Variation in olfactory neuron repertoires is genetically controlled and environmentally modulated

Ximena Ibarra-Soria[1], Thiago S. Nakahara[2], Jingtao Lilue[1], Yue Jiang[3], Casey Trimmer[4], Mateus A. A. Souza[2], Paulo H. M. Netto[2], Kentaro Ikegami[3], Nicolle R. Murphy[4], Mairi Kusma[1], Andrea Kirton[1], Luis R. Saraiva[1], Thomas M. Keane[1], Hiroaki Matsunami[3,5], Joel D. Mainland[4,6], Fabio Papes[2], Darren W. Logan[1,4,7].

[1]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge, CB10 1SA, UK.

[2]Department of Genetics and Evolution, Institute of Biology, University of Campinas, Rua Monteiro Lobato, Campinas, SP 13083-862, Brazil.

[3]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, 27710, USA.

[4]Monell Chemical Senses Center, Philadelphia, Pennsylvania, 19104, USA.

[5]Department of Neurobiology, Duke Institute for Brain Sciences, Duke University Medical Center, Durham, North Carolina, 27710, USA

[6]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA

[7]Corresponding author

Darren W. Logan
Wellcome Trust Sanger Institute,
Wellcome Genome Campus,
Hinxton, Cambridge, CB10 1SA, United Kingdom.
Tel: +44 1223 496854
E-mail: dl5@sanger.ac.uk

SUMMARY

The mouse olfactory sensory neuron (OSN) repertoire is composed of 10 million cells and each expresses one olfactory receptor (OR) gene from a pool of over 1000. Thus, the nose is sub-stratified into more than a thousand OSN subtypes. Here, we employ and validate an RNA-sequencing based method to quantify the abundance of all OSN subtypes in parallel, and investigate the genetic and environmental factors that contribute to neuronal diversity. We find that the OSN subtype distribution is stereotyped in genetically identical mice, but varies extensively between different strains. Further, we identify *cis*-acting genetic variation as the greatest component

influencing OSN composition and demonstrate independence from OR function. However, we show that olfactory stimulation with particular odorants results in modulation of dozens of OSN subtypes in a subtle but reproducible, specific and time-dependent manner. Together, these mechanisms generate a highly individualized olfactory sensory system by promoting neuronal diversity.

KEYWORDS

Olfactory receptor, RNAseq, variation, genetic regulation, environmental regulation.

INTRODUCTION

Mapping the neuronal diversity within a brain remains a fundamental challenge of neuroscience. Quantifying variance in a population of neurons within and between individuals first requires precise discrimination of cellular subtypes, followed by an accurate method of counting them. While this has been achieved in a simple invertebrate model containing hundreds of neurons (White et al., 1986), applying the same approach to mammalian brains that encompass many millions of neurons represents a significant challenge (Wichterle et al., 2013).

The main olfactory epithelium (MOE) is an essential component of the olfactory sensory system. It contains olfactory sensory neurons (OSNs) that express olfactory receptors (ORs), the proteins that bind odorants (Buck and Axel, 1991, Zhao et al., 1998). The mouse genome codes for over a thousand functional OR genes, but each mature OSN expresses only one abundantly, in a monoallelic fashion (Hanchate et al., 2015, Saraiva et al., 2015b, Tan et al., 2015, Chess et al., 1994). This results in a highly heterogeneous repertoire of approximately 10 million OSNs (Kawagishi et al., 2014) within the nose of a mouse, stratified into more than a thousand functionally distinct subpopulations, each one characterized by the particular OR it expresses. The monogenic nature of OR expression serves as a molecular barcode for OSN subtype identity. Thus, the MOE offers a unique opportunity to generate a comprehensive neuronal map of a complex mammalian sensory organ, and investigate the mechanisms that influence its composition and maintenance.

To date only a few studies have quantified the number of OSNs that express a given OR (Bressel et al., 2016, Fuss et al., 2007, Khan et al., 2011, Rodriguez-Gil et al.,

2010, Royal and Key, 1999, Young et al., 2003). For the scarce data available (<10% of the full repertoire) reproducible differences in abundance have been observed between OSNs expressing different ORs (Bressel et al., 2016, Fuss et al., 2007, Khan et al., 2011, Young et al., 2003). This suggests variance in the representation of OSN subtypes exists within an individual, but the extent of variation between individuals is unknown. Moreover, the mechanisms that dictate the abundance of OSN subtypes are poorly understood. Most promoters of OR genes contain binding sites for Olf/Ebf1 (O/E) and homeodomain (HD) transcription factors (Young et al., 2011) and these are involved in determining the probability with which the OR genes are chosen for expression (Rothman et al., 2005, Vassalli et al., 2011). Enhancer elements also regulate the gene choice frequencies of nearby, but not distally located, ORs (Khan et al., 2011). To date these studies have focused only on a handful of OSN subtypes.

In addition to OR gene choice regulation exerted by genetic elements, it is conceivable that the olfactory system adapts to the environment. The MOE is continually replacing its OSN pool and the birth of every neuron presents an opportunity to shape the proportion of different subpopulations. It is also possible that relative OSN abundances could be altered by regulating the lifespan of each OSN subtype. Indeed activation extends a sensory neuron's life-span (Santoro and Dulac, 2012), suggesting that persistent exposure to particular odorants may, over time, increase the relative proportions of the OSNs responsive to them. Some OSN subtypes do reportedly increase in number in response to odor activation, but others do not (Cadiou et al., 2014, Cavallin et al., 2010, Watt et al., 2004). Whether this variation reflects differences in the biology of OSN subtypes or experimental procedures is unclear.

Here we fully map OSN diversity in the MOE and characterize the influence of both genetic and environmental factors on its regulation. We show that RNA sequencing (RNAseq) is an accurate proxy for measuring the number of OSNs expressing a particular OR type, and use this approach to quantify, in parallel, the composition of 1,115 OSN subtypes in the MOE. We report that, while the repertoire of OSN subtypes is stable across individuals from the same strain, it reproducibly and extensively differs between genetically divergent strains of laboratory mice. We show that under controlled environmental conditions these stereotypic differences in OSN

abundance are directed by genetic variation within regulatory elements of OR genes that act in *cis*, and are independent of the sequence or function of the OR protein. However, we find that persistent, but not continuous, exposure to specific odorants can also subtly alter abundance of the OSN subtypes that are responsive to such stimuli. Taken together, these results show that the OSN repertoire is shaped by both genetic and environmental influences to generate a unique nose for each individual.

RESULTS

Olfactory Sensory Neuron Diversity Measured by RNAseq.

Previously, we characterized the transcriptional profile of the whole olfactory mucosa (WOM) in adult C57BL/6J animals (hereafter termed B6) to generate hundreds of new, extended OR gene annotations (Ibarra-Soria et al., 2014). As each OR gene is expressed in only a small fraction of cells within WOM, differences in their abundance are difficult to distinguish from sampling bias. We hypothesized that mapping RNAseq data to significantly extended OR transcripts should increase detection sensitivity. With these models, OR gene mRNA level estimates in adult WOM increase, on average, 2.3-fold, but some increase almost 20-fold (**Figure S1A**). Despite this improvement, most OR mRNAs still have relatively low expression values (**Figure 1A**). Nevertheless, they show a dynamic range of abundance levels (**Figure 1A, inset**) that are consistent between biological replicates, as indicated by their very high correlation values (median rho = 0.89, $p < 2.2 \times 10^{-16}$).

To assess whether these low OR mRNA expression values are biologically meaningful or if they represent low-level technical artifacts of RNAseq analysis, we sequenced RNA from WOM of a mouse strain, *Δ*Olfr7Δ*, that has a targeted homozygous deletion of the *Olfr7* OR gene cluster on chromosome 9 (Xie et al., 2000, Khan et al., 2011), and compared their gene expression to control mice. From the 94 OR genes of the cluster that have been deleted, 83 (88.3%) have no counts in any of the three biological replicates. The 11 remaining genes have just one or two fragments mapped in only one of the replicates **(Figure 1B)**, resulting in normalized counts of less than 0.4. In contrast, the control mice have from 14.2 to 498.1 normalized counts for the same genes. Together these experiments demonstrate that the use of extended gene models significantly increases the sensitivity to detect OR

mRNA expression in WOM, and that the full dynamic range of abundances reflects true measures of OR gene expression.

