

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

Yue Jiang¹⁻³, Naihua Natalie Gong¹, Xiaoyang Serene Hu¹, Mengjue Jessica Ni¹, Radhika Pasi¹ & Hiroaki Matsunami^{1,4}

The mammalian olfactory system uses a large family of odorant receptors (ORs) to detect and discriminate amongst a myriad of volatile odor molecules. Understanding odor coding requires comprehensive mapping between ORs and corresponding odors. We developed a means of high-throughput *in vivo* identification of OR repertoires responding to odorants using phosphorylated ribosome immunoprecipitation of mRNA from olfactory epithelium of odor-stimulated mice followed by RNA-Seq. This approach screened the endogenously expressed ORs against an odor in one set of experiments using awake and freely behaving mice. In combination with validations in a heterologous system, we identified sets of ORs for two odorants, acetophenone and 2,5-dihydro-2,4,5-trimethylthiazoline (TMT), encompassing 69 OR-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. Our results provide a method for understanding the combinatorial coding of odors *in vivo*.

In mammals, olfactory sensation starts with the detection of odor ligands predominantly by ORs, a large family of seven transmembrane G protein-coupled receptors (GPCRs)^{1,2}. ORs are individually expressed in olfactory sensory neurons (OSNs) located in the olfactory epithelium (OE). The mouse genome encodes over 1,000 intact ORs². 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 odorant⁵. 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 mice⁸ and *in vitro* OR expression⁹ 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 humans¹²⁻¹⁶, 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 activation¹⁷. 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 neurons¹⁷. This approach has been successfully applied in brain regions such as the hypothalamus to identify markers of neurons that respond to feeding, starvation and high salt¹⁷.

We found that S6 phosphorylation occurs in the mouse OE following odor stimulation and is a reliable marker for OSN activation. pS6-IP revealed enrichment in subsets of OR mRNAs in mice stimulated with odorants. Using both high-throughput *in vivo* mapping and *in vitro* validation, we identified 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 1,000 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 would likely be 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 respond 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

¹Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA. ²University Program in Genetics and Genomics, Duke University Medical Center, Durham, North Carolina, USA. ³Department of Statistical Science, Duke University, Durham, North Carolina, USA. ⁴Department of Neurobiology, Duke Institute for Brain Sciences, Duke University Medical Center, Durham, North Carolina, USA. Correspondence should be addressed to H.M. (hiroaki.matsunami@duke.edu).

Received 19 June; accepted 9 August; published online 31 August 2015; doi:10.1038/nn.4104

enclosing a piece of filter paper spotted with 10 μ l undiluted (100%) or 1% acetophenone (odor), or distilled water (control) in a clean disposable cage. We killed the animals 1 h later and stained coronal sections from the OE with anti-pS6 antibody (Fig. 1a). Although 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: $16 \pm 3\%$, $n = 3$ images; 1% acetophenone: $6 \pm 2\%$, $n = 3$ 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 Olfr160-expressing OSNs show S6 phosphorylation in response to acetophenone stimulation. To determine the specificity of pS6 induction

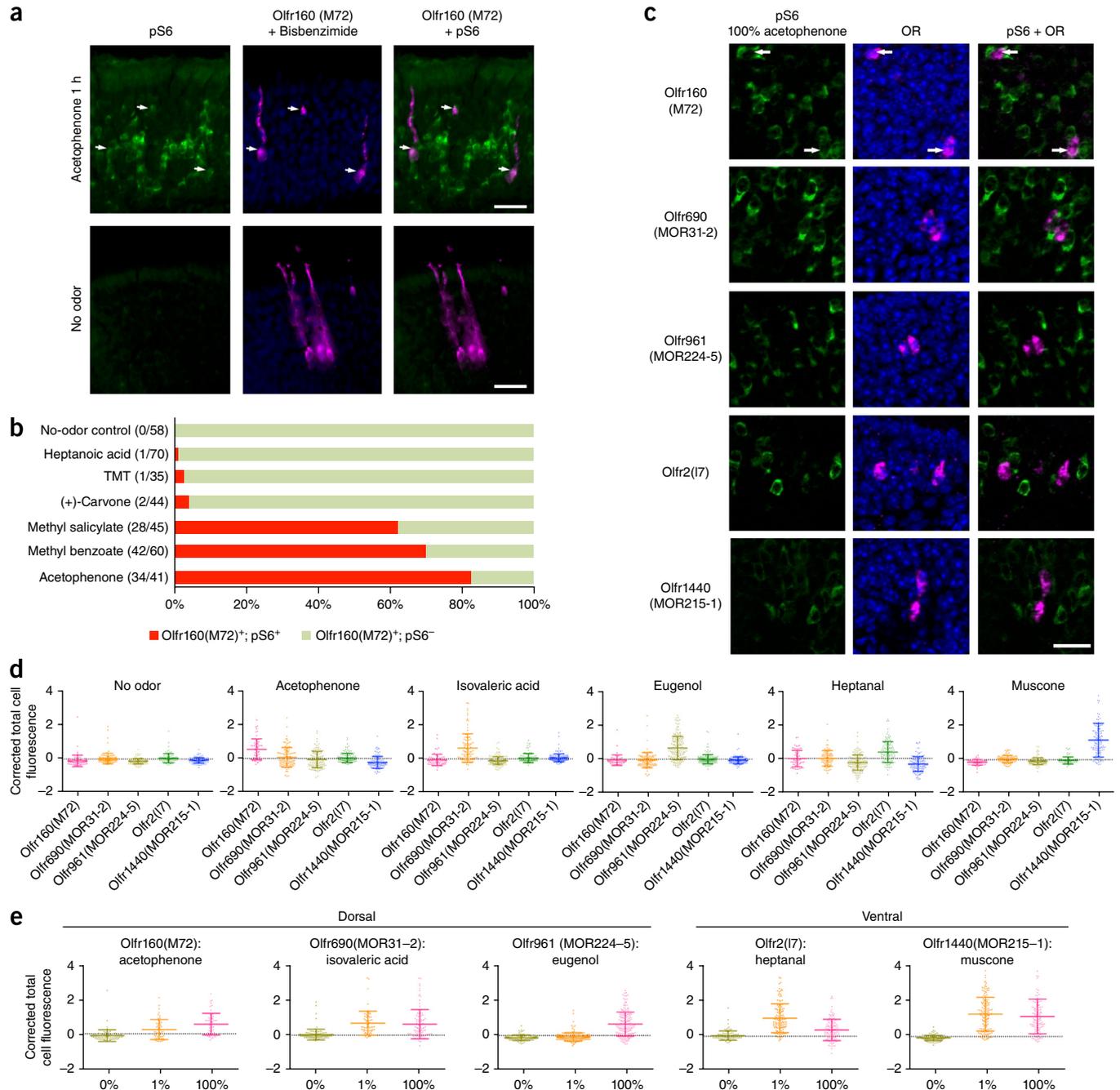
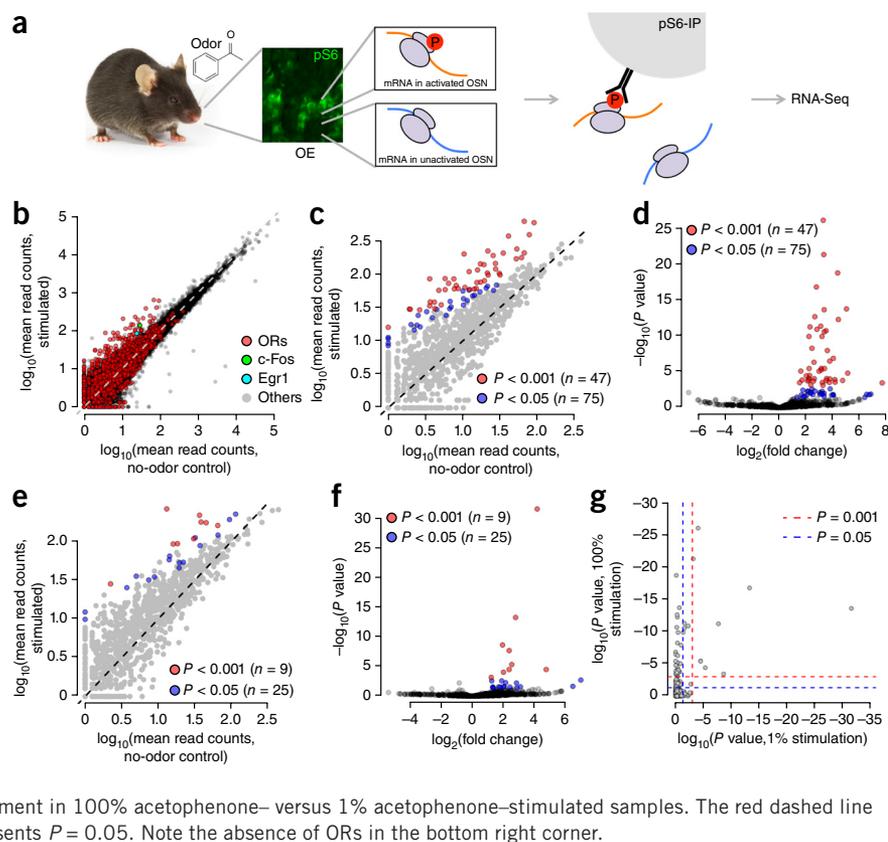


