORs not enriched by 100% acetophenone ($p > 0.05$), 449 ORs were tested in vitro and only 13% (58 ORs) responded. 74% (43 ORs) of these 58 ORs showed trends of enrichment *in vivo* but did not reach statistical significance. To quantify how well *in vivo* and *in vitro* results predict each other, we generated receiver operating characteristic (ROC) curves by plotting the true positive rate against the false positive rate as the discrimination cut–off of the predictor is varied. Enrichment in pS6–IP RNA–Seq predicted whether the OR responded to acetophenone *in vitro* (area under the ROC curve (AUC) = 0.754, $p = 6 \times 10^{-16}$, Wilcoxon rank–sum test, one tailed against H_0: classifier performs no better than random, Fig. 3e). *In vitro* response also predicted the p–

value of enrichment in pS6–IP RNA–Seq (AUC = 0.845, $p = 7 \times 10^{-21}$, Wilcoxon rank–sum test, one tailed against H_0: classifier performs no better than random, Fig. 3f).

Identification of acetophenone ORs

After establishing the concordance between *in vivo* and *in vitro* responses, we set out to further investigate the acetophenone–activated ORs. We selected ORs that were enriched by S6 phosphorylation analysis with acetophenone, expressed these ORs in Hana3A cells [9] and challenged the cells with 10^{-9} – 10^{-3} M acetophenone. The dose–dependent luciferase response profiles confirmed that 48 ORs were indeed activated by acetophenone (Fig. 4a, b). Out of the 75 ORs enriched in 100% acetophenone experiment ($p < 0.05$), 71 were tested *in vitro* and 45 (63%) responded. Out of the 25 ORs enriched in 1% acetophenone experiment ($p < 0.05$), 24 were tested *in vitro* and 15 (63%) responded. Out of the 15 confirmed acetophenone ORs identified for 1% acetophenone, 12 (80%) were also enriched by 100% acetophenone, while the other three did not reach statistical significance although they showed trends towards enrichment. The 1% and 100% stimulation experiments thus identify at total of 48 acetophenone receptors (Supplementary table 2). We reason that these 48 receptors, supported by both *in vivo* and *in vitro* evidence, are likely to be *bona fide* acetophenone receptors.

To confirm that the acetophenone ORs we identified do respond to acetophenone *in vivo*, we performed pS6 staining along with fluorescent RNA in situ hybridization using specific probes for six of the newly identified acetophenone ORs and a control OR that shows no *in vitro* response or enrichment in pS6–IP. Indeed, acetophenone exposure resulted in pS6 induction in OSNs expressing identified acetophenone ORs as compared to the control OR (Fig. 4c,d and Supplementary Fig. S3).

While the binary responses of *in vivo* and *in vitro* results correlate well, it is not clear to what extent the pS6 method captures the information of OR sensitivity. To address this question, we plotted the *in vitro* EC50 values against the *in vivo* fold of transcript enrichment for the 48 acetophenone ORs. Indeed, EC50 value negatively correlated with transcript fold change on log scales following 1% acetophenone stimulation ($p = 0.005$, linear regression and ANOVA, Spearman's rho = –0.395, Fig. 5a), suggesting that ORs more sensitive *in vitro* tend to be more enriched *in vivo.* However, this correlation was not observed for the 100% acetophenone stimulation ($p = 0.993$, linear regression and ANOVA, Spearman's rho = 0.00137, Fig. 5b), presumably due to saturation of the olfactory signaling pathways in OSNs following high intensity odor stimulation. In line with this, the set of ORs enriched by 1% acetophenone tended to contain a higher fraction of more sensitive ORs as compared to ORs not enriched or enriched only by 100% acetophenone ($p < 2 \times 10^{-16}$, Chi–square test, Fig. 5c). In addition, the median EC50 value of ORs enriched by 1% acetophenone tended to be lower than that of those only enriched by 100%, although these results were not statistically significant ($p = 0.207$, Wilcoxon rank–sum test, Fig. 5d). These results suggest that while pS6–IP with high odorant concentrations is useful in capturing the maximum number of responding ORs, lower odorant concentrations better preserve information regarding OR sensitivity to the odorant.

Sequence–function relationship of acetophenone ORs

Having a large set of acetophenone ORs, we next asked if these ORs have greater overall homology with each other. Surprisingly, when plotted on the OR phylogenetic tree, the acetophenone ORs were not clustered in one or a few subfamilies, though there were a few notable examples of closely related ORs (Fig. 6a). Rather, their sequences were strikingly diverse: at least one member of 29 out of the 286 receptor families defined by Zhang and Firestein [28] were represented.

Even though their overall sequences are diverse, it is possible that particular domains or sites are shared among acetophenone ORs. We therefore asked whether particular amino acid residues were conserved at a given site among the acetophenone ORs as a group. Comparison of mean Grantham distances [29] of amino acid properties at individual sites among the acetophenone ORs with that of random OR sets identified 44 sites with higher amino acid similarity among acetophenone ORs than random (Fig. 6b, n = 44, $p < 0.05$; n = 17, $p < 0.01$; FDR adjusted and Supplementary Fig. S4. See methods for details). Notably, these sites are not highly conserved among ORs as a whole and the majority of them are located within transmembrane helices implicated in odorant binding.

The conserved sites among acetophenone ORs indicate that primary protein sequences may contain sufficient information to predict whether or not a given OR responds to acetophenone. Indeed, a principle component analysis on the amino acid properties [29] of ORs identified 4 clusters using the first three principle components, and the acetophenone ORs were mainly present in one of them (Fig. 6c). To build models linking OR protein sequences to their responsiveness to acetophenone, we took two different approaches, and tested them in both 10–fold cross–validation and external validation schemes. The data for modeling contained the 48 ORs we identified as acetophenone receptors, and 367 ORs that showed no significant response to acetophenone both *in vivo* and *in vitro*. The first model was built using the support vector machine (SVM) with radial basis kernel as commonly used in the machine learning community [30]. Briefly, the SVM algorithm finds the hyper surface in the parameter space that best separates the ORs responding and not responding to acetophenone. A second model was built using logistic regression with variable selection using elastic net penalty [31]. The variable selection step was introduced to reduce over fitting. As a comparison, we also implemented a classifier using the overall sequence similarity to predict OR responses. All models were tested in a 10–fold cross–validation scheme. In addition, we also obtained the *in vitro* acetophenone response profiles of 27 human ORs and used this separate dataset for external validation of models (Supplementary Fig. S5). These human ORs have been functionally expressed in our heterologous system successfully to exclude confounding factors such as failures of functional expression *in vitro* [12]. Indeed, the SVM and elastic net models showed significant predictive values in both cross and external validation and correctly predicting responsiveness to acetophenone 71.6% to 81.5% of the time. In contrast, the overall similarity based approach did not predict responsiveness to acetophenone with statistical significance (Fig. 6d, e and Table 1).

Identification of TMT ORs

Thus far, we used acetophenone to comprehensively identify OR repertoire, but whether the method can be used for other odorants is not clear. We chose the odor 2,5–dihydro–2,4,5–trimethylthiazoline (TMT), a fox feces component that induces fear–related responses in predator–naïve mice [32, 33]. We next applied our OR mapping protocol to identify ORs responsible for detecting TMT *in vivo*. We stimulated mice with 100% TMT for 1 hour and performed pS6–IP in the dissected OE (n = 4 pairs). TMT stimulation led to enrichment of 43 ORs with $p < 0.05$. 1%TMT enriched 4 ORs with $p < 0.05$, a subset of the 43 100% TMT–enriched ORs (Fig. 7a and Supplementary Fig. S6). We cloned 42 of these ORs, and 21 of them displayed significant responses to TMT when expressed *in vitro* (Fig. 7b). Thus, we have strong evidence that these 21 ORs are indeed TMT receptors (Supplementary table 2). Somewhat unexpectedly, 6 of the TMT ORs were also identified as acetophenone receptors. Similar to acetophenone ORs, the TMT ORs were also diverse in sequence (Fig. 7c). Like the majority of conserved residues among acetophenone ORs, the conserved residues among TMT ORs were also located within transmembrane helices, and may be involved in ligand binding (Fig. 7d). Five of these TMT ORs (Olfr20, Olfr30, Olfr57, Olfr376, Olfr491) were expressed in the dorsal OE, a region implicated in mediating the fear–related behavioral responses of TMT [33] (Supplementary Fig. S6e). These ORs represent candidates for future investigation to understand the basis of TMT–induced behaviors. Together, our data suggests that pS6–IP in combination with RNA–Seq can be used to identify active ORs for diverse odorants.