The wide range of stereotypic OR gene expression can be explained by two scenarios, acting alone or in combination (**Figure 1C**): either 1) OR genes with high expression values are monogenically expressed in more OSNs than those with low expression values; and/or 2) OR genes are consistently expressed at different levels per OSN. To differentiate between these possibilities, we performed *in situ* hybridization (ISH) of probes specific to 6 OR genes with expression values distributed across the dynamic range. We then counted the number of OSNs in which each OR is expressed (**Figure 1D)**. We find a perfect correlation between OSN number and RNAseq expression value (rho = 1, p = 0.0028). We additionally compared OR gene RNAseq expression levels with three independent measures of the number of OSNs expressing the same ORs (Bressel et al., 2016, Fuss et al., 2007, Khan et al., 2011). In all three cases we find high correlations (**Figure S1B-C**). We next collected 63 single mature OSNs from WOM, and determined the OR gene most abundantly expressed in each using a single-cell RNAseq approach (Saraiva et al., 2015b). If OR expression levels in WOM reflect the proportion of OSNs that express each receptor (**Figure 1C**), the probability of isolating each OSN type is not equal. Indeed, we find a strong selection bias towards OSNs that express OR genes with high RNAseq levels in WOM (hypergeometric test, $p = 6.44 \times 10^{-9}$; **Figure 1E**), suggesting those OSN types are more numerous in the olfactory epithelium. Thus, consistent with a recent analysis in zebra fish (Saraiva et al., 2015a), OR mRNA levels are an accurate measure of the number of each OSN subtype in the mouse WOM (scenario 1). But do consistent differences in OR mRNA levels per cell also contribute (scenario 2)? To test this we quantified the mRNA levels of the most abundant OR gene in each of the 63 single, mature OSNs, normalized to three stably expressed OSN marker genes (Khan et al., 2013). We find OR mRNA levels vary within the single cells, but this does not correlate with expression levels across the WOM (rho = -0.04, p = 0.7518) (**Figure 1F**). Analysis of ERCC spike-ins confirmed that the levels of OR mRNAs in single OSNs are reliable. Moreover, the single OSN transcript levels also positively correlate with transcript levels in pools of millions of OSNs (Saraiva et al., 2015b), demonstrating they reliably reflect the transcriptome within single cells. Together

these data demonstrate that OR mRNA levels obtained by RNAseq are an accurate proxy for quantifying the diversity of OSN subtypes that express each receptor.

The OSN Repertoire Differs Between Strains of Mouse.

The relative proportion of each OSN subtype is stable between genetically identical animals. We have previously reported the expression of OR genes in B6 male and female mice (Ibarra-Soria et al., 2014). By applying full gene models to these data, here we confirm their OSN distribution profiles are remarkably similar (**Figure 2A**); only 1.2% of the OR gene repertoire is significantly differentially expressed (**Figure 2B**). To investigate whether this OSN distribution is a stereotypic feature of the species, we next reconstructed the WOM transcriptome of a different laboratory strain, 129S5SvEv (hereafter termed 129). The 129 genome has 4.4 million single nucleotide polymorphisms (SNPs) and 0.81 million small indels compared to B6 (Keane et al., 2011), of which we find 13,484 SNPs and 1,936 indels within our extended OR gene transcripts. As OR genes are particularly variable in coding sequence between strains of mice (Logan, 2014), mapping RNAseq data from other strains to a B6 reference genome results in biases in OR gene expression estimates (**Figure S2A**). We therefore generated a pseudo-129 genome on which to map the RNAseq data, by editing the reference genome at all polymorphic sites. From the 1,249 OR genes, we find 462 are significantly differentially expressed (DE) compared to B6 (FDR < 5%), representing 37% of the whole repertoire (**Figure 2C,D**).

To determine whether greater genetic diversity influences the variance in OSN repertoire, we repeated this experiment using a wild-derived strain from the *Mus musculus castaneus* subspecies (CAST/EiJ, henceforth CAST). This strain has more than 17.6 million SNPs and 2.7 million indels relative to B6 (Keane et al., 2011); of these, we calculated that 45,688 SNPs and 6,303 indels are found within our extended OR transcripts. After mapping to a pseudo-CAST genome (**Figure S2B**), 634 OR genes are significantly differentially expressed (FDR < 5%) compared to B6, constituting 50.8% of the whole OR repertoire (**Figure 2E,F**). The changes in expression for some OR genes are dramatic: 132 genes have differences of at least 8-fold. Taking all pairwise comparisons into account (including 129 vs CAST, **Figure S2C,D**), 821 OR genes (65.7%) are DE between at least two strains. 136 of these are DE in all three pairwise comparisons (**Figure 2G**); for example there are consistently

different numbers of *Olfr6*-expressing OSNs in each strain (**Figure 2H)**.

To determine if the DE OR genes between strains reflect differences in the number of OSN subtypes, we performed ISH with probes specific to OR genes with significantly different expression values between B6 and 129 (**Figure 2I**). We then counted the number of OSNs that 6 different OR mRNAs are expressed in, in each strain, and compared this with their RNAseq expression values (**Figure 2J**). We find a high correlation between the difference in OSN number and the difference in RNAseq expression values between B6 and 129 (r^2 = 0.97, p = 0.0003; **Figure 2K**), demonstrating our RNAseq-based approach accurately measures the difference in OSN repertoires between strains.

OR gene clusters are enriched in copy number variants (CNVs) between individual human (Nozawa et al., 2007, Young et al., 2008) and mouse strain genomes (Graubert et al., 2007). Thus it is possible that variance in OSN subtype representations are a consequence of different numbers of highly similar OR genes between strains. To assess this, we mined CAST genome sequence data (Keane et al., 2011) for heterozygous SNPs within annotated OR genes. We identified 51 ORs that contain ten or more heterozygous SNPs, an indication of additional alleles. Using genome sequencing data from these genes we identified 30 novel or misassembled OR genes. We remapped the CAST RNAseq data to a pseudo-CAST genome incorporating these new OR alleles and re-estimated the expression of the OR repertoire. The overall abundance profile remains unchanged except for 36 genes (**Figure S2E**). To assess whether this accounts for the observed differential expression between strains, we compared these estimates to B6. Only 12 of 634 OR genes lose their DE status, and 4 OR genes now become DE (**Figure S2F**). Thus, while differences in OR gene copy number minimally contribute to the diversity in OSN repertoire between three strains of mice, other mechanisms are responsible for most of the variation.

Genetic Background Instructs OSN Diversity Independent of Odor Environment.

Genetically divergent mouse strains produce different chemical odortypes in their urine (Kwak et al., 2012, Yamaguchi et al., 1981) and amniotic fluid (Logan et al., 2012). Therefore each strain of mouse, when housed in homogeneous groups, is exposed to a unique pre- and post-natal olfactory environment. As odor exposure

alters the life-span of OSNs in an activity dependent manner (Francois et al., 2013, Santoro and Dulac, 2012, Watt et al., 2004), genetic variation could regulate OSN population dynamics either directly or indirectly, via odortype. We therefore devised an experiment to test and differentiate the influence of the olfactory environment from the genetic background.

We transferred 4- to 8-cell stage B6 and 129 zygotes to F1 mothers to ensure they experienced an identical *in utero* environment. At birth, B6 litters were cross-fostered to B6 mothers and 129 litters to 129 mothers. In addition, B6 litters received a single 129 pup, and 129 litters a single B6 pup. Therefore, each litter experienced a characteristic olfactory environment, but one animal (the *alien*) had a different genetic background from the others (**Figure 3A**). At 10 weeks of age we quantified the OSN repertoires of six alien animals and six cage-mates using RNAseq. We found that the OSN repertoires cluster in two groups, clearly defined by genetic background (**Figure 3B**). The correlation coefficient for any two B6 samples was on average 0.97, with no significant difference between the environments (t-test, $p = 0.09$). In contrast, the correlation for any B6 with a 129 sample had a mean of 0.89, which is significantly lower (t-test, $p = 3.8 \times 10^{-12}$). 507 OR genes, among 5,475 other genes are DE between these mice when grouped by strain (**Figure S3A**). In striking contrast, across the whole transcriptome we find only mRNA from two genes that show differences in expression according to odor environment, both of which encode ORs (**Figure 3C, S3B**). These data demonstrate that the diversity in OSN repertoire we observe between strains is almost entirely dictated by direct genetic effects. In a controlled environment the influence of odortype on the development and maintenance of the MOE is minimal, perhaps restricted to only a few OSN subtypes.

OSN Diversity Profiles are Independent of OR Function and are Controlled in *cis*.

The indifference of the OSN repertoire to the olfactory environment suggests its development and maintenance is not influenced by the specific activity of OR proteins or, by inference, their protein coding sequence. To further test this, we analyzed the OSN repertoire of newborn pups. We identify the presence of 1,198 (95.9%) OSN subtypes across a dynamic range of abundance (**Figure 4A**). The differential proportions of OSNs expressing particular OR genes are therefore already

present during the development of the MOE, suggesting that it is not dependent on the activity of the OSNs nor on differences in OSN life-span.