Figure 1 Odor stimulation induces S6 phosphorylation in the mouse OE. (a) Coronal section of OE stimulated with 100% acetophenone (upper) and no odor (control) condition for 1 h. Green represents antibody staining for pS6, magenta represents antibody staining for the known acetophenone receptor Olfr160, and blue represents bisbenzimidazole staining showing the nuclei. Arrowheads indicate colocalization of Olfr160 and pS6 signals. Scale bars represent 25 μ m. (b) Quantification of pS6 induction in Olfr160-expressing OSNs following odor stimulation by Olfr160 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 h. Green represents antibody staining for pS6, magenta represents RNA FISH for 5 ORs, and blue represents bisbenzimidazole staining showing the nuclei. Arrows indicate colocalization of Olfr160 and pS6 signals. Scale bar represents 25 μ m. (d) Quantification of pS6 staining intensity for five known OR-odorant pairs. (e) Quantification of pS6 staining intensity following 1% and 100% odorant stimulation for five known OR-odorant pairs. Error bars indicate s.d.

Figure 2 pS6-IP enriches OR mRNAs from odor stimulated OE. (a) Scheme of the experiment. When the animal was exposed to odor, ribosome subunit S6 underwent phosphorylation in odor-responding OSNs. pS6-IP enriched for mRNA species expressed in the activated OSNs, which could then be 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 the broken line represents the unit slope. (c) Differential enrichment calling of OR mRNA. 75 ORs were enriched in the 100% acetophenone-stimulated group with P values smaller than 0.05, after adjusting for multiple comparisons across the detected OR repertoire. Broken line represents unit slope. (d) Volcano plot showing enrichment of OR mRNA in 100% acetophenone-stimulated group. (e) Differential enrichment calling of OR mRNA. 25 ORs were 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 represents 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. The red dashed line indicates $P = 0.001$ and the blue dashed line represents $P = 0.05$. Note the absence of ORs in the bottom right corner.



in response to odor exposure, we stimulated the mice with other known Olfr160 agonists²¹ along with control odors (Supplementary Fig. 1a) and quantified the proportion of Olfr160-expressing OSNs that showed 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, whereas, in animals stimulated by control odors (heptanoic acid, TMT, (+)-carvone), pS6 signals showed little overlap with Olfr160 signals (Fig. 1b and Supplementary Fig. 1b).

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 μ l, undiluted) and performed *in situ* hybridization with each of the five OR probes followed by pS6 immunostaining (Fig. 1c,d and Supplementary Fig. 2a). 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 with 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. 3 and Supplementary Table 1). Time-course analysis of pS6 signals after acetophenone exposure revealed that pS6 signals started by 30 min and reached a plateau