Discussion

One odorant can activate a subset of ORs, but which ORs are activated *in vivo* is largely unclear. Though the search for OR ligands *in vivo* has been facilitated by transgenic animals that co–express fluorescent proteins with the ORs of interest [8], these animal resources are limited to several well–studied ORs. Here we describe a novel approach aiming to comprehensively identify odor–activated mammalian ORs *in vivo*. Applying this approach to acetophenone and TMT identified 48 and 21 OR–odorant pairs, respectively. The pS6–IP method has a few advantages over the currently available techniques. First, the experiment is performed in awake, freely behaving animals; thus, the results obtained are more likely to be physiologically relevant. Secondly, pS6–IP screens most if not all expressed mouse ORs in one experiment, greatly enhancing the throughput while providing a global view of odor–induced OR activation patterns. Recently, McClintock et al. reported a method called the Kentucky assay to deorphanize ORs using a reporter S100a5 to drive GFP expression in activated OSNs, followed by microarray analysis of sorted GFP^+ cells [24]. Three ORs were identified for eugenol and two ORs for muscone, supported by both *in vivo* and *in vitro* methods. The Kentucky assay may be suitable for identifying ORs that are most active, while the pS6–IP method enables a large–scale identification of ORs for a given odor. As compared to S100a5–tauGFP, pS6 induction shows faster onset (30 minute – 1 hour versus 14 hours of odor stimulation time). In addition, the pS6–IP does not require specific transgenic animals, which may enable researchers using non–model species to identify ORs activated by odorants of interest.

Von der Weid et al. reported a method that takes advantage of the decrease in OR transcript abundance following odor stimulation to identify ORs responding to acetophenone and ethyl isobutyrate [34]. As compared to pS6–IP, this method involves fewer steps and only requires working with whole RNA from OE while pS6 induction occurs much faster than down regulation of OR transcripts. We examined the set of acetophenone ORs (n = 22) identified by von der Weid et al. in our pS6-IP RNA-Seq and heterologous expression data. 20/22 (91%) ORs show positive log fold change in at least one concentration, of which 18 (81%) ORs show $p < 0.05$ (FDR adjusted for 22 ORs) at least in one concentration. In addition, 15/17 (88%) ORs show in vitro activation. Overall, 21/22 (95%) ORs are supported by at least one of the criteria in our dataset (Supplementary table 3). Quantification by generating an ROC curve showed that enrichment in pS6–IP RNA–Seq predicted the ORs identified by von der Weid et al. (area under the ROC curve (AUC) = 0.86, $p=4\times10^{-9}$, Wilcoxon rank–sum test, one tailed against H_0: classifier performs no better than random) (Supplementary Fig. S7). Together, our data supports data presented by von der Weid et al. as groups. The different stimulation conditions, different signaling pathways and kinetics underlying the two methods as well as method–specific statistical cutoffs for defining positive receptors likely contribute to the differences in outcome. Future research is necessary to clarify why certain ORs are only identified in one of the two methods.

In general we found high correlations between *in vivo* and *in vitro* mapping results, strongly suggesting that OR activation leads to phosphorylation of S6. Future study is necessary to identify the signaling pathway leading to S6 phosphorylation. Our quantification of pS6 immunostaining signal intensity suggests that ligand concentrations and ligand affinities for each OR both influence phosphorylation levels of S6. While higher phosphorylation levels are usually associated with higher concentrations of a given ligand, the levels may saturate for high–affinity receptors. We established a stringent criteria requiring both *in vivo* and *in vitro* support to determine ORs activated by a given odorant. Nonetheless, there are some discrepancies between *in vivo* and *in vitro* OR activity worth addressing in the future. One possible explanation for these differences is that odor molecules are converted into other chemicals in the nasal mucus and activate additional ORs [35, 36]. Another possibility is poor functional expression of some ORs *in vitro*. Though unlikely, one also cannot exclude the possibility that some OSNs express multiple ORs at high levels, in which case non–responding ORs may be enriched if the co–expressed OR responded to the odor. Finally, the statistical definitions of activation and enrichment used may not fully capture the biological significance. Increasing the sequencing depth (currently ~10 million reads/sample) might give more reads to better test against the null hypothesis for ORs less frequently chosen and expressed at low levels in the OE. Despite these potential shortcomings, our study represents a powerful strategy to identify active ORs for an odor of interest.

We found good correlation between *in vitro* sensitivities and *in vivo* enrichment at a low stimulation concentration. Differences between the *in vivo* and *in vitro* systems may help explain the noise in this correlation. For instance, the anatomical structure of the nasal passage can influence airflow, which in turn can affect the odorant concentration dissolved in the nasal mucus [37]. Because ORs are expressed in a zonal pattern in the olfactory epithelium, the exact odorant concentration each OR is exposed to can vary. Furthermore,

the existence of odorant binding proteins in the nasal mucus may also change the kinetics of OR–odorant interactions. In contrast, we did not see correlation between *in vitro* sensitivities and *in vivo* enrichment at a high stimulation concentration. It is possible that for many olfactory neurons, the olfactory signaling pathways are saturated by the high concentration of ligand regardless of OR sensitivity.

Higher concentrations of a given odor not only increase the perceived odor intensity, but may affect the perceived odor quality in some cases [38]. Our unbiased mapping of ORs activated at different concentration of odors provides identities of more sensitive ORs as well as additional ORs activated by a higher odor concentration in freely behaving mice (Fig. 2 and Fig. 7). Our approach will lead to future experiments addressing the roles of each of the active ORs in odor detection and perception.

We showed acetophenone and TMT activate a large set of 48 and 21 ORs, respectively. 6 ORs were activated by both odorants, raising the possibility that these ORs are broadly tuned, which can be determined by future experiments with an expanded range of odors. Unlike acetophenone and TMT, some odors such as muscone, however, may only activate a few ORs [15, 24]. Following with the logic of narrowly–tuned and broadly–tuned ORs [4, 10, 11, 13, 39, 40], the number of ORs activated by different odorants may vary dramatically.

Though the overall sequences of acetophenone and TMT ORs are diverse, we identified residues that are similar within the two types of ORs to their respective activating odors, the majority of which were unique to and non–overlapping between the ORs of each tested odor. Most of these conserved residues are located at the transmembrane domains implicated in ligand binding. Furthermore, PCA performed on OR sequences suggests that OR protein sequences contain information about their responsiveness to acetophenone. To further pursue the relationship between OR sequence and responsiveness, we developed models using two independent approaches that both predict whether or not a given OR responds to acetophenone based on its primary sequence information. The shared residues may directly or indirectly contribute to ligand selectivity of the ORs. One possibility is that ORs activated by the same odor form a similar binding pocket even though the overall sequences are different. On the other hand, multiple unrelated binding pockets could accommodate the same molecule. As structural information of ORs is currently lacking, molecular modeling studies should help answer the question of how ORs with diverse structures can interact with the same odor ligand. Our receptor–based informatics studies will complement ligand–based chemoinformatics studies [41] to help understand how odors and ORs interact.

In summary, odor–induced S6 phosphorylation in activated OSNs, pS6–IP RNA–Seq, the ‘one neuron–one receptor’ rule, and *in vitro* validation enable us to efficiently and comprehensively identify ORs responsive to specific odor stimuli in awake, behaving animals. OR mapping for additional odors is likely to guide deeper understanding of olfactory information coding.

Experimental Procedures

Animals

WT C57BL/6 mice were purchased from Jackson. Animals of either sex were used at the age of 3–4 weeks. The procedures of animal handling and tissue harvesting were approved by the Institutional Animal Care and Use Committee of Duke University.