Next we analyzed the expression of ORs that are pseudogenized and do not produce receptor proteins capable of odor-mediated activity, but can be co-expressed with functional ORs. These are represented in OSNs with a very similar distribution to functional OR genes (**Figure 4B**). Moreover we analyzed the OR genes that encode identical protein coding sequences between different strains. 36.3% of the OSN subtypes that express identical ORs are differentially represented between 129 and B6. 44.8% are differentially represented between CAST and B6. Together these results suggest that the proportion of each OSN subtype is not dependent on receptor activity.

To directly test whether the abundance of a particular OSN subtype is influenced by the identity of the receptor protein it expresses, we used CRISPR-Cas9 to replace only the coding sequence of *Olfr1507* with that of *Olfr2* (referred to as *Olfr2>Olfr1507*), in a pure B6 genetic background (**Figure 4C**). OSNs expressing *Olfr1507* are the most common subtype in B6 while *Olfr2* expressing OSNs are ranked 334th by decreasing abundance. Homozygous *Olfr2>Olfr1507* animals have 47 fold more *Olfr2*-expressing OSNs compared to controls, and is the highest subtype in these animals (**Figure 4D**). DE analysis of OR genes supports the striking reciprocal differences in *Olfr1507* and *Olfr2*-expressing OSNs in the *Olfr2>Olfr1507* animals, but we also find 122 other OSN subtypes with significant, albeit subtle, differences (over 90% have fold-changes <2) (**Figure 4E**). Taken together these data indicate that the extensive variance in OSN subtype composition we observe within and between mice is determined by the wider genetic architecture of the animal, and is independent of the coding sequence and function of the OR proteins.

To investigate how genetic background influences OSN subtype abundances, we mined 129 and CAST whole genome sequences (Keane et al., 2011) for SNPs and short indels in regulatory regions of OR genes. We find that differentially represented OSN subtypes express OR genes with significantly greater amounts of variation in their coding sequence, whole transcript and regions of 300bp or 1kb upstream of the transcription start site, for both the 129 and CAST genomes (Mann-Whitney one tail, $p < 0.02$ for 129 and $p < 0.0002$ for CAST; **Figure S4**). Further, we scanned OR gene

promoters for O/E and HD binding sites. In the CAST genome, 58 and 310 putative OR promoters have gains or losses of O/E and HD binding sites respectively, compared to the B6 genome. In contrast, only 12 and 46 OR promoters show differences in the number of O/E and HD binding sites respectively when comparing the 129 and B6 genomes.

We therefore hypothesized that OSN subtype repertoires are generated via sequence variance in OR gene promoter and/or local enhancer elements, which dictate the frequency of OR gene choice. For two OR gene clusters it has been demonstrated that enhancer/promoter interactions act in *cis* and do not influence the expression of the homologous OR allele on the other chromosome (Fuss et al., 2007, Khan et al., 2011, Nishizumi et al., 2007). However, recent chromosome conformation capture experiments revealed interchromosomal interactions between OR enhancer elements (Markenscoff-Papadimitriou et al., 2014). Moreover, the differential representation of 122 other OSN subtypes in the *Olfr2>Olfr1507* line (**Figure 4E**), 112 of which express ORs that are located on a different chromosome from *Olfr1507*, is consistent with the possibility that genetic modification of one OR locus directly influences the probability of choice in other ORs, in *trans*.

To determine whether the genetic elements that instruct the whole OSN repertoire are *cis*- or *trans*-acting, we carried out an analysis at the OR allele level in B6×CAST F1 hybrids. Following the logic of (Goncalves et al., 2012), if the genetic elements act in *cis* then we would expect the OSN subtypes that differ between B6 and CAST to be maintained between OSNs expressing the corresponding B6 and CAST alleles within an F1 hybrid. On the other hand, if the elements act in *trans* the number of OSNs that express the B6 derived allele in the F1 would not differ from those that express the CAST allele.

Within F1 mice 840 OSN subtypes (67.2%) expressed OR mRNAs that could be distinguished at the allelic level. Among these, the ratios between B6 and CAST OSN subtype abundance (F0) correlate with the ratios between alleles in the F1 hybrids at approximately 1:1 (**Figure 4F**). In other words, the OSN subtypes expressing a B6 OR allele in F1 animals have the same repertoire as the B6 parent, while the subtypes expressing the CAST OR allele match the CAST parent (**Figure 4G**). Thus, we demonstrate that the genetic elements dictating the abundance of over 800 OSN

subtypes act in *cis*.

Taken together, these data are consistent with a model where genetic variation in local, non-coding regulatory elements determines the probability with which each OR gene is chosen early in OSN neurogenesis.

Acute but not Chronic Odor Exposure Affects OR mRNA Expression in the WOM.

Previous studies have shown that OSNs activated by their cognate ligands have increased life-span (Francois et al., 2013, Santoro and Dulac, 2012, Watt et al., 2004). With time, longer survival rates should translate into enrichment in the neuronal population, compared to those OSN types that are mostly inactive (Santoro and Dulac, 2012). However, we found no evidence of different strain- or sex-derived odors influencing the OSN repertoire (**Figure 2A, 3B).** Because these odor exposures were temporally constant, we hypothesized that the absence of an observed environmental influence on OSN repertoire could be due to olfactory adaptation (a reduction of specific olfactory sensitivity due to prolonged odor exposure, reviewed in (Zufall and Leinders-Zufall, 2000)).

To test this, we exposed mice to a mix of four chemically distinct odorants (acetophenone, eugenol, heptanal and (R)-carvone). The odorant mixture was added to the drinking water supplied to the animals to avoid adaptation, such that they could smell the odor mixture when they approached the bottle to drink (**Figure 5A**). We collected the WOM from animals exposed to the odorants for 24 weeks from birth, along with water-exposed controls, and performed RNAseq. DE analysis reveals 36 OR genes with significantly different mRNA levels (FDR < 5%), with similar numbers more or less abundant in the exposed animals (**Figure 5B, S5**). We selected seven OR genes with the biggest fold-changes in mRNA level for which specific TaqMan qPCR probes were available, and validated their expression levels in an independent cohort. The results indicate that all the tested genes have mRNA levels that are statistically significantly different from controls (t-test, FDR < 5%) and the direction of the expression changes are concordant with the RNAseq data (**Figure 5C**).

To characterize the temporal dynamics of these OR mRNAs, we tested their

expression after different periods of exposure (1, 4 and 10 weeks) in independent samples. After one week of treatment none showed significant differences from controls, which is expected since young pups do not drink from the odorized water bottle. After 4 weeks, three of the OR genes are DE from controls, and at 10 weeks five out of the seven receptors are DE (t-test, FDR < 5%; **Figure 5C**). To assess the plasticity of these changes, we stimulated a group of animals for four weeks, and then removed the stimuli for an additional six weeks. In these mice none of the OR genes are DE from controls (**Figure 5D**). Thus, the overall abundance of specific OR types in WOM is increasingly altered, over a period of weeks to months, upon frequent environmental exposure to defined olfactory cues. These differences are reversible, however, and require persistent stimulation to be maintained.

To investigate whether olfactory adaptation blocks this effect, we presented the same odor mixture on a cotton ball inside a tea strainer (**Figure 5E**), such that the stimuli are present in a sustained manner. None of the same seven OR genes are DE after 24 weeks, nor are any consistently dysregulated during the course of the exposure experiment (t-test, FDR < 5%; **Figure 5E**). Therefore, when odorants are present in the environment in a constant manner (similar to those differentially produced by gender or strains of mice), the OR mRNA abundance levels most responsive to acute exposure remain unchanged.

Differential regulation of OR gene mRNAs is odorant-specific.

If temporal differences in OR mRNA abundance are a consequence of odorant-specific activity, exposure to different odorants should lead to the differential expression of discrete subsets of OR genes. To test this we odorized the drinking water with (R)-carvone alone, heptanal alone, or with the combination of both (**Figure 6A**). After 10 weeks of exposure we tested the expression of the seven DE OR mRNAs that were responsive to the four odor mix (acetophenone, eugenol, heptanal and (R)-carvone), by TaqMan qRT-PCR. None of the genes are significantly DE in the animals exposed to (R)-carvone alone. However, four of the seven OR genes have mRNA levels significantly different in the animals exposed to heptanal, or to the combination of both odorants (t-test, FDR < 5%; **Figure S6A**). We next carried out a transcriptome-wide analysis by RNAseq, finding 43 OR genes significantly DE

in at least one of the conditions (FDR < 5%) of which 32 (74.4%) are upregulated in the odor-stimulated animals (**Figure 6B**). Exposure to (R)-carvone or heptanal resulted in a change in mRNA expression of 15 and 20 OR genes, respectively. These sets of receptors are almost completely independent, with only one OR mRNA significantly upregulated in both groups (*Olfr538*; **Figure 6C-D**). The animals that were exposed to both odorants simultaneously showed significant changes in mRNA levels for 24 OR genes, 15 of which are shared with the individually exposed groups (hypergeometric test, $p = 1.87 \times 10^{-19}$). Almost 40% of the ORs that show significant changes when exposed to all four odorants **(Figure 5B)** are also significantly altered in one or more of the groups exposed to (R)-carvone, heptanal or their combination. Thus, together these data demonstrate that acute environmental exposure to the odorants alters the global expression of around 1.2-1.6% of OR genes in the WOM. These changes are odor-specific and reproducible in isolation and in increasingly complex mixtures.