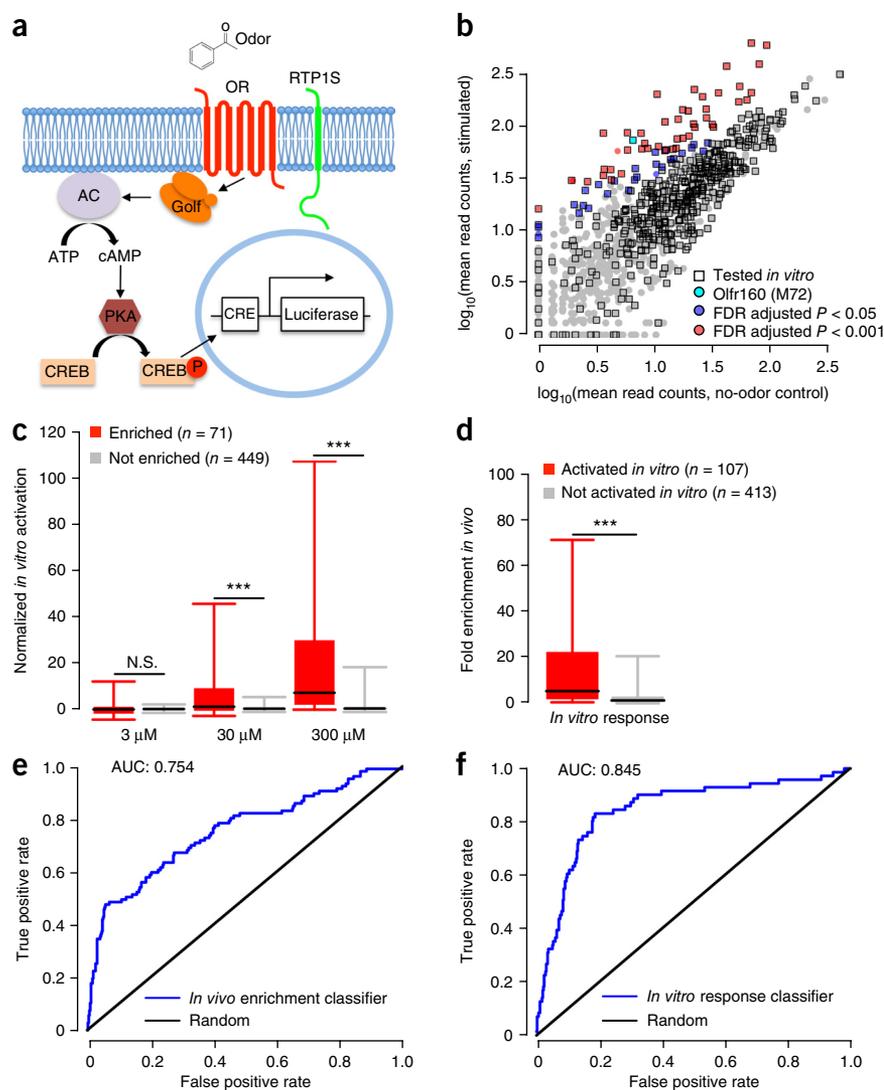
at 1 h (Supplementary Fig. 2b). These results suggest that pS6 is a reliable marker for OSN activation following odor stimulation.

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 consisting of littermates of the same gender and age, were used for each pS6-IP RNA-Seq experiment. In each pair, mice were exposed to either acetophenone or no odor control. We dissected the OE 1 h later, homogenized the tissue, immunoprecipitated pS6 along with associated RNAs and analyzed the purified RNA. Comparison of the samples from mice stimulated with 10 μ l of acetophenone versus control revealed a higher level of the immediate early genes *c-Fos* and *Egr1* in the stimulated samples, suggesting that pS6-IP enriches mRNA species in activated OSNs (Fig. 2b). We measured expression levels for ORs by counting reads that mapped to 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 μ l of undiluted (100%) acetophenone (47 and 75 ORs with $P < 0.001$ and $P < 0.05$, respectively, false discovery rate (FDR) adjusted; Fig. 2c,d). Stimulation with 10 μ l of 1% acetophenone (Fig. 2e,f) enriched 9 and 25 ORs ($P < 0.001$ and $P < 0.05$, respectively, FDR adjusted). Most of these ORs were also enriched following 100% acetophenone stimulation (9 of 9 for ORs enriched with $P < 0.001$, 14 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}.

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 unenriched ORs. *y* axis, normalized fold of increase in luciferase signals. 100% was determined by the fold of increase of Olfr1126 stimulated with 300 μ M acetophenone. 0% was determined by the fold of increase of empty vector stimulated with 3 μ M acetophenone. Red indicates ORs enriched at $P < 0.05$ and gray indicates ORs not enriched ($P > 0.05$). Black bars represent median, boxes represent 25th and 75th percentiles, and whiskers represent 5th and 95th percentiles. Wilcoxon rank-sum test for difference of *in vitro* responses between enriched and unenriched: 3 μ M $P = 0.1$, 30 μ M $P = 0.005$, 300 μ M $P = 7 \times 10^{-21}$. (d) *In vivo* enrichment of 107 activated and 413 not activated ORs. *y* axis, fold of enrichment of transcripts by pS6 IP. Red indicates ORs activated *in vitro* and gray indicates ORs not activated *in vitro*. Black bars represent median, boxes represent 25th and 75th percentiles, and whiskers represent 5th and 95th percentiles. Wilcoxon rank-sum test for difference of *in vivo* enrichment between activated and not activated: $P = 10^{-14}$. N.S., $P > 0.05$; *** $P < 0.01$. (e) ROC curves illustrating performance of classifiers using *in vivo* enrichment P values to predict whether the OR responds to acetophenone *in vitro*. AUC: 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 was enriched at $P < 0.05$. AUC: 0.845, $P = 7 \times 10^{-21}$, Wilcoxon rank-sum test (one tailed against H_0 : classifier performance no better than random).



Enriched ORs tend to respond to acetophenone *in vitro*

To confirm 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. 3a). To systematically test whether the enriched ORs are more likely to respond to acetophenone, we tested 71 ORs enriched by 100% acetophenone (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 an HEK293T-derived cell line that supports the robust expression of various transiently expressed ORs^{9,26}, and stimulated the cells with 3 μ M, 30 μ M or 300 μ M 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 with control stimulation. As expected, acetophenone-induced *in vitro* activation of enriched ORs was higher than that of control ORs ($P = 0.005$ at 30 μ M, $P = 7 \times 10^{-21}$ at 300 μ M, 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 ($P = 10^{-14}$, Wilcoxon rank-sum test; Fig. 3d). The cut-off for defining positive response to acetophenone *in vitro* was a 2.33-fold increase in luciferase induction at 300 μ M (see Online Methods for the determination of the cut-off, which was

set to exclude 99% of negative receptors). Of the 75 ORs enriched in the 100% acetophenone experiment ($P < 0.05$), 71 ORs were tested *in vitro* and 69% (49 ORs) responded. In contrast, of the 1,047 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 our *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 in the S6 phosphorylation analysis with acetophenone, expressed these ORs in