Odor stimulation

3–4week old C57BL/6 mice (Jackson Labs) were placed individually into sealed containers (volume ≈ 2.7L) inside a fume hood and allowed to rest for 1 hour in an odorless environment. For odor stimulus, 10μl odor solution or 10μl distilled water (control) was applied to 1cm × 1cm filter paper held in a cassette (Tissue–Tek). The cassette was placed into a new mouse container into which the mouse was also transferred, and the mouse was exposed to the odor solution or control for 1 hour. Experiments were performed in triplicates or quadruplicates, and within each replication the experimental and control mice were littermates of the same sex.

pS6–IP from the OE

pS6–IP was conducted as described with modifications [17]. The pS6 240–containing peptide that was used by Knight et al. to improve signal–noise ratio was not included in our protocol because we observed relatively low pS6 background in the OE, and pilot experiments with the peptide significantly decreased the RNA yield. Following odor stimulation, the mouse was sacrificed and the OE was dissected in 25 ml of dissection buffer (1 × HBSS [Gibco, with Ca^{2+} and Mg^{2+}], 2.5mM HEPES [pH 7.4 adjusted with KOH], 35mM glucose, 100 μg/ml cycloheximide, 5mM sodium fluoride, 1mM sodium orthovanadate, 1mM sodium pyrophosphate, 1mM beta–glycerophosphate) on ice. The dissected OE was transferred to 1.35ml homogenization buffer (150mM KCl, 5mM $MgCl_2$, 10mM HEPES [pH 7.4 adjusted with KOH], 100nM Calyculin A, 2mM DTT, 100 U/ml RNasin [Promega], 100 μg/ml cycloheximide, 5mM sodium fluoride, 1mM sodium orthovanadate, 1mM sodium pyrophosphate, 1mM beta–glycerophosphate, protease inhibitor [Roche, 1 tablet/10ml]) and homogenized 3 times at 250 rpm and 9 times at 750 rpm (Glas–Col). The homogenate was transferred to a 1.5 ml lobind tube (Eppendorf), and centrifuged at 4600 rpm for 10 minutes at 4°C. The supernatant was then transferred to a new 1.5 ml lobind tube, to which 90 μl 10%NP–40 and 90 μl 300 mM DHPC (Avanti Polar Lipids) was added. The mixture was centrifuged at 13000 rpm for 10 minutes at 4°C. The supernatant was transferred to a new 1.5 ml lobind tube, and mixed with 20 μl pS6 antibody (Cell signaling, #2215). Antibody binding was allowed by incubating the mixture for 1.5 hours at 4°C with rotation. During antibody binding, Protein A Dynabeads (Invitrogen, 100 μl/sample) was washed 3 times with 900 μl beads wash buffer 1 (150mM KCl, 5mM $MgCl_2$, 10mM HEPES [pH 7.4 adjusted with KOH], 0.05% BSA, 1% NP–40). After antibody binding, the mixture was added to the washed beads and gently mixed, followed by incubation for 1 hour at 4°C with rotation. After incubation, the RNA–bound beads were washed 4 times with 700 μl beads wash buffer 2 (RNase free water containing 350mM KCl, 5mM $MgCl_2$, 10 mM HEPES [pH 7.4 adjusted with KOH], 1% NP–40, 2mM DTT, 100U/ml recombinant RNasin [Promega], 100 μg/ml cycloheximide, 5 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM

sodium pyrophosphate, 1 mM beta–glycerophosphate). During the final wash, beads were placed onto the magnet and moved to room temperature. After removing supernatant, RNA was eluted by mixing the beads with 350 μl RLT (Qiagen). The eluted RNA was purified using RNeasy Micro kit (Qiagen). Chemicals were purchased from Sigma if not specified otherwise.

cDNA synthesis, PCR amplification and library preparation for next–generation sequencing

RT–PCR from a small amount of RNA was conducted as described (Saito et al. 2004) with modifications. Briefly, 1.5 μl purified RNA was mixed with 5 μl reaction mix (1× PCR buffer [Roche], 1.5 mM $MgCl_2$, 50 μM dNTPs, 2 ng/μl poly–T primer (TAT AGA ATT CGC GGC CGC TCG CGA TTT TTT TTT TTT TTT TTT TTT TTT), 0.04 U/μl RNase inhibitor [Qiagen], 0.4 U/μl recombinant RNasin [Promega]). This mixture was heated at 65°C for 1 min and cooled to 4°C. 0.3μl RT mix (170 U/μl Superscript II [Invitrogen], 0.4 U/μl RNase inhibitor [Qiagen], 4 U/μl recombinant RNasin [Promega], 3 μg/μl T4 gene 32 protein [Roche]) was added to each tube and incubated at 37°C for 10 minutes then 65°C for 10 minutes. 1μl ExoI mix (2 U/μl ExoI [NEB], 1× PCR buffer [Roche], 1.5 mM $MgCl_2$) was added to each tube and incubated at 37°C for 15 minutes then 80°C for 15 minutes. 5μl TdT mix (1.25U/μl TdT [Roche], 0.1 U/μl RNase H [Invitrogen], 1× PCR buffer [Roche], 3 mM dATP, 1.5 mM $MgCl_2$) was added to each tube and incubated at 37°C for 20 minutes then 65°C for 10 minutes. 3.5 μl of the product was added to 27.5 μl PCR mix (1× LA Taq reaction buffer [TaKaRa], 0.25mM dNTPs, 20 ng/μl poly–T primer, 0.05 U/μl LA Taq [TaKaRa]) and incubated at 95°C for 2 minutes, 37°C for 5 minutes, 72°C for 20 minutes, then 16 cycles of 95°C for 30 seconds, 67°C for 1 minute, 72°C for 3 minutes with 6 seconds extension for each cycle, and then 72°C for 10 minutes. The PCR product was purified by gel purification, and 50 ng of the purified product was used for library preparation with Nextera DNA Sample Prep kits (Illumina). Libraries were sequenced on MiSeq (for individual libraries) or HiSeq 2000/2500 (for pooled indexed libraries, 12 libraries per lane) in 50 base pair single read mode. Raw reads and quantification results can be accessed at GEO: GSE59324.

Reads mapping and test for enrichment

Short reads were aligned to the mouse reference genome mm10 using Bowtie [42]. The reads mapped to annotated genes were then counted using BEDTools [43]. Because (to date) the UTRs of many ORs are not well defined, we only used the coordinates of coding exons for all the genes to count the mapped reads. Because the similarity between ORs can result in multiple alignments, we implemented a stepwise mapping scheme that allows reads aligned to multiple regions to be distributed proportionally to these regions based on the previous read counts from these regions [44]. Briefly, the reads were aligned to mm10 by Bowtie with –m 20 –a (reports all reads that map to at most 20 different regions on the reference genome), and the resulting mapping file was separated into 20 sub–files that included reads mapped only to 1, 2, …, 20 regions respectively. BEDTools was used to first count reads mapped to coding exons of genes using only uniquely mapped reads. Reads mapped to two regions were distributed based on read counts from uniquely mapped reads. For example, if a read was mapped to two different genes, and the read counts of these two genes were 3 and 7

respectively based on uniquely mapped read counts, then 0.3 and 0.7 were added to the previous read counts and the new counts became 3.3 and 7.7. The iteration was allowed for 20 cycles and the resulted read counts were then rounded. Iteration number 20 was chosen because pilot parameter tuning indicated that the majority of read counts stopped increasing after 20 cycles.

The read counts table for all the genes were then imported into R and analyzed using the negative–binomial model based tool EdgeR [45]. A similar result was obtained using a separate tool DESeq2 employing similar model assumptions [46]. Littermate information was included in the design matrix to enhance the power of detection [47]. Multiple comparison was adjusted within the detected OR group by controlling false discovery rate (FDR) [48]. DESeq was used to calculate the size factors of individual libraries [46].

DNA and Vector Preparation

The open reading frames of ORs were amplified using Phusion polymerase (Finnzymes) following manufacturer's protocols. Amplified fragments were cloned into pCI (Promega) or Rho–pCI for sequence verification. Unless otherwise noted, ORs were cloned into Rho–pCI to introduce a Rho–tag (first 20 amino acids of rhodopsin) at the N–termini. The Rho–tag is known to facilitate the heterologous expression of ORs [23].

Cell Culture

Hana3A cells [9] were maintained in minimal essential medium (MEM) containing 10% fetal bovine serum with penicillin–streptomycin and amphotericin B, at 37°C and 5% CO_2.