To investigate whether DE OR genes are directly activated by the environmental odorants, we expressed a subset (*Olfr538*, *Olfr902*, *Olfr916*, *Olfr1182*, *Olfr347* and *Olfr524*) in a heterologous system (Zhuang and Matsunami, 2008) and challenged them with increasing concentrations of (R)-carvone and heptanal. Half of the DE ORs we tested were responsive *in vitro* (**Figure 6E-F, S6B**): for example *Olfr538* displayed a dose-dependent response to (R)-carvone (**Figure 6E**) and *Olfr524* was responsive to heptanal (**Figure 6F**).

Some odorants, including heptanal, are known to be decomposed by enzymes present in the nasal mucus (Nagashima and Touhara, 2010) such that *in vivo* exposure to an odorant may result in stimulation of the OSNs with chemically distinct byproducts. We therefore employed a recently published deorphanization system to identify the ORs that respond to heptanal stimulation *in vivo* (Jiang et al., 2015). This strategy exploits the phosphorylation of the S6 ribosomal subunit when an OSN is activated. Thus by coupling pS6-immunoprecipitation (ps6-IP) and RNAseq, the OR mRNAs expressed in the activated OSNs can be identified. We exposed mice to two concentrations of heptanal for an hour, and sequenced the mRNAs from OSNs after pS6-IP. Seven and 191 OR mRNAs were significantly enriched (FDR < 5%) upon exposure to 1% and 100% heptanal, respectively, compared to controls. Half of DE

ORs after 10-week acute exposure to heptanal (**Figure 6B**) are also DE in the $pS6^{+}$ cells (**Figure 6G**), which is significantly more than expected by chance (hypergeometric test, p = 0.0003). Thus, using both *in vitro* and *in vivo* methods, we conclude that long-term odor-mediated changes in OR gene expression occurs via direct activation of OSNs expressing those receptors.

DISCUSSION

We have exploited the power of RNA sequencing and the monogenic and monoallelic nature of OR gene expression, to comprehensively characterize the full neuronal diversity of a mammalian nose. We have discovered that the representation of OSN subtypes within an individual is highly unequal but stereotyped between animals of the same genetic background. We show that non-coding genetic variation results in high divergence of the relative proportions of different OSN subtypes, with most being susceptible to altered abundance. OSN diversity is controlled by genetic elements that act in *cis*, and is not affected by receptor sequence or function in a sustained olfactory environment. However, the persistent but interleaved presentation of olfactory stimuli can alter the representation of specific ORs in an activity dependent manner, thus subtly shaping the genetically-encoded neuronal repertoire to the olfactory environment.

The MOE is a genetically-determined mosaic of OSN subtypes.

The process of OR gene choice, stabilization and exclusion during OSN maturation is poorly understood. Occasionally, it is referred to as a *random* process (McClintock, 2010, Rodriguez, 2013), suggesting there is no pattern or predictability to the outcome. However our data indicates that, at the OSN population level, the result of this process is deterministic. A particular genetic background in controlled environmental conditions reproducibly generates an OSN population with fixed, unequal proportions of the different OSN subtypes. Thus the process that generates this profile is more accurately described as *stochastic*. Despite divergence in the profiles generated by different genomes, all show a similarly shaped distribution: a small proportion of OSN subtypes are present at high levels with a rapid decay in abundance thereafter. In fact, 3.6% or less of the OSN subtypes contribute to 25% of the overall neuronal content of the WOM. We find that unequal OSN distributions are

already present at birth (**Figure 4A**), suggesting the genetic influence is on the process of OR gene choice/stabilization rather than modulating neuronal survival.

Proximity to the H element, a cluster-specific enhancer, increases the frequency in which an OR is represented within the OSN population (Khan et al., 2011). Therefore the most highly represented OSN types may express OR genes located close to other strong enhancers. However, we propose that genetic variation in enhancers is not sufficient to account for the full diversity of differences in OSN subtypes between strains, as different ORs located adjacent to one another within a cluster are frequently represented very differently. Recently it has been proposed that higher levels of OR transcription per cell may result in more OSNs expressing that receptor due to increased success in a post-selection refinement process (Abdus-Saboor et al., 2016). Our measurements of OR mRNA expression levels in 63 single OSNs (**Figure 1F**) do not support this hypothesis. Instead our data are consistent with a model where non-uniform probabilities of OR choice are instructed by genetic variation in both specific OR promoters and enhancers. Supporting this model, we identified many putative promoters for differentially represented ORs where genetic variation has altered the number of Olf1/Ebf1 (O/E) and homeodomain (HD) transcription factors binding sites between the mouse strains, both sequences known to influence OR choice (D'Hulst et al., 2016, Vassalli et al., 2011). Moreover, through analysis of F1 hybrids we confirmed the finding that the choice of OR genes linked to the H element are regulated in *cis* (Fuss et al., 2007), and extended this to find no evidence of *trans*-acting enhancers regulating choice for over 800 additional OR genes distributed throughout the genome. Instead, the haploid CAST- and B6-derived OR alleles within an F1 are each regulated almost identically as they are in a diploid state within their original genetic backgrounds (**Figure 4F,G**). Our data are inconsistent with *trans*-interactions of multiple enhancers acting additively to regulate OR choice (Markenscoff-Papadimitriou et al., 2014). These *trans* interactions may, however, stabilize or maintain OR singularity after choice has been instructed in *cis*.

Many existing studies into OR gene choice, especially those utilizing transgenic mouse lines, use animals with a mixed 129/B6 genetic background. The remarkable diversity in the OSN repertoire between these strains (**Figure 2**) suggests caution should be exercised in their interpretation. Here we created a mouse line that carries

the coding sequence of *Olfr2* in the locus of *Olfr1507*, the most frequently selected OR gene, in a pure B6 genetic background. Olfr2-expressing OSNs, which rank 334th across the repertoire in the original B6 strain, are the most abundant OSN subtype in this modified line, demonstrating the critical importance of the genetic context in the regulation of OR gene choice. Curiously, we also observed that ~10% of other OSN types show relatively subtle but reproducible differences in abundance. The mechanism underlying these differences is unclear. It could potentially be a consequence of the olfactory system adapting to an OR gene being misexpressed in a non-endogenous zone.

Odor-mediated plasticity in the olfactory system.

The main olfactory epithelium regenerates throughout the life of an animal. It has been suggested that activity-mediated mechanisms may shape the olfactory system by increasing OSN survival (Francois et al., 2013, Santoro and Dulac, 2012, Watt et al., 2004, Zhao and Reed, 2001), though other studies have found that the number of specific OSN subtypes decrease or are unaffected by odor-exposure (Cadiou et al., 2014, Cavallin et al., 2010). Each of these studies focused on one or two OSN subtypes and the odor exposure procedures varied significantly in frequency, persistence and length. Here we took a comprehensive approach, measuring the response of over 1000 ORs to four odorants, after different types of exposure from 1 week to 6 months. We find that mice living in stable chemical environments maintain the olfactory transcriptomes of their genetically-dictated OSN repertoire. However, when frequently recurring odor stimulation is introduced, the abundances of responsive ORs are modified. We propose that this difference is a result of olfactory adaptation in the presence of continuous stimulation. However, other factors may also contribute, for example the continuous exposure odors were likely detected orthonasally while the intermittently exposure odors were retronasally detected.

The lengthy timeframe for odor-mediated differences to emerge is consistent with modulation of OSN lifespan. It is mechanistically unlikely that odor-activity could influence OR gene choice, but it could promote OR expression stabilization. Compared to the dramatic influence of genetic variation on OSN repertoire, odor-evoked changes are subtle, typically less than a 2-fold change after 6 months of exposure. The limited effect magnitude and long-time scales precluded a more

detailed analysis to confirm a correlative alteration in OSN number.