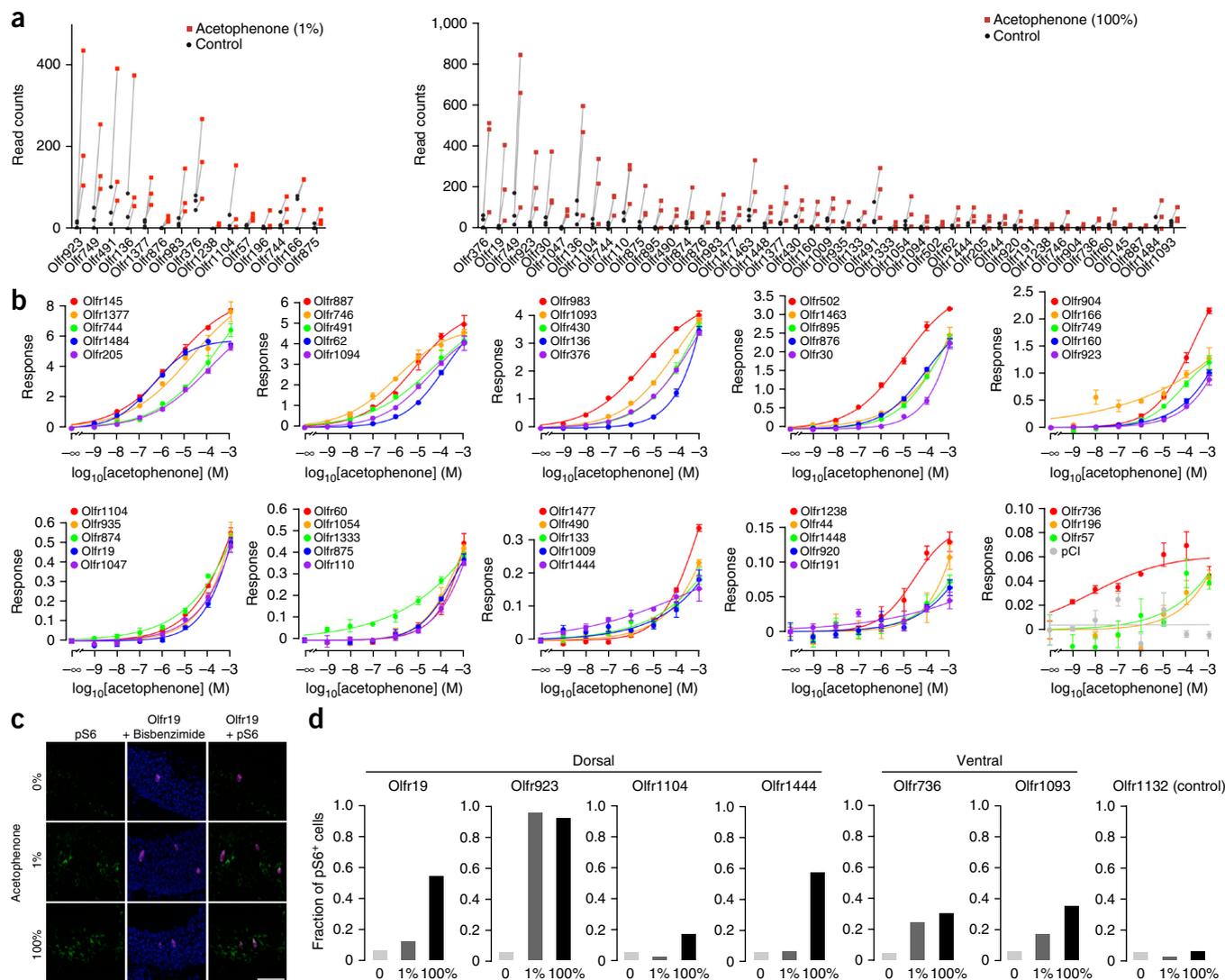


Figure 4 Identification of acetophenone receptors. **(a)** Left, read counts of acetophenone receptors in control and 1% acetophenone–stimulated samples ($n = 3$ pairs). Right, read counts of acetophenone receptors in control and 100%–stimulated samples ($n = 3$ pairs). **(b)** *In vitro* responses of the acetophenone receptors. Responses were scaled to Olfr160 (M72). The maximum response of Olfr160 was defined as 1. **(c)** Coronal section of OE following acetophenone stimulation for 1 h. Green indicates antibody staining for pS6, magenta indicates RNA FISH for the newly identified acetophenone OR Olfr19 and blue indicates bisbenzimidazole staining showing the nuclei. Scale bar represents 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.

Hana3A cells⁹ and challenged the cells with 10^{-9} to 10^{-3} M acetophenone. The dose-dependent luciferase response profiles confirmed that 48 ORs were indeed activated by acetophenone (**Fig. 4a,b**). Of the 75 ORs enriched in 100% acetophenone experiment ($P < 0.05$), 71 were tested *in vitro* and 45 (63%) responded. Of the 25 ORs enriched in 1% acetophenone experiment ($P < 0.05$), 24 were tested *in vitro* and 15 (63%) responded. Of the 15 confirmed acetophenone ORs identified for 1% acetophenone, 12 (80%) were also enriched by 100% acetophenone, whereas the other three did not reach statistical significance, although they showed trends toward enrichment. The 1% and 100% stimulation experiments thus identified a total of 48 acetophenone receptors (**Supplementary Table 2**). We reasoned 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 that 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 showed no *in vitro* response or enrichment in pS6-IP. Indeed, acetophenone exposure resulted in pS6 induction in OSNs expressing identified acetophenone ORs as compared with the control OR (**Fig. 4c,d** and **Supplementary Fig. 3**).

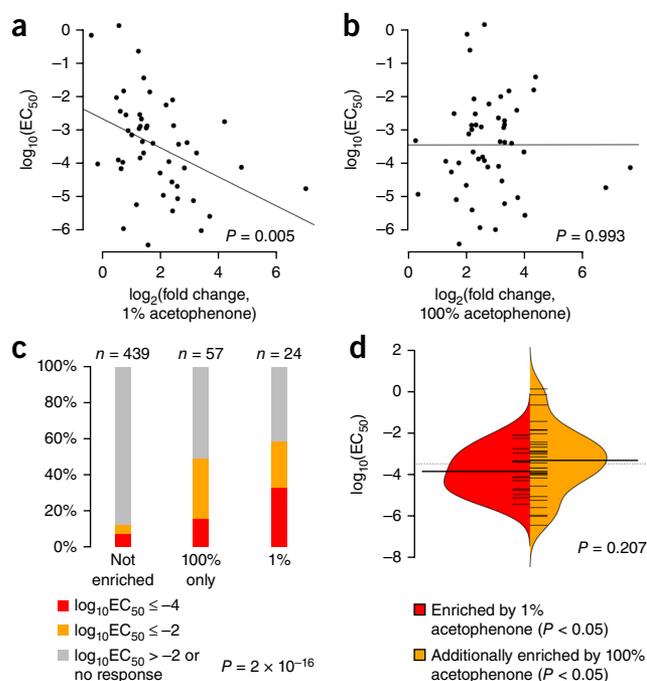
Although the binary responses of *in vivo* and *in vitro* results correlate well, the extent to which the pS6 method captures the information of OR sensitivity is not clear. To address this question, we plotted the *in vitro* EC_{50} values against the *in vivo* fold of transcript enrichment for the 48 acetophenone ORs. Indeed, EC_{50} 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

Figure 5 Correlation between *in vivo* and *in vitro* sensitivities of ORs. (a) Correlation between EC_{50} 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 EC_{50} 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 were relatively more sensitive ($\log_{10}EC_{50} \leq -4$), moderately sensitive ($\log_{10}EC_{50}$ between -2 and -4) and not sensitive ($\log_{10}EC_{50} > -2$) 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 $\log_{10}EC_{50}$ values in ORs only enriched by 100% acetophenone as compared with those enriched by 1% acetophenone. $P = 0.207$, Wilcoxon rank-sum test.