Dual–Glo assay

The Dual–Glo assay for ORs was performed as described previously [26]. Briefly, Hana3A cells were plated on 96–well plates. 18 to 24 hours after plating, cells were transfected with plasmids coding for OR, M3–R, RTP1S, CRE–luciferase and pRL–SV40. 18 to 24 hours later, cells were stimulated by incubation with odorants diluted in CD293 for 4 hours at 37°C and 5% CO_2 to allow for CRE–luciferase expression. Luciferase and *Renilla* luciferase activities were measured using Dual–Glo kit following manufacturer's protocols (Promega). For dose–response curves, the background *Renilla* luciferase activity was subtracted from each data set. For fold of induction, the fold of signal increase was scaled so that luciferase measurements from cells transfected with empty vector was 0, and positive control (Olfr1126) 100. Data was analyzed using GraphPad Prism 5 or custom R scripts. The cut–off to determine positive or negative for *in vitro* responses to acetophenone was constructed as follows: A two–component Gaussian mixture model was used to fit the fold of luciferase activity increase at 300μM acetophenone stimulation. The cut–off was chosen such that it excluded 99% of the Gaussian component that represents the non–responding group. When dose–response curves with a wide range of stimulation concentrations were available (Fig. 4), we defined positive *in vitro* responses as the lower bounds of the 95% confidence intervals of the estimated maximum responses above 0. ORs with small response amplitude were independently confirmed by additional experiments using $10^{-5.5}$, 10^{-5}, $10^{-4.5}$, 10^{-4}, $10^{-3.5}$, and 10^{-3}M acetophenone concentrations for stimulation.

Protein sequence analysis of acetophenone ORs

1090 mouse OR protein sequences (pseudogenes with one or more predicted transmembrane regions disrupted were excluded from the analysis) were aligned by Clustalx with manual adjustments. To identify residues that were more conserved among the acetophenone ORs than by chance, we first simulated the distributions of mean Grantham distances within random OR sets for each residue in the alignment. To simulate these distributions, 48 random ORs were randomly sampled from the mouse OR repertoire. At each amino acid residue, the pairwise Grantham distances across these 48 ORs were calculated and their mean was computed. Random sampling was iterated 20000 times. The per–residue mean Grantham distances were also computed for the group of 48 acetophenone ORs, and used to find p–values under the null–hypothesis that they were samples from the distributions of random OR sets. Locations with gaps for more than 10% of the ORs were excluded from the study. Multiple comparison was adjusted by controlling false discovery rate (FDR) [48]. Visualization of residue frequencies were generated by WebLogo [49].

SVM and logistic regression with elastic net penalty were used to build models using OR sequence properties to predict responsiveness to acetophenone. Using the aforementioned multiple alignment of mouse OR proteins, we calculated amino acid properties (polarity, composition and volume) as defined by Grantham [29] for 291 amino acid residues common to at least 10% of the ORs, resulting in $3 \times 291 = 873$ predictors. The missing values in this matrix (0.17%) resulting from occasional gaps in alignment were imputed from column means. To construct 0–1 responses, we used ORs that were supported by both *in vivo* and *in vitro* evidence, with the 48 ORs that were enriched in our pS6 RNA–Seq and responded to acetophenone *in vitro* labeled as 1's, and 367 ORs that were not enriched in pS6 RNA–Seq ($p > 0.05$ in both 1% and 100% acetophenone stimulation experiments) and did not respond to acetophenone *in vitro* labeled as 0's. Elastic net logistic regression was performed using the glmnet package[50]. SVM was performed using the e1071 package. A 10–fold cross–validation scheme was performed on the mouse OR data. Each time, the data was divided into two subsets, for 100 iterations: 90% as training subset and 10% as testing subset. Logistic regression and SVM models were trained using the training set, before trained models were asked to predict the testing set and the predictions were compared with truth. True positive and false positive rates were calculated at all possible cut–offs to generate the ROC curves and AUC. Parameters ($\gamma = 0.001$, cost = 10 for RBF kernel SVM; $\alpha = 0.07$ for elastic net logistic regression) were tuned by grid search. External validation was performed using 27 human ORs that could be functionally expressed in our heterologous system.

Immunohistochemistry and in situ hybridizations

For immunohistochemical staining, 20 μm frozen sections of the olfactory epithelium and surrounding tissues of 3 week–old mice were incubated with rabbit anti–phospho–S6 (240/244) (Science Signaling) overnight at 4 °C. Following washing, donkey Cy3–conjugated anti–rabbit IgG (Jackson Immunologicals) was used to visualize pS6 protein signals.

Fluorescent in situ hybridization (FISH) was performed to detect the mRNA of the OR of interest, followed by immunohistochemical staining for pS6 induction. In situ hybridization

was carried out as previously described [9]. Briefly, digoxigenin (DIG)–labeled complementary RNA probes were hybridized overnight at 58 °C to target mRNAs in 20 μm frozen tissue sections of the olfactory epithelium. After washing, the sections were incubated with a horseradish peroxidase (HRP)–conjugated antibody against DIG (Roche). Hybridization signals were detected using tyramide signal amplification (TSA) using fluorescein (Perkin–Elmer) as the fluorophore. Following FISH, staining for pS6 induction was performed as described above.

pS6 signal intensities were quantified as follows. Z–stacked images with 2 μm intervals between each slice, were obtained at 200x magnification using the Zeiss Axiocam MRm and upright inverted fluorescent microscope with ApoTome functionality to take optical sections that reduced background fluorescence. The filter sets used were as follows: Zeiss filter set #38 for fluorescein, #43 for Cy3, and #49 for bisbenzimide. Following acquisition, the OR signal was merged as a maximum intensity projection in ImageJ, and each OR area was selected manually. For each slice in a Z-stacked image, the average intensity of the pixels of a selected OR area was multiplied by the area in arbitrary square units to yield an integrated density value. The maximum integrated density measurement among the Z–stacks for the selected OR area was used for signal intensity quantification. The pS6 signal was merged as a maximum intensity projection, and used for measuring the average fluorescence of the entire olfactory epithelium. The corrected total cell fluorescence was calculated as follows:

Whole–cell signal = maximum integrated density of a cell

Background signal = average signal per pixel for the entire olfactory epithelium

Corrected total cell fluorescence = Whole–cell signal – (area of selected cell X background signal)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Gilad Barnea and Dr. Richard Axel for antibodies against M72, Dr. Hanyi Zhuang for some of the OR constructs, Dr. Douglas Marchuk for sharing of equipment, the Duke IGSP Genome Sequencing and Analysis Core Resource for providing Illumina sequencing service, and Dr. Sayan Mukherjee and Dr. Joel Mainland for helpful discussions and comments on the manuscript. This work is supported by grants from the NIH.

References

1. Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell. 1991; 65:175–187. [PubMed: 1840504]
2. Niimura Y, Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome research. 2014; 24:1485–1496. [PubMed: 25053675]
3. Malnic B, Hirono J, Sato T, Buck LB. Combinatorial receptor codes for odors. Cell. 1999; 96:713–723. [PubMed: 10089886]
4. Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. Odor coding by a mammalian receptor repertoire. Science signaling. 2009; 2:ra9. [PubMed: 19261596]