Interestingly, we identified ORs that became more abundant after exposure to specific odorants and, within the same animals, others that became less abundant. Both types were marked by phosphorylation of the S6 ribosomal subunit, a feature of activated OSNs (Jiang et al., 2015), indicating that the differential expression is mediated by OSN subtype-specific olfactory stimulation. This may explain why very different conclusions were drawn from previous exposure studies on a small number of ORs (Cadiou et al., 2014, Cavallin et al., 2010, Francois et al., 2013, Watt et al., 2004). Short-term odor exposures (30 min to 24 hr) result in a temporary downregulation of activated OR mRNA (von der Weid et al., 2015), presumably as part of the olfactory adaptation process. It is possible that our analyses are capturing this dynamic short-term response in addition to changes resulting from long-term exposures. We could not identify any phylogenetic or chromosomal predictor of the ORs that responded with contrasting directional effects, and at this time the logic underpinning the difference in the direction of expression changes remains unexplained.

An individually unique olfactory nose.

Genetic variation has great impact on individual phenotypic traits. Humans differ in up to a third of their OR alleles by functional variation (Mainland et al., 2014), which contributes to an individually unique sense of smell (Secundo et al., 2015). Segregating OR alleles have been functionally linked to perceptual differences of their odor ligands, by altering intensity, valence or detection threshold (Jaeger et al., 2013, Keller et al., 2007, Mainland et al., 2014, McRae et al., 2013, Menashe et al., 2007). However, in most cases these OR coding genetic variants explain only a small proportion of the observed phenotypic variance (reviewed in (Logan, 2014)), suggesting that other factors contribute to individual differences in perception. Recently it has been demonstrated that increasing the number of a particular OSN subtype in a mouse nose increases olfactory sensitivity to its ligand (D'Hulst et al., 2016). Therefore the very different OSN repertoires present between strains of mice are likely to result in significant phenotypic variation in olfactory thresholds, and thus contribute to the individualization of olfaction.

Though it remains to be determined whether human OSN repertoires are as variable

as the mice reported here, an array-based study of OR expression in 26 humans found unequal expression of ORs within and between individual noses (Verbeurgt et al., 2014). Moreover, a recent systematic survey of olfactory perception in humans found high levels of individual variability in reporting the intensity of some odors (for example, benzenethiol and 3-pentanone) but not others (Keller and Vosshall, 2016). Further, a non-coding variant within an OR cluster associated with insensitivity to 2-heptanone has been shown to be dominant to the sensitive allele (McRae et al., 2013). As OR genes are regulated monoallelicaly, this implies that a 50% reduction in the sensitive OR allele dosage is, in some cases, sufficient to influence perception. On the other hand, because many odorants activate multiple OSN subtypes (Malnic et al., 1999), a differential representation of one subtype may have a limited influence on the overall perception of its odor.

Further investigation into the functional consequence of diverse OSN repertoires will be necessary to determine the full extent to which they individualize the sense of smell.

EXPERIMENTAL PROCEDURES

RNA sequencing and data analysis.

Animal experiments were carried out under the authority of a UK Home Office license (80/2472), after review by the Wellcome Trust Sanger Institute Animal Welfare and Ethical Review Board. Details of the strain, age and sex of each animal sequenced can be found in Table S1. MOEs were dissected and immediately homogenized in lysis RLT buffer (Qiagen). Total RNA was extracted using the RNeasy mini kit (Qiagen) with on-column DNAse digestion, following the manufacturer's protocol.

mRNA was prepared for sequencing using the TruSeq RNA sample preparation kit (Illumina). All RNA sequencing was paired-end and produced 100-nucleotide-long reads. Sequencing data were aligned with STAR 2.3 (Dobin et al., 2013) to the GRCm38 mouse reference genome (B6) or to pseudo-genomes created for the different strains using Seqnature (Munger et al., 2014). The annotation used was from the Ensembl mouse genome database version 72 (http://jun2013.archive.ensembl.org/info/data/ftp/index.html), modified to include all

reconstructed gene models for OR genes as reported in (Ibarra-Soria et al., 2014). The numbers of fragments uniquely aligned to each gene were obtained using the HTSeq 0.6.1 package, with the script htseq-count, mode *intersection-nonempty* (Anders et al., 2015). Data analysis, statistical testing and plotting were carried out in R (http://www.R-project.org). OR expression values are provided in Files S1-S3.

Differential expression analysis was performed with DESeq2 1.8.1 (Love et al., 2014) with standard parameters. Genes were considered significantly DE if they had an adjusted p-value of 0.05 or less. Detailed results are provided in Files S1-S3.

Genetic or environmental effects on OR expression.

To dissect the influence of the genetic background from the olfactory environment, C57BL/6N and 129S5 4 to 8-cell stage embryos were transferred into F1 (C57BL/6J×CBA) pseudo-pregnant females. One day after birth, the C57BL/6N and 129S5 litters were cross-fostered to C57BL/6N and 129S5 wild-type mothers, respectively. Then, a single pup from the other strain was transferred to the cross-fostered litter (the *alien*). At 10 weeks of age, the WOM was collected form the alien and a randomly selected cage-mate, and RNA was extracted as described.

Generation of the *Olfr2>Olfr1507* mice.

CRISPR-Cas9 technology was used to generate double strand breaks on either side of the *Olfr1507* coding sequence and facilitate homologous recombination with a DNA vector containing the coding sequence of *Olfr2* and homology arms for the *Olfr1507* locus of ~1kb. All components were microinjected into the cytoplasm of 112 C57BL/6N zygotes at the following concentrations: 25 ng/ul for each gRNA, 100 ng/ul of Cas9 RNA and 200 ng/ul of vector DNA.

Odor-exposure *in vivo*.

B6 mice were exposed to a mix of heptanal, (R)-carvone, eugenol and acetophenone at 1mM concentration each. For the acute exposure experiments, the odor mix was added to the water bottles of the animals; mineral oil alone was used for controls. The exposure started from at least E14.5 and the WOM was collected from age-matched exposed and control groups at different time-points after the start of the treatment. For the chronic exposure experiments, the odor mixture, or mineral oil only, were applied to a cotton ball with a plastic pasteur pipette; these were put into metal tea strainers

that were then introduced into the cage of the animals. The cotton ball was replaced fresh daily. The exposure started from birth and the WOM was collected from age-matched exposed and control groups at different time-points after the start of the treatment.

qRT-PCR expression estimation.

1 μg of WOM RNA was reversed-transcribed into cDNA using the High-Capacity RNA-to-cDNA kit (Applied Biosystems) with the manufacturer's protocol. Predesigned TaqMan gene expression assays were used on a 7900HT Fast Real-Time PCR System (Life Technolo- gies) following the manufacturer's instructions. Mean cycle threshold (Ct) values were obtained from two technical replicates, each normalized to *Actb* using the ΔCt method. Relative quantity (RQ) values were calculated using the formula $RQ = 2^{\Delta Ct}$. Differential expression between groups was assessed in R, by a two-tailed t-test, with multiple-testing correction by the Benjamini & Hochberg (FDR) method.

***In vitro* screening.**

For OR response in vitro, a Dual-Glo Luciferase Assay System (Promega) was employed using the previously described method (Zhuang and Matsunami, 2008). 10-fold serial dilutions of each odorant, from 1mM to 1nM, were assayed in triplicate. The data were fit to a sigmoidal curve and every OR-odorant pair was compared to a vector-only control using an extra sums-of-squares *F* test. Data were analyzed with GraphPad Prism 7.00 and R.

In vivo screening.

pS6 immunoprecipitation, RNAseq and analysis for differentially represented ORs were performed as described in Jiang et al. 2015 (Jiang et al., 2015).

AUTHOR CONTRIBUTIONS

X.I.-S. and D.W.L. conceived the project, designed experiments, interpreted results and wrote the manuscript. X.I.-S. carried out most experiments, processed the data and performed analyses. T.S.N., M.A.A.S., P.H.M.N. and F.P. performed *in situ* hybridization experiments. J.L. and T.M.K. analyzed CAST olfactory receptor sequences and identified novel alleles. Y.J., K.I. and H.M. performed the *in vivo*

receptor deorphanization and analyzed the corresponding data. C.T., N.R.M. and J.D.M. performed the *in vitro* receptor deorphanization. M.K. and A.K. performed the chronic odor exposure experiment. L.R.S collected and prepared single cell sequencing libraries. D.W.L. oversaw experiments and supervised the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank M Khan and P Mombaerts for providing RNA from the *ΔOlfr7Δ* mouse line that was used in Figure 1B; G Gurria and SS Gerety for experimental support and insightful discussions; L Stowers for her critical review of the manuscript; and the Sanger Institute Research Support Facility, microinjection and sequencing pipeline staff for invaluable technical support.

REFERENCES

ABDUS-SABOOR, I., AL NUFAL, MOHAMMED J., AGHA, MAHA V., RUINART DE BRIMONT, M., FLEISCHMANN, A. & SHYKIND, BENJAMIN M. 2016. An Expression Refinement Process Ensures Singular Odorant Receptor Gene Choice. *Current Biology,* 26**,** 1083-1090.

ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics,* 31**,** 166-9.

BRESSEL, O. C., KHAN, M. & MOMBAERTS, P. 2016. Linear correlation between the number of olfactory sensory neurons expressing a given mouse odorant receptor gene and the total volume of the corresponding glomeruli in the olfactory bulb. *J Comp Neurol,* 524**,** 199-209.

BUCK, L. & AXEL, R. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell,* 65**,** 175-187.

CADIOU, H., AOUDE, I., TAZIR, B., MOLINAS, A., FENECH, C., MEUNIER, N. & GROSMAITRE, X. 2014. Postnatal odorant exposure induces peripheral olfactory plasticity at the cellular level. *J Neurosci,* 34**,** 4857-70.

CAVALLIN, M. A., POWELL, K., BIJU, K. C. & FADOOL, D. A. 2010. State-dependent sculpting of olfactory sensory neurons is attributed to sensory enrichment, odor deprivation, and aging. *Neurosci Lett,* 483**,** 90-5.

CHESS, A., SIMON, I., CEDAR, H. & AXEL, R. 1994. Allelic inactivation regulates olfactory receptor gene expression. *Cell,* 78**,** 823-834.

D'HULST, C., MINA, RAENA B., GERSHON, Z., JAMET, S., CERULLO, A., TOMOIAGA, D., BAI, L., BELLUSCIO, L., ROGERS, MATTHEW E., SIROTIN, Y. & FEINSTEIN, P. 2016. MouSensor: A Versatile Genetic Platform to Create Super Sniffer Mice for Studying Human Odor Coding. *Cell Reports*.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

FRANCOIS, A., LAZIZ, I., RIMBAUD, S., GREBERT, D., DURIEUX, D., PAJOT-AUGY, E. & MEUNIER, N. 2013. Early survival factor deprivation in the olfactory epithelium enhances activity-driven survival. *Front Cell Neurosci,* 7**,** 271.

FUSS, S. H., OMURA, M. & MOMBAERTS, P. 2007. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell,* 130**,** 373-84.

GONCALVES, A., LEIGH-BROWN, S., THYBERT, D., STEFFLOVA, K., TURRO, E., FLICEK, P., BRAZMA, A., ODOM, D. T. & MARIONI, J. C. 2012. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res,* 22**,** 2376-84.

GRAUBERT, T. A., CAHAN, P., EDWIN, D., SELZER, R. R., RICHMOND, T. A., EIS, P. S., SHANNON, W. D., LI, X., MCLEOD, H. L., CHEVERUD, J. M. & LEY, T. J. 2007. A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet,* 3**,** e3.

HANCHATE, N. K., KONDOH, K., LU, Z., KUANG, D., YE, X., QIU, X., PACHTER, L., TRAPNELL, C. & BUCK, L. B. 2015. Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science,* 350**,** 1251-5.

IBARRA-SORIA, X., LEVITIN, M. O., SARAIVA, L. R. & LOGAN, D. W. 2014. The olfactory transcriptomes of mice. *PLoS Genet,* 10**,** e1004593.

JAEGER, S. R., MCRAE, J. F., BAVA, C. M., BERESFORD, M. K., HUNTER, D., JIA, Y., CHHEANG, S. L., JIN, D., PENG, M., GAMBLE, J. C., ATKINSON, K. R., AXTEN, L. G., PAISLEY, A. G., TOOMAN, L., PINEAU, B., ROUSE, S. A. & NEWCOMB, R. D. 2013. A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Curr Biol,* 23**,** 1601-5.

JIANG, Y., GONG, N. N., HU, X. S., NI, M. J., PASI, R. & MATSUNAMI, H. 2015. Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat Neurosci,* 18**,** 1446-54.

KAWAGISHI, K., ANDO, M., YOKOUCHI, K., SUMITOMO, N., KARASAWA, M., FUKUSHIMA, N. & MORIIZUMI, T. 2014. Stereological quantification of olfactory receptor neurons in mice. *Neuroscience,* 272**,** 29-33.

KEANE, T. M., GOODSTADT, L., DANECEK, P., WHITE, M. A., WONG, K., YALCIN, B., HEGER, A., AGAM, A., SLATER, G., GOODSON, M., FURLOTTE, N. A., ESKIN, E., NELLAKER, C., WHITLEY, H., CLEAK, J., JANOWITZ, D., HERNANDEZ-PLIEGO, P., EDWARDS, A., BELGARD, T. G., OLIVER, P. L., MCINTYRE, R. E., BHOMRA, A., NICOD, J., GAN, X., YUAN, W., VAN DER WEYDEN, L., STEWARD, C. A., BALA, S., STALKER, J., MOTT, R., DURBIN, R., JACKSON, I. J., CZECHANSKI, A., GUERRA-ASSUNCAO, J. A., DONAHUE, L. R., REINHOLDT, L. G., PAYSEUR, B. A., PONTING, C. P., BIRNEY, E., FLINT, J. & ADAMS, D. J. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature,* 477**,** 289-94.

KELLER, A. & VOSSHALL, L. B. 2016. Olfactory perception of chemically diverse molecules. *bioRxiv*.

KELLER, A., ZHUANG, H., CHI, Q., VOSSHALL, L. B. & MATSUNAMI, H. 2007. Genetic variation in a human odorant receptor alters odour perception. *Nature,* 449**,** 468-72.

KHAN, M., VAES, E. & MOMBAERTS, P. 2011. Regulation of the probability of mouse odorant receptor gene choice. *Cell,* 147**,** 907-21.

KHAN, M., VAES, E. & MOMBAERTS, P. 2013. Temporal patterns of odorant receptor gene expression in adult and aged mice. *Mol Cell Neurosci,* 57**,** 120-9.

KWAK, J., GRIGSBY, C. C., RIZKI, M. M., PRETI, G., KOKSAL, M., JOSUE, J., YAMAZAKI, K. & BEAUCHAMP, G. K. 2012. Differential binding between volatile ligands and major urinary proteins due to genetic variation in mice. *Physiol Behav,* 107**,** 112-20.

LOGAN, D. W. 2014. Do you smell what I smell? Genetic variation in olfactory perception. *Biochem Soc Trans,* 42**,** 861-5.

LOGAN, D. W., BRUNET, L. J., WEBB, W. R., CUTFORTH, T., NGAI, J. & STOWERS, L. 2012. Learned recognition of maternal signature odors mediates the first suckling episode in mice. *Curr Biol,* 22**,** 1998-2007.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

MAINLAND, J. D., KELLER, A., LI, Y. R., ZHOU, T., TRIMMER, C., SNYDER, L. L., MOBERLY, A. H., ADIPIETRO, K. A., LIU, W. L., ZHUANG, H., ZHAN, S., LEE, S. S., LIN, A. & MATSUNAMI, H. 2014. The missense of smell: functional variability in the human odorant receptor repertoire. *Nat Neurosci,* 17**,** 114-20.

MALNIC, B., HIRONO, J., SATO, T. & BUCK, L. B. 1999. Combinatorial receptor codes for odors. *Cell,* 96**,** 713-23.

MARKENSCOFF-PAPADIMITRIOU, E., ALLEN, W. E., COLQUITT, B. M., GOH, T., MURPHY, K. K., MONAHAN, K., MOSLEY, C. P., AHITUV, N. & LOMVARDAS, S. 2014. Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell,* 159**,** 543-57.

MCCLINTOCK, T. S. 2010. Achieving Singularity in Mammalian Odorant Receptor Gene Choice. *Chemical Senses,* 35**,** 447-457.

MCRAE, J. F., JAEGER, S. R., BAVA, C. M., BERESFORD, M. K., HUNTER, D., JIA, Y., CHHEANG, S. L., JIN, D., PENG, M., GAMBLE, J. C., ATKINSON, K. R., AXTEN, L. G., PAISLEY, A. G., WILLIAMS, L., TOOMAN, L., PINEAU, B., ROUSE, S. A. & NEWCOMB, R. D. 2013. Identification of regions associated with variation in sensitivity to food-related odors in the human genome. *Curr Biol,* 23**,** 1596-600.

MENASHE, I., ABAFFY, T., HASIN, Y., GOSHEN, S., YAHALOM, V., LUETJE, C. W. & LANCET, D. 2007. Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol,* 5**,** e284.

MUNGER, S. C., RAGHUPATHY, N., CHOI, K., SIMONS, A. K., GATTI, D. M., HINERFELD, D. A., SVENSON, K. L., KELLER, M. P., ATTIE, A. D., HIBBS, M. A., GRABER, J. H., CHESLER, E. J. & CHURCHILL, G. A. 2014. RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. *Genetics,* 198**,** 59-73.