regression and ANOVA, Spearman's rho = 0.00137 ; **Fig. 5b**), presumably as a result of saturation of the olfactory signaling pathways in OSNs following high-intensity odor stimulation. Consistent with this, the set of ORs enriched by 1% acetophenone tended to contain a higher fraction of more sensitive ORs as compared with ORs not enriched or enriched only by 100% acetophenone ($P < 2 \times 10^{-16}$, Chi-square test; **Fig. 5c**). In addition, the median EC_{50} 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, although pS6-IP with high odorant concentrations is useful for 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 whether these ORs have greater overall homology with each other. To our surprise,



when we plotted the acetophenone ORs on the OR phylogenetic tree, they were not clustered in one or a few subfamilies, although there were a few notable examples of closely related ORs (**Fig. 6a**). Rather, their sequences were markedly diverse: at least one member of 29 of the 286 receptor families defined previously²⁸ 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 distances²⁹ 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

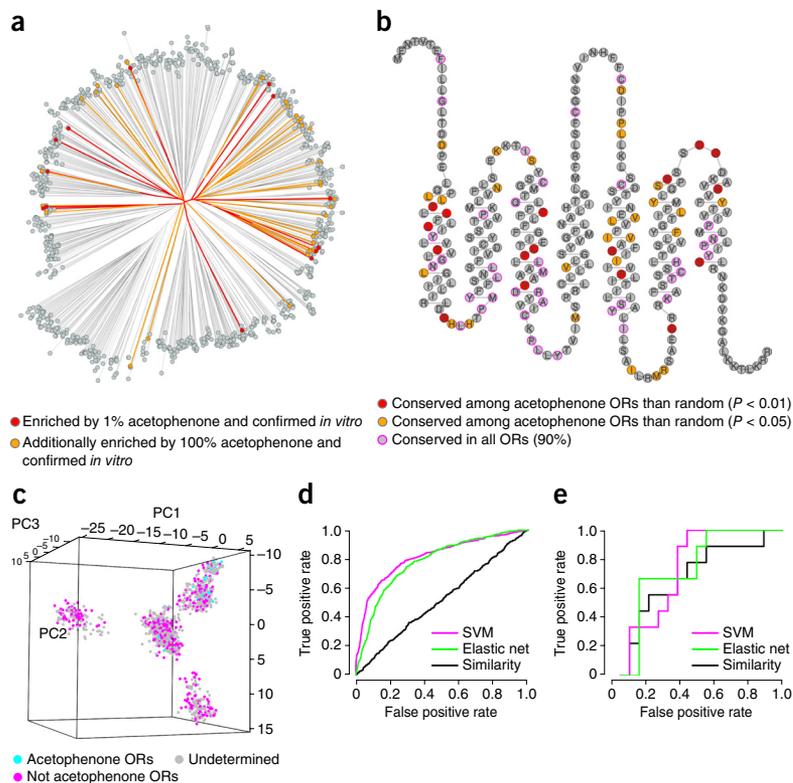


Figure 6 Sequence-function analysis of the identified acetophenone receptors. (a) Distance

tree of OR protein sequences. Red indicates ORs enriched by 1% acetophenone and confirmed *in vitro* and orange indicates ORs additionally enriched by 100% acetophenone and confirmed *in vitro*. (b) Amino acid residues that were more conserved in the acetophenone receptors than random OR sets of the same size. Red, $P < 0.01$. Orange, $P < 0.05$.

(c) Principal component analysis of amino acid properties of mouse ORs. Shown is a plot of the first three principal components (variance explained: 6.8%, 3.7%, 2.6%). Cyan represents acetophenone ORs, magenta represents ORs that did not respond to acetophenone both *in vivo* and *in vitro*, and gray represents 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.

Table 1 Performance of models in tenfold cross-validation and external validation

	Cross-validation		External validation	
	AUC	<i>P</i> value	AUC	<i>P</i> 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

ORs than would be expected by random chance ($n = 44$, $P < 0.05$, $n = 17$, $P < 0.01$, FDR adjusted; **Fig. 6b** and **Supplementary Fig. 4**; see Online Methods for details). Notably, these sites are not highly conserved among ORs as a whole and the majority of them are located in 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 principal component analysis on the amino acid properties²⁹ of ORs identified four clusters using the first three principal 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 tenfold cross-validation and external validation schemes. The data for modeling contained the 48 ORs that 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 community³⁰. 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 penalty³¹. 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 of the models were tested in a tenfold cross-validation scheme. In addition, we also obtained the *in vitro* acetophenone response profiles of 27 human ORs and used this separate data set for external validation of models (**Supplementary Fig. 5**). These human ORs were 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 ($P = 10^{-120}$, $P = 3 \times 10^{-97}$) and external ($P = 0.04$, $P = 0.04$) validation and correctly predicting responsiveness to acetophenone 71.6–81.5% of the time. In contrast, the overall similarity based approach did not predict responsiveness to acetophenone with statistical significance ($P = 0.1$, $P = 0.1$; **Fig. 6d,e** and **Table 1**).

Identification of TMT ORs

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 TMT, a fox feces component that induces fear-related responses in predator-naive 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 h and performed pS6-IP on the dissected OE ($n = 4$ pairs). TMT stimulation led to enrichment of 43 ORs ($P < 0.05$). 1% TMT enriched 4 ORs ($P < 0.05$), a subset of the 43 100% TMT-enriched ORs (**Fig. 7a** and **Supplementary Fig. 6**). We cloned 42 of these ORs, and 21 of them displayed significant responses to TMT when expressed *in vitro* (**Fig. 7b**). Thus, we found strong evidence that these 21 ORs are indeed TMT receptors (**Supplementary Table 2**). Somewhat unexpectedly, six of the TMT ORs were also