5. Peterlin Z, Firestein S, Rogers ME. The state of the art of odorant receptor deorphanization: A report from the orphanage. The Journal of general physiology. 2014; 143:527–542. [PubMed: 24733839]
6. Zhao H, et al. Functional expression of a mammalian odorant receptor. Science. 1998; 279:237–242. [PubMed: 9422698]
7. Touhara K, et al. Functional identification and reconstitution of an odorant receptor in single olfactory neurons. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:4040–4045. [PubMed: 10097159]
8. Bozza T, Feinstein P, Zheng C, Mombaerts P. Odorant receptor expression defines functional units in the mouse olfactory system. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2002; 22:3033–3043. [PubMed: 11943806]
9. Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H. RTP family members induce functional expression of mammalian odorant receptors. Cell. 2004; 119:679–691. [PubMed: 15550249]
10. Araneda RC, Kini AD, Firestein S. The molecular receptive range of an odorant receptor. Nature neuroscience. 2000; 3:1248–1255. [PubMed: 11100145]
11. Grosmaitre X, et al. SR1, a mouse odorant receptor with an unusually broad response profile. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009; 29:14545–14552. [PubMed: 19923288]
12. Mainland JD, et al. The missense of smell: functional variability in the human odorant receptor repertoire. Nature neuroscience. 2014; 17:114–120. [PubMed: 24316890]
13. Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H. Genetic variation in a human odorant receptor alters odour perception. Nature. 2007; 449:468–472. [PubMed: 17873857]
14. Menashe I, et al. Genetic elucidation of human hyperosmia to isovaleric acid. PLoS biology. 2007; 5:e284. [PubMed: 17973576]
15. Shirasu M, et al. Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. Neuron. 2014; 81:165–178. [PubMed: 24361078]
16. Dong, PY.; Gong, NN.; Matsunami, H. Cell-Based System for Identification of Olfactory Receptors. In: Park, TH., editor. Bioelectronic Nose. Springer; Netherlands: 2014. p. 83-95.
17. Knight ZA, et al. Molecular profiling of activated neurons by phosphorylated ribosome capture. Cell. 2012; 151:1126–1137. [PubMed: 23178128]
18. Haga S, et al. The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. Nature. 2010; 466:118–122. [PubMed: 20596023]
19. Isogai Y, et al. Molecular organization of vomeronasal chemoreception. Nature. 2011; 478:241–245. [PubMed: 21937988]
20. Chess A, Simon I, Cedar H, Axel R. Allelic inactivation regulates olfactory receptor gene expression. Cell. 1994; 78:823–834. [PubMed: 8087849]
21. Zhang J, Huang G, Dewan A, Feinstein P, Bozza T. Uncoupling stimulus specificity and glomerular position in the mouse olfactory system. Molecular and cellular neurosciences. 2012; 51:79–88. [PubMed: 22926192]
22. Oka Y, et al. Odorant receptor map in the mouse olfactory bulb: in vivo sensitivity and specificity of receptor-defined glomeruli. Neuron. 2006; 52:857–869. [PubMed: 17145506]
23. Krautwurst D, Yau KW, Reed RR. Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell. 1998; 95:917–926. [PubMed: 9875846]
24. McClintock TS, et al. In vivo identification of eugenol-responsive and muscone-responsive mouse odorant receptors. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014; 34:15669–15678. [PubMed: 25411495]
25. Meister M, Bonhoeffer T. Tuning and topography in an odor map on the rat olfactory bulb. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2001; 21:1351–1360. [PubMed: 11160406]
26. Zhuang H, Matsunami H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. Nature protocols. 2008; 3:1402–1413. [PubMed: 18772867]

27. Li YR, Matsunami H. Activation state of the M3 muscarinic acetylcholine receptor modulates mammalian odorant receptor signaling. Science signaling. 2011; 4:ra1. [PubMed: 21224444]
28. Zhang X, Firestein S. The olfactory receptor gene superfamily of the mouse. Nature neuroscience. 2002; 5:124–133. [PubMed: 11802173]
29. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185:862–864. [PubMed: 4843792]
30. Bishop, CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc; 2006.
31. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B. 2005; 67:301–320.
32. Takahashi LK, Nakashima BR, Hong H, Watanabe K. The smell of danger: a behavioral and neural analysis of predator odor-induced fear. Neuroscience and biobehavioral reviews. 2005; 29:1157–1167. [PubMed: 16095694]
33. Kobayakawa K, et al. Innate versus learned odour processing in the mouse olfactory bulb. Nature. 2007; 450:503–508. [PubMed: 17989651]
34. Weid B, et al. In vivo identification of receptor-ligand pairs via large-scale transcriptional profiling of chemosensory neurons. Nature neuroscience. 2015
35. Thiebaud N, et al. Odorant metabolism catalyzed by olfactory mucosal enzymes influences peripheral olfactory responses in rats. PloS one. 2013; 8:e59547. [PubMed: 23555703]
36. Nagashima A, Touhara K. Enzymatic conversion of odorants in nasal mucus affects olfactory glomerular activation patterns and odor perception. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2010; 30:16391–16398. [PubMed: 21123585]
37. Scott JW, Sherrill L, Jiang J, Zhao K. Tuning to odor solubility and sorption pattern in olfactory epithelial responses. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014; 34:2025–2036. [PubMed: 24501345]
38. Grossisseroff R, Lancet D. Concentration-Dependent Changes of Perceived Odor Quality. Chemical senses. 1988; 13:191–204.
39. Li J, Haddad R, Chen S, Santos V, Luetje CW. A broadly tuned mouse odorant receptor that detects nitrotoluenes. Journal of neurochemistry. 2012; 121:881–890. [PubMed: 22443178]
40. Nara K, Saraiva LR, Ye X, Buck LB. A large-scale analysis of odor coding in the olfactory epithelium. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011; 31:9179–9191. [PubMed: 21697369]
41. Boyle SM, McInally S, Ray A. Expanding the olfactory code by in silico decoding of odor-receptor chemical space. eLife. 2013; 2:e01120. [PubMed: 24137542]
42. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10:R25. [PubMed: 19261174]
43. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
44. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 2008; 5:621–628. [PubMed: 18516045]
45. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
46. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11:R106. [PubMed: 20979621]
47. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic acids research. 2012; 40:4288–4297. [PubMed: 22287627]
48. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57:289–300.
49. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome research. 2004; 14:1188–1190. [PubMed: 15173120]

50. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software. 2010; 33:1–22. [PubMed: 20808728]

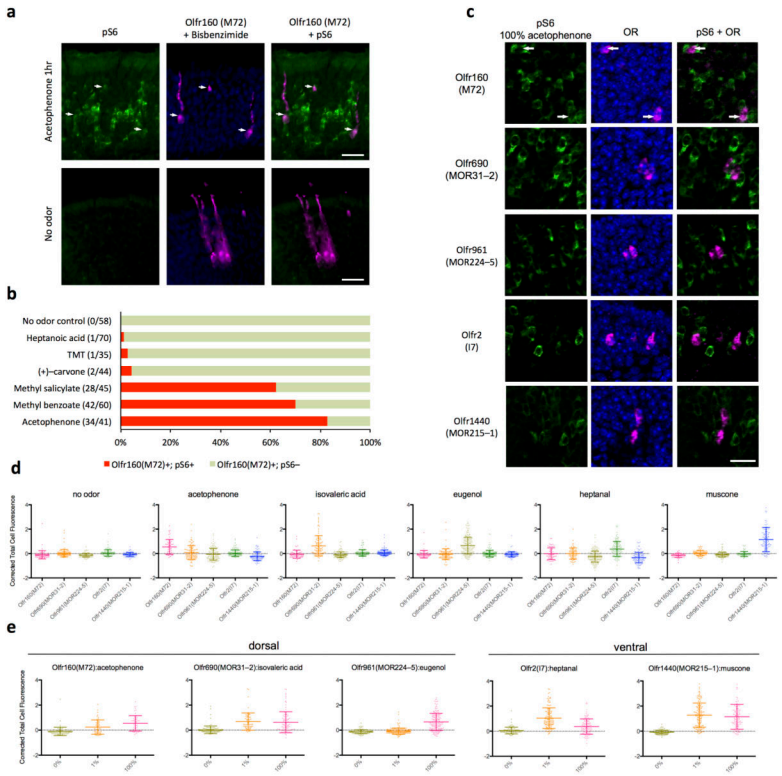

Figure 1.

Odor stimulation induces S6 phosphorylation in the mouse OE.

(a) Coronal section of OE stimulated with 100% acetophenone (upper) and no odor (control) condition for 1 hour. Green: Antibody staining for pS6. Magenta: Antibody staining for a known acetophenone receptor, M72. Blue: Bisbenzimide staining showing the nuclei. Arrowheads indicate colocalization of M72 and pS6 signals. Scale bar, 25 μm.

(b) Quantification of pS6 induction in M72–expressing OSNs following odor stimulation by M72 agonists (methyl salicylate, methyl benzoate, acetophenone) and controls [no odor, heptanoic acid, TMT, (+)–carvone].

(c) Coronal section of OE stimulated with 100% acetophenone for 1 hour. Green: Antibody staining for pS6. Magenta: RNA FISH for 5 ORs, respectively. Blue: Bisbenzimide staining showing the nuclei. Arrowheads indicate colocalization of M72 and pS6 signals. Scale bar, 25 μm.

(d) Quantification of pS6 staining intensity for 5 known OR–odorant pairs.

(e) Quantification of pS6 staining intensity following 1% and 100% odorant stimulation for 5 known OR–odorant pairs.