NAGASHIMA, A. & TOUHARA, K. 2010. Enzymatic conversion of odorants in nasal mucus affects olfactory glomerular activation patterns and odor perception. *J Neurosci,* 30**,** 16391-8.

NISHIZUMI, H., KUMASAKA, K., INOUE, N., NAKASHIMA, A. & SAKANO, H. 2007. Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in cis. *Proc Natl Acad Sci U S A,* 104**,** 20067-72.

NOZAWA, M., KAWAHARA, Y. & NEI, M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci U S A,* 104**,** 20421-6.

RODRIGUEZ, I. 2013. Singular Expression of Olfactory Receptor Genes. *Cell,* 155**,** 274-277.

RODRIGUEZ-GIL, D. J., TRELOAR, H. B., ZHANG, X., MILLER, A. M., TWO, A., IWEMA, C., FIRESTEIN, S. J. & GREER, C. A. 2010. Chromosomal location-dependent non-stochastic onset of odor receptor expression. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 30**,** 10067-10075.

ROTHMAN, A., FEINSTEIN, P., HIROTA, J. & MOMBAERTS, P. 2005. The promoter of the mouse odorant receptor gene M71. *Molecular and Cellular Neuroscience,* 28**,** 535-546.

ROYAL, S. J. & KEY, B. 1999. Development of P2 Olfactory Glomeruli in P2-Internal Ribosome Entry Site-Tau-LacZ Transgenic Mice. *The Journal of Neuroscience,* 19**,** 9856-9864.

SANTORO, S. W. & DULAC, C. 2012. The activity-dependent histone variant H2BE modulates the life span of olfactory neurons. *Elife,* 1**,** e00070.

SARAIVA, L. R., AHUJA, G., IVANDIC, I., SYED, A. S., MARIONI, J. C., KORSCHING, S. I. & LOGAN, D. W. 2015a. Molecular and neuronal homology between the olfactory systems of zebrafish and mouse. *Sci Rep,* 5**,** 11487.

SARAIVA, L. R., IBARRA-SORIA, X., KHAN, M., OMURA, M., SCIALDONE, A., MOMBAERTS, P., MARIONI, J. C. & LOGAN, D. W. 2015b. Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Sci Rep,* 5**,** 18178.

SECUNDO, L., SNITZ, K., WEISSLER, K., PINCHOVER, L., SHOENFELD, Y., LOEWENTHAL, R., AGMON-LEVIN, N., FRUMIN, I., BAR-ZVI, D., SHUSHAN, S. & SOBEL, N. 2015. Individual olfactory perception reveals meaningful nonolfactory genetic information. *Proc Natl Acad Sci U S A,* 112**,** 8750-5.

TAN, L., LI, Q. & XIE, X. S. 2015. Olfactory sensory neurons transiently express multiple olfactory receptors during development. *Mol Syst Biol,* 11**,** 844.

VASSALLI, A., FEINSTEIN, P. & MOMBAERTS, P. 2011. Homeodomain binding motifs modulate the probability of odorant receptor gene choice in transgenic mice. *Mol Cell Neurosci,* 46**,** 381-96.

VERBEURGT, C., WILKIN, F., TARABICHI, M., GREGOIRE, F., DUMONT, J. E. & CHATELAIN, P. 2014. Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS One,* 9**,** e96333.

VON DER WEID, B., ROSSIER, D., LINDUP, M., TUBEROSA, J., WIDMER, A., COL, J. D., KAN, C., CARLETON, A. & RODRIGUEZ, I. 2015. Large-scale transcriptional profiling of chemosensory neurons identifies receptor-ligand pairs in vivo. *Nat Neurosci,* 18**,** 1455-63.

WATT, W. C., SAKANO, H., LEE, Z. Y., REUSCH, J. E., TRINH, K. & STORM, D. R. 2004. Odorant stimulation enhances survival of olfactory sensory neurons via MAPK and CREB. *Neuron,* 41**,** 955-67.

WHITE, J. G., SOUTHGATE, E., THOMSON, J. N. & BRENNER, S. 1986. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci,* 314**,** 1-340.

WICHTERLE, H., GIFFORD, D. & MAZZONI, E. 2013. Mapping neuronal diversity one cell at a time. *Science,* 341**,** 726-7.

XIE, S. Y., FEINSTEIN, P. & MOMBAERTS, P. 2000. Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. *Mamm Genome,* 11**,** 1070-8.

YAMAGUCHI, M., YAMAZAKI, K., BEAUCHAMP, G. K., BARD, J., THOMAS, L. & BOYSE, E. A. 1981. Distinctive urinary odors governed by the major histocompatibility locus of the mouse. *Proc Natl Acad Sci U S A,* 78**,** 5817-20.

YOUNG, J. M., ENDICOTT, R. M., PARGHI, S. S., WALKER, M., KIDD, J. M. & TRASK, B. J. 2008. Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet,* 83**,** 228-42.

YOUNG, J. M., LUCHE, R. M. & TRASK, B. J. 2011. Rigorous and thorough bioinformatic analyses of olfactory receptor promoters confirm enrichment of O/E and homeodomain binding sites but reveal no new common motifs. *BMC Genomics,* 12**,** 1-22.

YOUNG, J. M., SHYKIND, B. M., LANE, R. P., TONNES-PRIDDY, L., ROSS, J. A., WALKER, M., WILLIAMS, E. M. & TRASK, B. J. 2003. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biology,* 4**,** R71-R71.

ZHAO, H., IVIC, L., OTAKI, J. M., HASHIMOTO, M., MIKOSHIBA, K. & FIRESTEIN, S. 1998. Functional Expression of a Mammalian Odorant Receptor. *Science,* 279**,** 237-242.

ZHAO, H. & REED, R. R. 2001. X Inactivation of the OCNC1 Channel Gene Reveals a Role for Activity-Dependent Competition in the Olfactory System. *Cell,* 104**,** 651-660.

ZHUANG, H. & MATSUNAMI, H. 2008. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat. Protocols,* 3**,** 1402-1413.

ZUFALL, F. & LEINDERS-ZUFALL, T. 2000. The cellular and molecular basis of odor adaptation. *Chem Senses,* 25**,** 473-81.