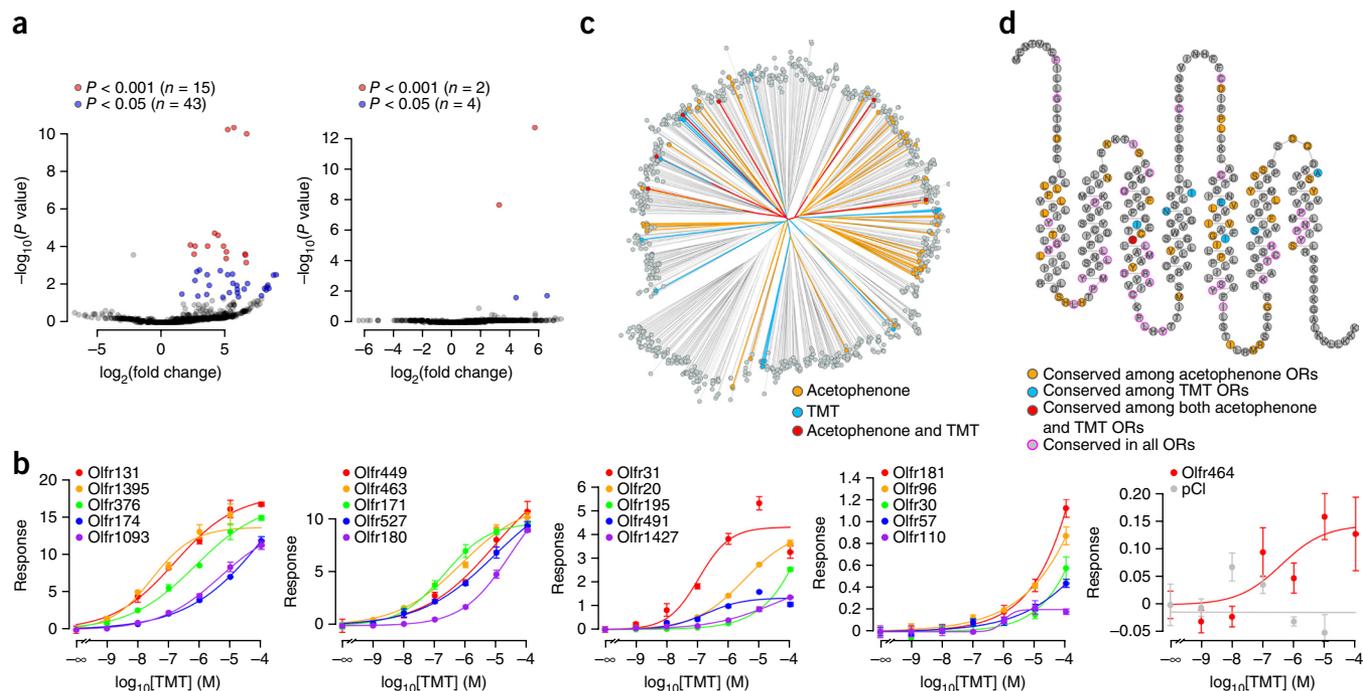


Figure 7 Identification of ORs activated by TMT. **(a)** Volcano plot showing enrichment of OR mRNA in the 100% TMT-stimulated group (left) and 1% acetophenone-stimulated group (right). **(b)** *In vitro* responses of the TMT receptors. Error bars indicate s.e.m. **(c)** Distance tree of OR protein sequences comparing acetophenone and TMT ORs. Orange represents ORs activated by acetophenone, blue represents ORs activated by TMT, and red represents ORs activated by both acetophenone and TMT. **(d)** Amino acid residues that were conserved in the acetophenone receptors (orange, $P < 0.05$), TMT receptors (blue, $P < 0.05$) and both (red). The amino acid residues that were conserved in all ORs are labeled for comparison (90%, magenta circles).

identified as acetophenone receptors. Similar to acetophenone ORs, the TMT ORs were also diverse in sequence (Fig. 7c). Similar to the majority of conserved residues among acetophenone ORs, the conserved residues among TMT ORs were also located in transmembrane helices, and may have been involved in ligand binding (Fig. 7d). Five of these TMT ORs (Olfr20, Olfr30, Olfr57, Olfr376 and Olfr491) were expressed in the dorsal OE, a region implicated in mediating the fear-related behavioral responses of TMT³³ (Supplementary Fig. 6e). These ORs represent candidates for future investigation to understand the basis of TMT-induced behaviors. Together, our data suggest that pS6-IP in combination with RNA-Seq can be used to identify active ORs for diverse odorants.

DISCUSSION

One odorant activates a certain subset of ORs, but which ORs are activated *in vivo* is largely unclear. Although the search for OR ligands *in vivo* has been facilitated by transgenic animals that coexpress fluorescent proteins with the ORs of interest⁸, these animal resources are limited to several well-studied ORs. We developed a new approach 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, experiments can be performed in awake, freely behaving animals; thus, the results obtained are more likely to be physiologically relevant. Second, 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. A recent study reported a method called the Kentucky assay to deorphanize ORs using the reporter S100a5 to drive GFP expression in activated OSNs, followed by microarray analysis of sorted GFP⁺ cells²⁴. 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 the ORs that are the most active, whereas our pS6-IP method enables the large-scale identification of ORs for a given odor. As compared with S100a5-tauGFP, pS6 induction showed faster onset (30 min to 1 h versus 14 h of odor stimulation time). In addition, pS6-IP does not require specific transgenic animals, which may enable researchers using non-model species to identify ORs activated by odorants of interest.

In this issue, Von der Weid *et al.*³⁴ report a method that takes advantage of the decrease in OR transcript abundance following odor stimulation to identify ORs responding to acetophenone and ethyl isobutyrate³⁴. As compared with pS6-IP, this method involves fewer steps and only requires working with whole RNA from OE, whereas pS6 induction occurs much faster than downregulation of OR transcripts. We examined the set of acetophenone ORs ($n = 22$) identified by von der Weid *et al.*³⁴ in our pS6-IP RNA-Seq and heterologous expression data. 20 of 22 (91%) ORs showed positive log fold change at least one concentration, of which 18 (81%) ORs were significantly enriched ($P < 0.05$, FDR adjusted for 22 ORs) at least in one concentration. In addition, 15 of 17 (88%) ORs showed *in vitro* activation. Overall, 21 of 22 (95%) ORs were supported by at least one of the criteria in our data set (Supplementary Table 3). Quantification by generating an ROC curve revealed that enrichment in pS6-IP RNA-Seq predicted the ORs identified by von der Weid *et al.*³⁴ ($AUC = 0.86$, $P = 4 \times 10^{-9}$, Wilcoxon rank-sum test, one tailed against H_0 : classifier performs no better than random; Supplementary Fig. 7). Together, our data support the findings of von der Weid *et al.*³⁴ as groups. The different stimulation conditions, different signaling pathways and kinetics underlying these 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 were 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 will be necessary to identify the signaling pathway leading to S6 phosphorylation. Our quantification of pS6 immunostaining signal intensity suggests that both ligand concentrations and ligand affinities for each OR influence phosphorylation levels of S6. Although 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*. Although 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 coexpressed 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 ~10,000,000 reads per 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 method represents a powerful strategy for identifying 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 to 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 mucus³⁷. 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 cases³⁸. Our unbiased mapping of ORs activated at different concentration of odors revealed the identities of more sensitive ORs as well as additional ORs activated by a higher odor concentration in freely behaving mice (Figs. 2 and 7). Our approach could allow future studies to address the roles of each of the active ORs in odor detection and perception.