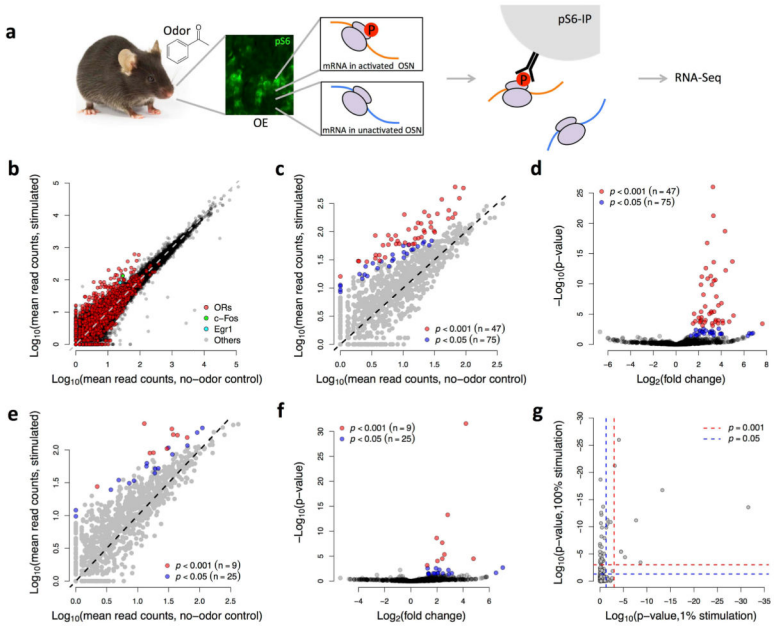

Figure 2.

pS6–IP enriches OR mRNAs from odor stimulated OE.

(a) Scheme of the experiment. When the animal is exposed to odor, ribosome subunit S6 undergoes phosphorylation in odor–responding OSNs. pS6–IP enriches for mRNA species expressed in the activated OSNs, which can be then profiled by RNA–Seq.

(b) Scatter plot comparing immunoprecipitated mRNA counts from stimulated sample (100% acetophenone) versus unstimulated sample. X–axis: Mean read counts of genes in unstimulated IP samples (n = 3). Y–axis: Mean read counts of genes in acetophenone stimulated IP samples (n = 3). Red dots represent ORs. Gray dots represent non–OR genes. Broken line: unit–slope.

(c) Differential enrichment calling of OR mRNA. 75 ORs are enriched in the 100% acetophenone stimulated group with p–value smaller than 0.05, after adjusting for multiple comparisons across the detected OR repertoire. Broken line: unit–slope.

(d) Volcano plot showing enrichment of OR mRNA in 100% acetophenone stimulated group.

(e) Differential enrichment calling of OR mRNA. 25 ORs are enriched in the 1% acetophenone stimulated group with p–value smaller than 0.05, after adjusting for multiple comparisons across the detected OR repertoire. Broken line: unit–slope.

(f) Volcano plot showing enrichment of OR mRNA in 1% acetophenone stimulated group.

(g) Scatter plot comparing p–values of enrichment in 100% acetophenone versus 1% acetophenone stimulated samples. Red dashed line: $p = 0.001$. Blue dashed line: $p = 0.05$. Note the absence of ORs in the bottom right corner.

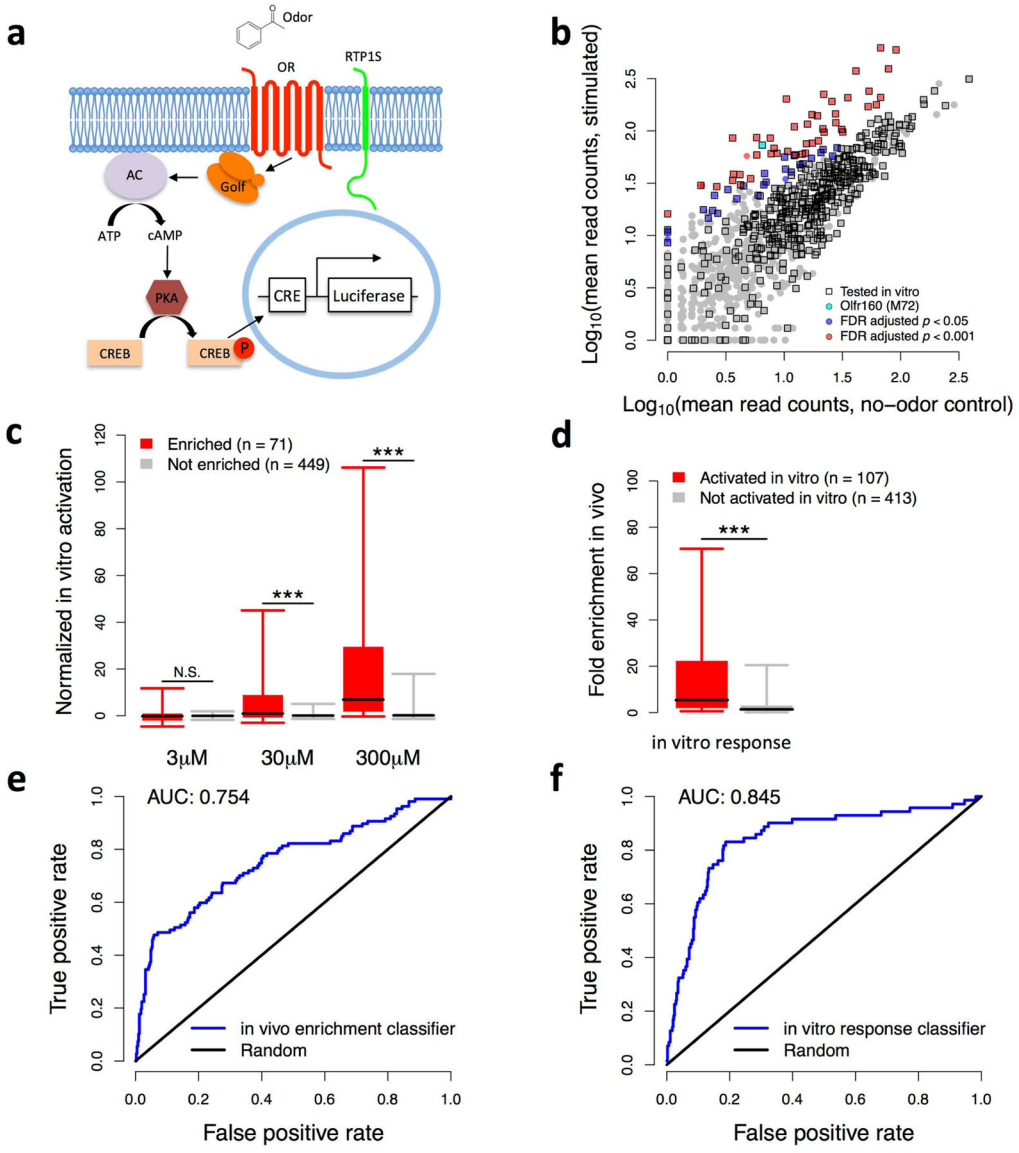

Figure 3.

Correlation between *in vivo* and *in vitro* responses.

(a) The heterologous OR signaling pathway. AC, adenylyl cyclase; CRE, cAMP response element; CREB, cAMP response element–binding protein; PKA, protein kinase A; RTP1S, receptor–transporting protein 1 (short).

(b) ORs tested for *in vitro* responses to acetophenone.

(c) *In vitro* activation of 71 enriched and 449 un–enriched ORs. Y–axis: normalized fold of increase in luciferase signals. 100% is determined by the fold of increase of Olfr1126 stimulated with 300 μM acetophenone. 0% is determined by the fold of increase of empty vector stimulated with 3 μM acetophenone. Red: ORs enriched at $p < 0.05$. Gray: ORs not called enriched at $p = 0.05$. Black bar: median; Box: 25 and 75 percentile; Whisker: 5 and 95 percentile. Wilcoxon rank sum test for difference of *in vitro* responses between enriched and not enriched: 3 μM $p = 0.1$, 30 μM $p = 0.005$, 300 μM $p = 7 \times 10^{-21}$.