We found that acetophenone and TMT activated a large set of 48 and 21 ORs, respectively. Six ORs were activated by both odorants, raising the possibility that these ORs are broadly tuned, which could be determined in the future 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 markedly.

Although the overall sequences of acetophenone and TMT ORs are diverse, we identified residues that are similar in 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, principal component analysis 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 to 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 cheminformatics studies⁴¹ 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.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Raw reads and quantification results can be accessed at GEO: [GSE59324](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank G. Barnea and R. Axel for antibodies to M72, H. Zhuang for some of the OR constructs, D. Marchuk for sharing of equipment, the Duke IGSP Genome Sequencing and Analysis Core Resource for providing Illumina sequencing service, and S. Mukherjee and J. Mainland for helpful discussions and comments on the manuscript. This work was supported by grants from the US National Institutes of Health (DC012095 and DC014423).

AUTHOR CONTRIBUTIONS

H.M. supervised all of the 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.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Nimura, 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 Res.* **24**, 1485–1496 (2014).
- Malnic, B., Hirono, J., Sato, T. & Buck, L.B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
- Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J.D. Odor coding by a mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).
- Peterlin, Z., Firestein, S. & Rogers, M.E. The state of the art of odorant receptor deorphanization: a report from the orphanage. *J. Gen. Physiol.* **143**, 527–542 (2014).

- Zhao, H. *et al.* Functional expression of a mammalian odorant receptor. *Science* **279**, 237–242 (1998).
- Touhara, K. *et al.* Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc. Natl. Acad. Sci. USA* **96**, 4040–4045 (1999).
- Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22**, 3033–3043 (2002).
- Saito, H., Kubota, M., Roberts, R.W., Chi, Q. & Matsunami, H. RTP family members induce functional expression of mammalian odorant receptors. *Cell* **119**, 679–691 (2004).
- Araneda, R.C., Kini, A.D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nat. Neurosci.* **3**, 1248–1255 (2000).
- Grosmaître, X. *et al.* SR1, a mouse odorant receptor with an unusually broad response profile. *J. Neurosci.* **29**, 14545–14552 (2009).
- Mainland, J.D. *et al.* The missense of smell: functional variability in the human odorant receptor repertoire. *Nat. Neurosci.* **17**, 114–120 (2014).
- Keller, A., Zhuang, H., Chi, Q., Vosshall, L.B. & Matsunami, H. Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–472 (2007).
- Menashe, I. *et al.* Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol.* **5**, e284 (2007).
- Shirasu, M. *et al.* Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron* **81**, 165–178 (2014).
- Dong, P.Y., Gong, N.N. & Matsunami, H. Cell-based system for identification of olfactory receptors. in *Bioelectronic Nose* (ed. T.H. Park) 83–95 (Springer Netherlands, 2014).
- Knight, Z.A. *et al.* Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* **151**, 1126–1137 (2012).
- Haga, S. *et al.* The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. *Nature* **466**, 118–122 (2010).
- Isogai, Y. *et al.* Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241–245 (2011).
- Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
- Zhang, J., Huang, G., Dewan, A., Feinstein, P. & Bozza, T. Uncoupling stimulus specificity and glomerular position in the mouse olfactory system. *Mol. Cell. Neurosci.* **51**, 79–88 (2012).
- Oka, Y. *et al.* Odorant receptor map in the mouse olfactory bulb: *in vivo* sensitivity and specificity of receptor-defined glomeruli. *Neuron* **52**, 857–869 (2006).
- Krautwurst, D., Yau, K.W. & Reed, R.R. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95**, 917–926 (1998).
- McClintock, T.S. *et al.* *In vivo* identification of eugenol-responsive and muscone-responsive mouse odorant receptors. *J. Neurosci.* **34**, 15669–15678 (2014).
- Meister, M. & Bonhoeffer, T. Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.* **21**, 1351–1360 (2001).
- Zhuang, H. & Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat. Protoc.* **3**, 1402–1413 (2008).
- Li, Y.R. & Matsunami, H. Activation state of the M3 muscarinic acetylcholine receptor modulates mammalian odorant receptor signaling. *Sci. Signal.* **4**, ra1 (2011).
- Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5**, 124–133 (2002).
- Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
- Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, 2006).
- Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).
- Takahashi, L.K., Nakashima, B.R., Hong, H. & Watanabe, K. The smell of danger: a behavioral and neural analysis of predator odor-induced fear. *Neurosci. Biobehav. Rev.* **29**, 1157–1167 (2005).
- Kobayakawa, K. *et al.* Innate versus learned odor processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
- Weid, B. *et al.* *In vivo* identification of receptor-ligand pairs via large-scale transcriptional profiling of chemosensory neurons. *Nat. Neurosci.* doi:10.1038/nn.4100 (31 August 2015).
- Thiebaud, N. *et al.* Odorant metabolism catalyzed by olfactory mucosal enzymes influences peripheral olfactory responses in rats. *PLoS ONE* **8**, e59547 (2013).
- Nagashima, A. & Touhara, K. Enzymatic conversion of odorants in nasal mucus affects olfactory glomerular activation patterns and odor perception. *J. Neurosci.* **30**, 16391–16398 (2010).
- Scott, J.W., Sherrill, L., Jiang, J. & Zhao, K. Tuning to odor solubility and sorption pattern in olfactory epithelial responses. *J. Neurosci.* **34**, 2025–2036 (2014).
- Grossisseroff, R. & Lancet, D. Concentration-dependent changes of perceived odor quality. *Chem. Senses* **13**, 191–204 (1988).
- Li, J., Haddad, R., Chen, S., Santos, V. & Luetje, C.W. A broadly tuned mouse odorant receptor that detects nitrotoluenes. *J. Neurochem.* **121**, 881–890 (2012).
- Nara, K., Saraiva, L.R., Ye, X. & Buck, L.B. A large-scale analysis of odor coding in the olfactory epithelium. *J. Neurosci.* **31**, 9179–9191 (2011).
- Boyle, S.M., McInally, S. & Ray, A. Expanding the olfactory code by *in silico* decoding of odor-receptor chemical space. *eLife* **2**, e01120 (2013).