(d) *In vivo* enrichment of 107 activated and 413 not activated ORs. Y–axis: fold of enrichment of transcripts by pS6 IP. Red: ORs activated *in vitro*. Gray: ORs not activated *in vitro*. Black bar: median; Box: 25 and 75 percentile; Whisker: 5 and 95 percentile. Wilcoxon rank sum test for difference of *in vivo* enrichment between activated and not activated: $p = 1 \times 10^{-14}$.

(e) ROC curves illustrating performance of classifiers using *in vivo* enrichment p–values to predict whether the OR responds to acetophenone *in vitro*. Area Under Curve: 0.754, $p = 6 \times 10^{-16}$, Wilcoxon rank–sum test (one tailed against H_0: Classifier performance no better than random)..

(f) ROC curves illustrating performance of classifiers using *in vitro* responses to predict whether the OR is enriched at $p < 0.05$. Area Under Curve: 0.845, $p = 7 \times 10^{-21}$, Wilcoxon rank–sum test (one tailed against H_0: Classifier performance no better than random).

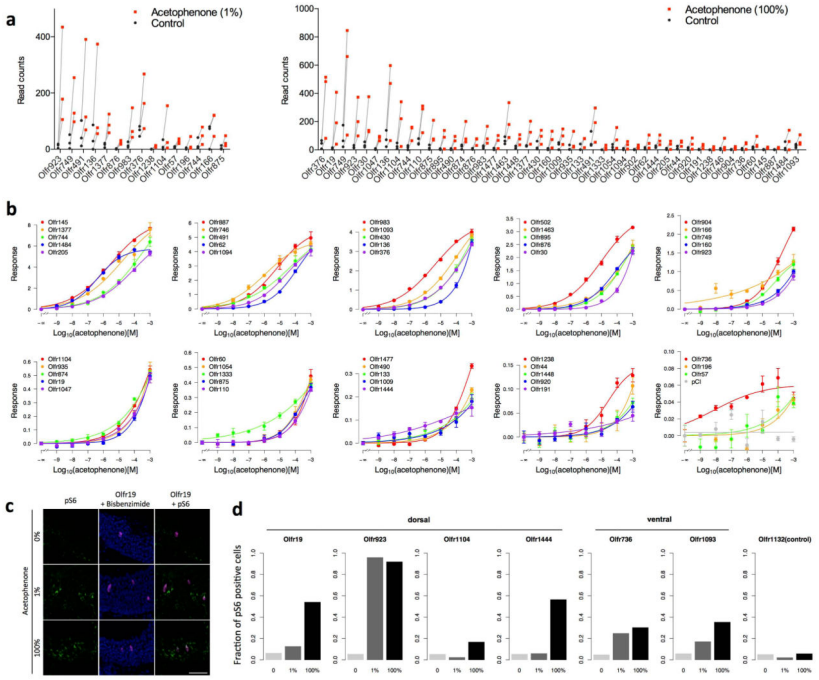

Figure 4.

Identification of acetophenone receptors

(a) Read counts of acetophenone receptors in control and 1% acetophenone stimulated samples (left, n = 3 pairs). Read counts of acetophenone receptors in control and 100% acetophenone stimulated samples (right, n = 3 pairs).

(b) *in vitro* responses of the acetophenone receptors. Responses are scaled to Olfr160 (M72). The maximum response of Olfr160 is defined as 1.

(c) Coronal section of OE following acetophenone stimulation for 1 hour. Green: Antibody staining for pS6. Magenta: RNA FISH for a newly identified acetophenone OR, Olfr19. Blue: Bisbenzimide staining showing the nuclei. Scale bar, 25 μm.

(d) Quantification of pS6 induction in OSNs expressing several newly–identified acetophenone ORs expressed in the dorsal OE (Olfr19, Olfr923, Olfr1104, Olfr1444) and ventral OE (Olfr736 and Olfr1093), along with a control OR (Olfr1132) following 1% and 100% acetophenone stimulation.

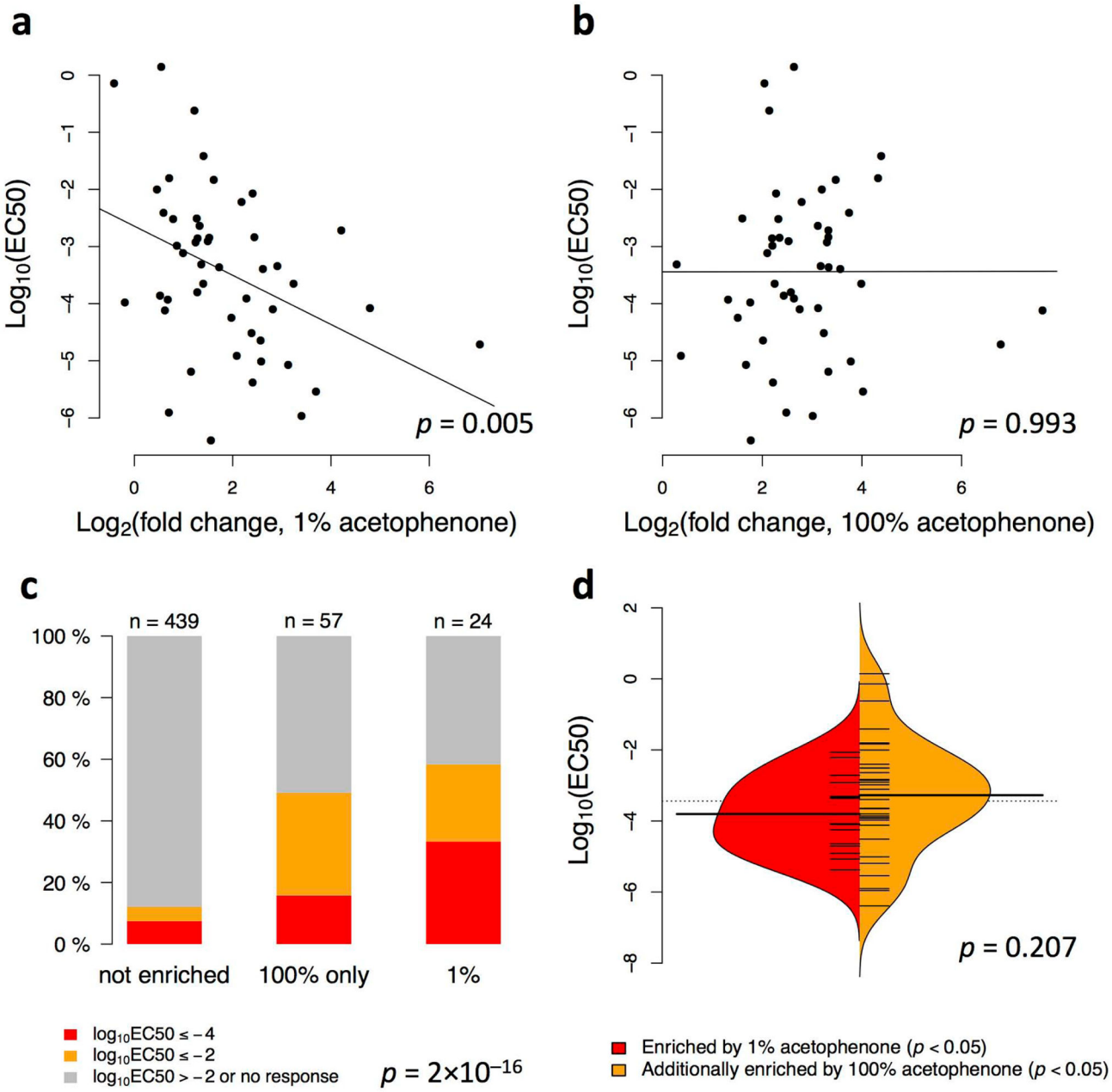

Figure 5.

Correlation between *in vivo* and *in vitro* sensitivities of ORs.

(a) Correlation between EC50 values measured for acetophenone *in vitro* and fold change of RNA transcript abundance following 1% acetophenone stimulation in the pS6–IP experiment *in vivo*. p = 0.005, linear regression and ANOVA. Spearman's rho=–0.395.

(b) Correlation between EC50 values measured for acetophenone *in vitro* and fold change of RNA transcript abundance following 100% acetophenone stimulation in the pS6–IP experiment *in vivo*. p = 0.993, linear regression and ANOVA. Spearman's rho=0.00137.

(c) Fractions of ORs that are relatively more sensitive ($\log_{10}$EC50 ≤–4), moderately sensitive ($\log_{10}$EC50 between –2 and –4), and not sensitive ($\log_{10}$EC50 > –2), in the groups

of ORs not enriched in pS6–IP, only enriched by 100% acetophenone, and enriched by 1% acetophenone. $p < 2 \times 10^{-16}$, Chi–square test.

(d) Beanplot showing the distribution of $\log_{10}$EC50 values in ORs only enriched by 100% acetophenone as compared to those enriched by 1% acetophenone. $p = 0.207$, Wilcoxon rank–sum test.

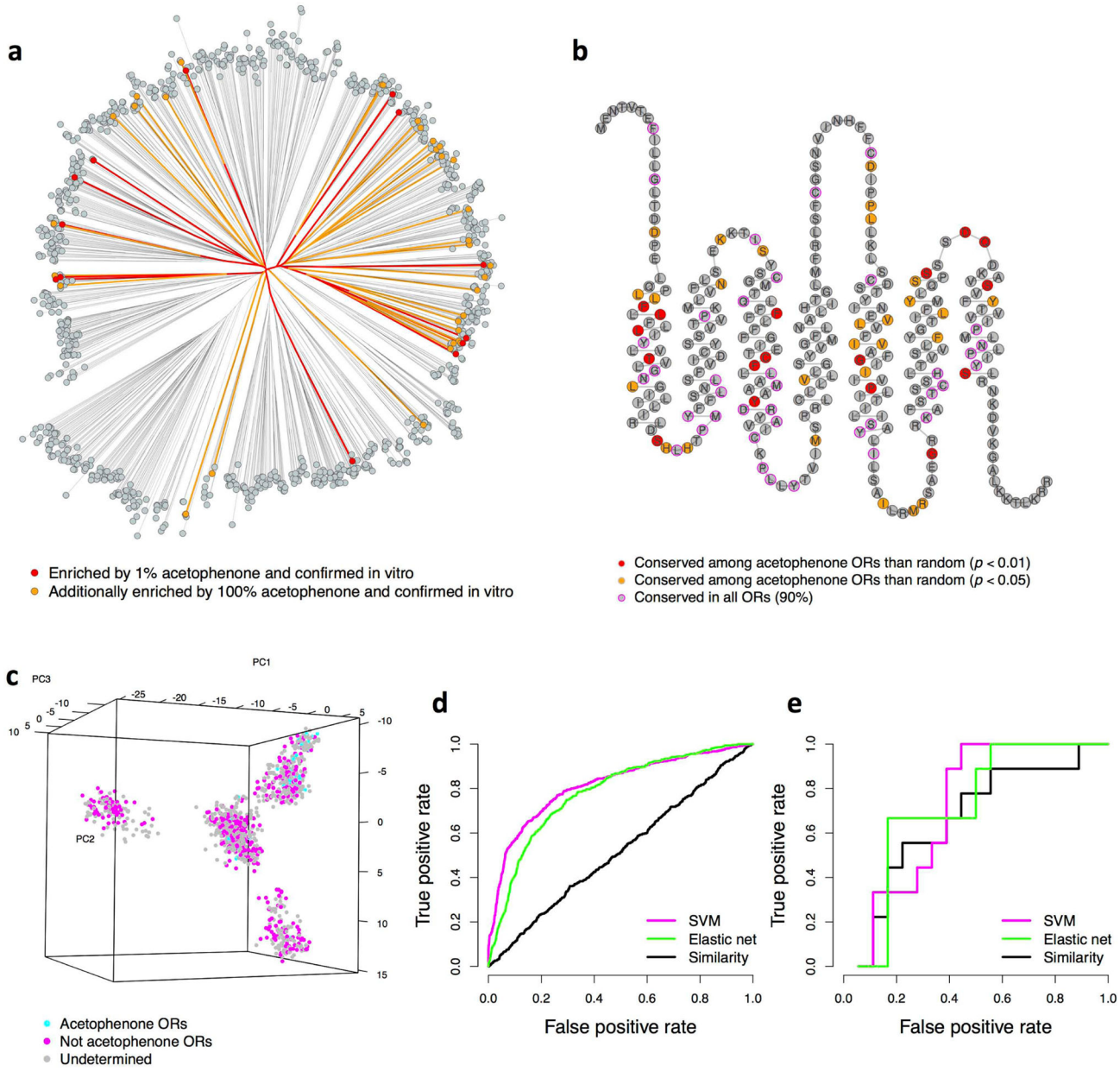

Figure 6.

Sequence–function analysis of the identified acetophenone receptors.

(a) Distance tree of OR protein sequences. Red, ORs enriched by 1% acetophenone and confirmed *in vitro*. Orange, ORs additionally enriched by 100% acetophenone and confirmed *in vitro*.

(b) Amino acid residues that are more conserved in the acetophenone receptors than random OR sets of the same size. Red, $p < 0.01$. Orange, $p < 0.05$.

(c) Principle component analysis of amino acid properties of mouse ORs. Plot showing the first three principle components (variance explained: 6.8%, 3.7%, 2.6%). Cyan: acetophenone ORs. Magenta: ORs that do not respond to acetophenone both *in vivo* and *in vitro*. Gray: other ORs.

(d) ROC curve illustrating cross–validation of SVM (magenta), elastic–net logistic regression (green) and overall sequence similarity based (black) models on mouse ORs. (e) ROC curve illustrating external validation of SVM (magenta), elastic–net logistic regression (green) and overall sequence similarity based (black) models using *in vitro* data of 27 human ORs.

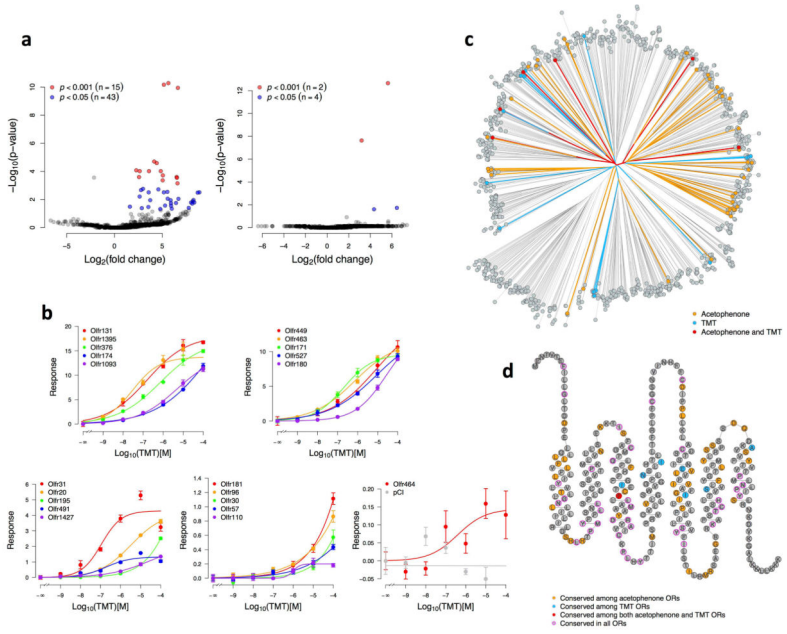

Figure 7.

Identification of ORs activated by TMT.

(a) Volcano plot showing enrichment of OR mRNA in 100% TMT stimulated group (left) and 1% acetophenone stimulated group (right).

(b) *in vitro* responses of the TMT receptors.

(c) Distance tree of OR protein sequences comparing acetophenone and TMT ORs. Orange, ORs activated by acetophenone. Blue, ORs activated by TMT. Red, ORs activated by both acetophenone and TMT.

(d) Amino acid residues that are conserved in the acetophenone receptors (orange, $p < 0.05$), TMT receptors (blue, $p < 0.05$), and both (red). The amino acid residues that are conserved in all ORs are labeled for comparison (90%, magenta circles).

Table 1

Performance of models in 10-fold cross-validation and external validation

	Cross-validation		External validation	
	AUC	**P–value**	**AUC**	**P–value**
SVM	0.815	1×10^{-120}	0.716	0.04
Elastic net	0.782	3×10^{-97}	0.716	0.04
Similarity	0.518	0.1	0.66	0.1