ONLINE METHODS

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. 3–4-week-old C57BL/6 mice (Jackson Labs) were placed individually into sealed containers (volume \approx 2.7 l) inside a fume hood and allowed to rest for 1 h in an odorless environment. For odor stimulus, 10 μ l of odor solution or 10 μ l of distilled water (control) was applied to 1-cm \times 1-cm 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 h. 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 modifications¹⁷. The pS6 240-containing peptide that was used by Knight *et al.*¹⁷ 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 killed and the OE was dissected in 25 ml of dissection buffer (1 \times HBSS (Gibco, with Ca²⁺ and Mg²⁺), 2.5 mM HEPES (pH 7.4 adjusted with KOH), 35 mM glucose, 100 μ g ml⁻¹ cycloheximide, 5 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium pyrophosphate, 1 mM beta-glycerophosphate) on ice. The dissected OE was transferred to 1.35 ml of homogenization buffer (150 mM KCl, 5 mM MgCl₂, 10 mM HEPES (pH 7.4 adjusted with KOH), 100 nM Calyculin A, 2 mM DTT, 100 U ml⁻¹ RNasin (Promega), 100 μ g ml⁻¹ cycloheximide, 5 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, protease inhibitor (Roche, 1 tablet per 10 ml)) and homogenized three times at 250 rpm and nine times at 750 rpm (Glas-Col). The homogenate was transferred to a 1.5-ml lobind tube (Eppendorf), and centrifuged at 2,000g for 10 min at 4 °C. The supernatant was then transferred to a new 1.5-ml lobind tube, to which 90 μ l 10% NP-40 (vol/vol) and 90 μ l 300 mM DHPC (Avanti Polar Lipids) were added. The mixture was centrifuged at 17,000g for 10 min at 4 °C. The supernatant was transferred to a new 1.5-ml lobind tube, and mixed with 20 μ l pS6 antibody (Cell Signaling, #2215). Antibody binding was allowed by incubating the mixture for 1.5 h at 4 °C with rotation. During antibody binding, Protein A Dynabeads (Invitrogen, 100 μ l per sample) was washed three times with 900 μ l beads wash buffer 1 (150 mM KCl, 5 mM MgCl₂, 10 mM HEPES (pH 7.4 adjusted with KOH), 0.05% BSA (wt/vol), 1% NP-40). After antibody binding, the mixture was added to the washed beads and gently mixed, followed by incubation for 1 h at 4 °C with rotation. After incubation, the RNA-bound beads were washed four times with 700 μ l beads wash buffer 2 (RNase-free water containing 350 mM KCl, 5 mM MgCl₂, 10 mM HEPES (pH 7.4 adjusted with KOH), 1% NP-40, 2 mM DTT, 100 U ml⁻¹ recombinant RNasin (Promega), 100 μ g ml⁻¹ 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 μ l 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 described⁹ with modifications. Briefly, 1.5 μ l purified RNA was mixed with 5 μ l reaction mix (1 \times PCR buffer (Roche), 1.5 mM MgCl₂, 50 μ M dNTPs, 2 ng μ l⁻¹ poly-T primer (TAT AGA ATT CGC GGC CGC TCG CGA TTT TTT TTT TTT TTT TTT TTT), 0.04 U μ l⁻¹ RNase inhibitor (Qiagen), 0.4 U μ l⁻¹ recombinant RNasin (Promega)). This mixture was heated at 65 °C for 1 min and cooled to 4 °C. 0.3 μ l RT mix (170 U μ l⁻¹ Superscript II (Invitrogen), 0.4 U μ l⁻¹ RNase inhibitor (Qiagen), 4 U μ l⁻¹ recombinant RNasin (Promega), 3 μ g μ l⁻¹ T4 gene 32 protein (Roche)) was added to each tube and incubated at 37 °C for 10 min then 65 °C for 10 min. 1 μ l ExoI mix (2 U μ l⁻¹ ExoI (NEB), 1 \times PCR buffer (Roche), 1.5 mM MgCl₂) was added to each tube and incubated at 37 °C for 15 min then 80 °C for 15 min. 5 μ l TdT mix (1.25 U μ l⁻¹ TdT (Roche), 0.1 U μ l⁻¹ RNase H (Invitrogen), 1 \times PCR buffer (Roche), 3 mM dATP, 1.5 mM MgCl₂)

was added to each tube and incubated at 37 °C for 20 min then 65 °C for 10 min. 3.5 μ l of the product was added to 27.5 μ l PCR mix (1 \times LA Taq reaction buffer (TaKaRa), 0.25 mM dNTPs, 20 ng μ l⁻¹ poly-T primer, 0.05 U μ l⁻¹ LA Taq (TaKaRa)) and incubated at 95 °C for 2 min, 37 °C for 5 min, 72 °C for 20 min, then 16 cycles of 95 °C for 30 s, 67 °C for 1 min, 72 °C for 3 min with 6-s extension for each cycle, and then 72 °C for 10 min. 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](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59324).

Reads mapping and test for enrichment. Short reads were aligned to the mouse reference genome mm10 using Bowtie⁴². The reads mapped to annotated genes were then counted using BEDTools⁴³. 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 regions⁴⁴. 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 EdgeR⁴⁵. A similar result was obtained using a separate tool DESeq2 employing similar model assumptions⁴⁶. Littermate information was included in the design matrix to enhance the power of detection⁴⁷. Multiple comparison was adjusted within the detected OR group by controlling FDR⁴⁸. DESeq was used to calculate the size factors of individual libraries⁴⁶.

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 ORs²³.

Cell culture. Hana3A cells⁹ were maintained in minimal essential medium (MEM) containing 10% FBS (vol/vol) with penicillin-streptomycin and amphotericin B, at 37 °C and 5% CO₂.

Dual-Glo assay. The Dual-Glo assay for ORs was performed as described previously²⁶. Briefly, Hana3A cells were plated on 96-well plates. 18–24 h after plating, cells were transfected with plasmids coding for OR, M3-R, RTP1S, CRE-luciferase and pRL-SV40. 18–24 h later, cells were stimulated by incubation with odorants diluted in CD293 for 4 h at 37 °C and 5% CO₂ 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 (Olfir1126) 100. Data was analyzed using GraphPad Prism 5 or custom R scripts. The cut-off to determine positive or negative *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 300 μ M 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. 1,090 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 20,000 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 FDR⁴⁸. Visualization of residue frequencies was generated by WebLogo⁴⁹.

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 Grantham²⁹ 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 package⁵⁰. SVM was performed using the e1071 package. A tenfold 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- μ m frozen sections of the olfactory epithelium and surrounding tissues of 3-week-old mice were incubated with rabbit anti-phospho-S6 (240/244) (Cell Signaling, 1:200) overnight at 4 °C. Following washing, donkey Cy3-conjugated anti-rabbit IgG (Jackson Immunologicals, 1:200) 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 described⁹. Briefly, digoxigenin (DIG)-labeled complementary RNA probes were hybridized overnight at 58 °C to target mRNAs in 20- μ m 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 (PerkinElmer) 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- μ m intervals between each slice, were obtained at 200 \times 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 bisbenzimidide. 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

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 \times background signal)

A **Supplementary Methods Checklist** is available.

- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
- Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
- McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
- Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).