FIGURES

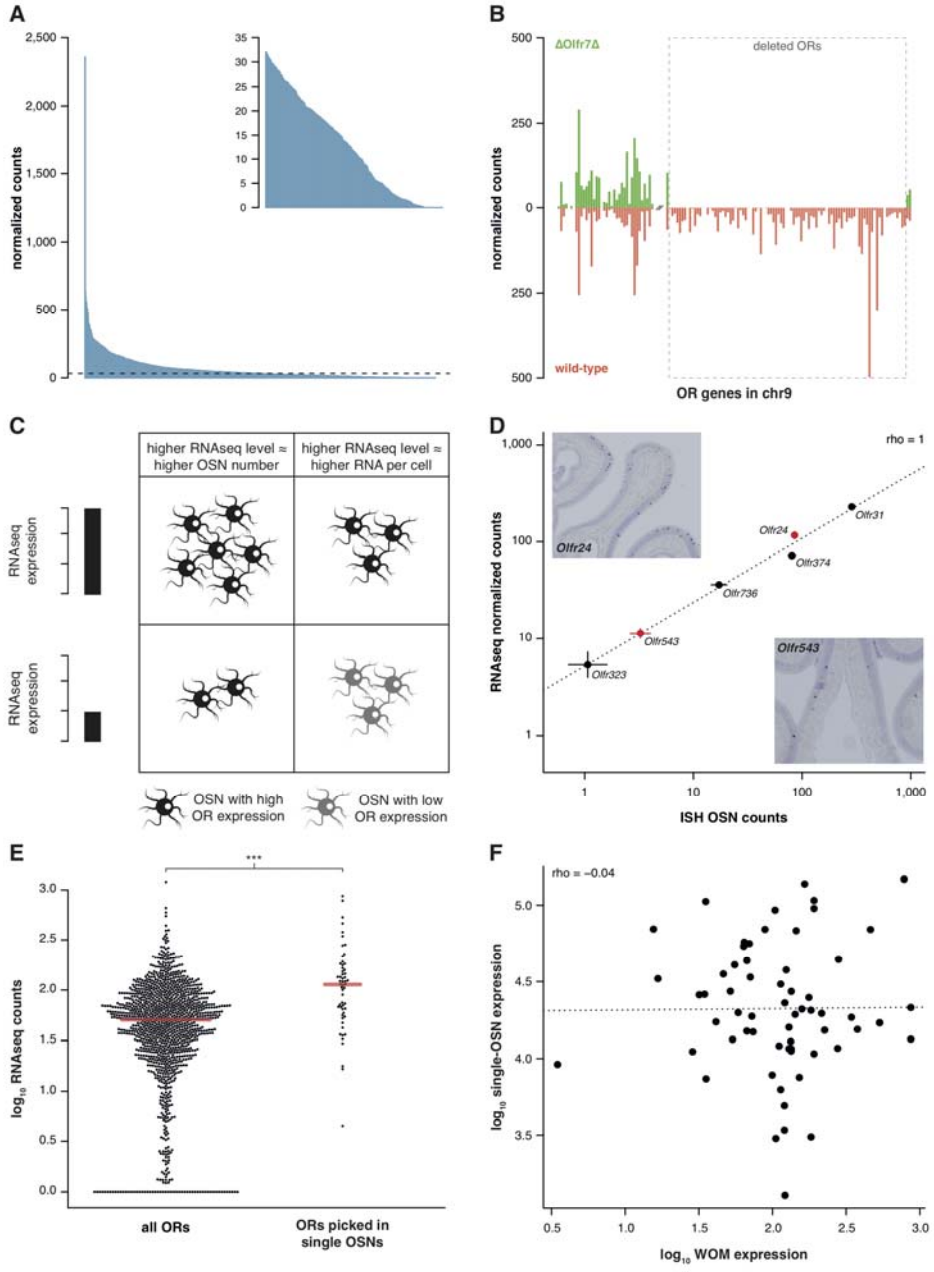

Figure 1: RNAseq is highly sensitive for OR mRNA detection and provides a measurement of OSN diversity. (A) Barplot of the mean normalized expression of 1249 OR genes from six biological replicates, accounting for gene length. Genes are ordered by decreasing abundance. The horizontal line is the median expression (32.06) and all the genes below it are shown in the inset. **(B)** Mean normalized mRNA expression values for the OR genes in chromosome 9 of the *ΔOlfr7Δ* mouse line

(green; n=3). The corresponding abundances in wild-type animals (orange) are shown as a mirror image (n=3). The break on the x-axis separates the two OR clusters. The dotted box encloses the deleted ORs. **(C)** Unequal RNAseq expression levels for different OR genes can be explained by two scenarios: (left) an OR gene with high RNAseq levels is expressed by a larger number of OSNs than a gene with low RNAseq abundance; and/or (right) an OR with high RNAseq values is expressed in the same number of OSNs as one with low RNAseq values, but at higher levels per OSN. **(D)** Comparison of the number of OSNs that express 6 OR genes assessed by *in situ* hybridization (ISH; x-axis) to the corresponding RNAseq values (y-axis). Error bars are the standard error of the mean (ISH n=4, RNAseq n=6). The line is the linear regression and the Spearman's correlation coefficient (rho) indicates a perfect correlation. Representative ISH images of two OR genes (in red) are shown. **(E)** In single-cell RNAseq experiments, 63 OSNs were randomly collected from the MOE. The distribution of OR mRNA expression in WOM samples is plotted (left), alongside the equivalent values for the ORs that were present in the picked single-OSNs (right). There is a significant enrichment ($p < 6.44 \times 10^{-9}$) towards collecting OSNs that express OR genes with high RNAseq counts in WOM. **(F)** Comparison of the normalized expression value for the highest OR detected in each of the 63 single-OSNs (y-axis) to the corresponding mean value in WOM (x-axis, n=3). The line is the linear regression and the Spearman's correlation coefficient (rho) indicates there is no correlation. See also Figure S1.

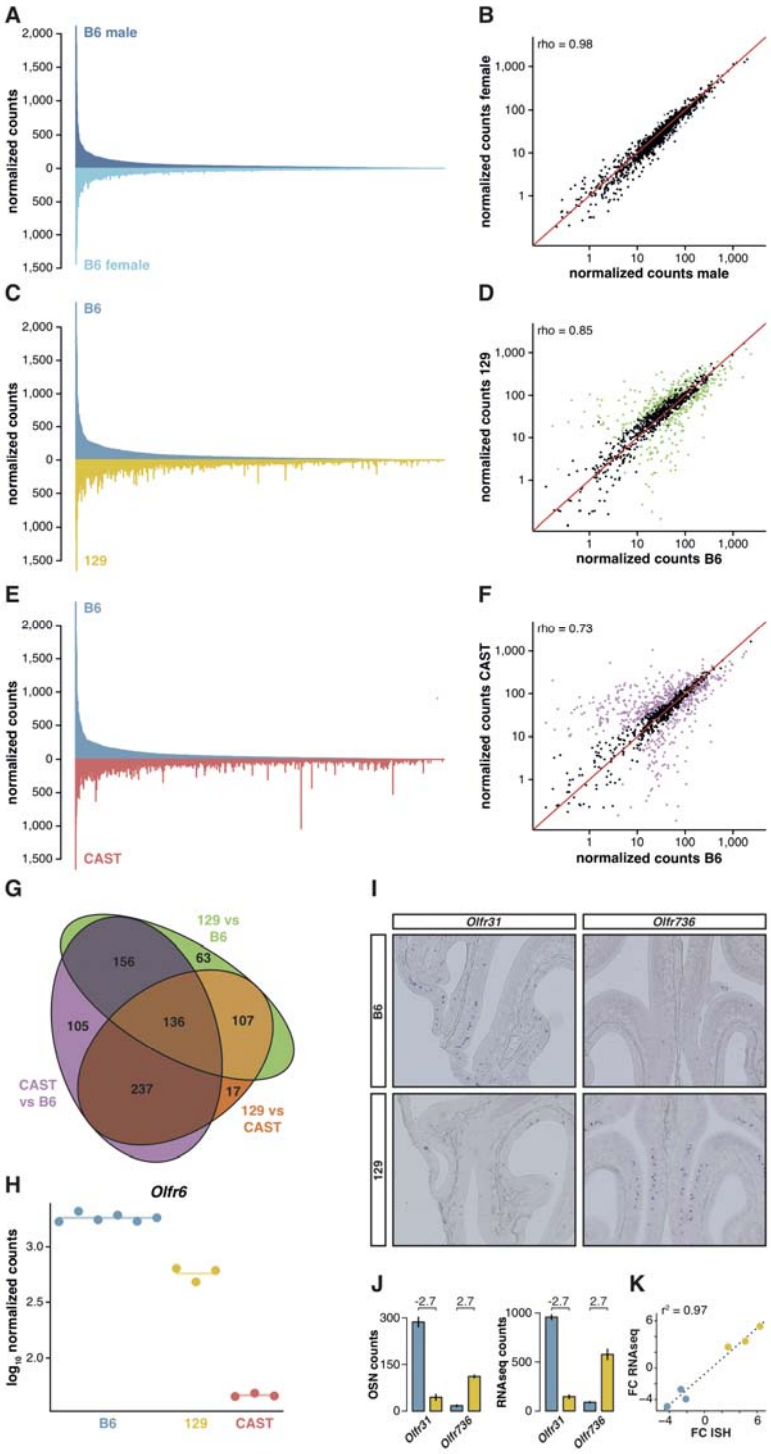

Figure 2: OSN diversity varies between mouse strains. (A) Mirrored barplot of the mean normalized RNAseq expression values for the OR genes in male (dark blue, top) and female (light blue, bottom) B6 animals (n=3). **(B)** Scatter plot for the same data, with the Spearman's correlation (rho) indicating a strong correlation. The red

line is the 1:1 diagonal. Significantly differentially expressed (DE) OR genes are represented in blue, non-DE genes are in black. **(C)** Same as in (A) but with the average B6 expression in blue (both males and females, n=6) compared to the corresponding 129 expression values in yellow (n=3). **(D)** Corresponding scatter plot, with the significant DE genes in green. **(E)** Same as in (A) but comparing the B6 expression in blue (n=6) to the CAST abundances in red (n=3). **(F)** Corresponding scatter plot, with DE genes in purple. **(G)** Venn diagram illustrating the intersection of DE OR genes between the pairwise comparisons of the three strains. **(H)** An example of an OR gene, *Olfr6*, that is DE in all strain comparisons. **(I)** Representative in situ hybridizations (ISH) on coronal slices of B6 and 129 MOEs for two OR genes, *Olfr31* and *Olfr736*, that are DE between these strains. **(J)** The quantification of OSNs expressing each OR gene in B6 (blue) and 129 (yellow) are plotted alongside the corresponding RNAseq counts. The $\log_2$ fold-changes between the strains are indicated. **(K)** Fold-change calculations from ISH data (x-axis) or RNAseq counts (y-axis) for six DE OR genes, including *Olfr31* and *Olfr736*. The line is the linear regression and the Pearson's coefficient (r^2) indicates a strong correlation between OSN and RNAseq counts. See also Figure S2.

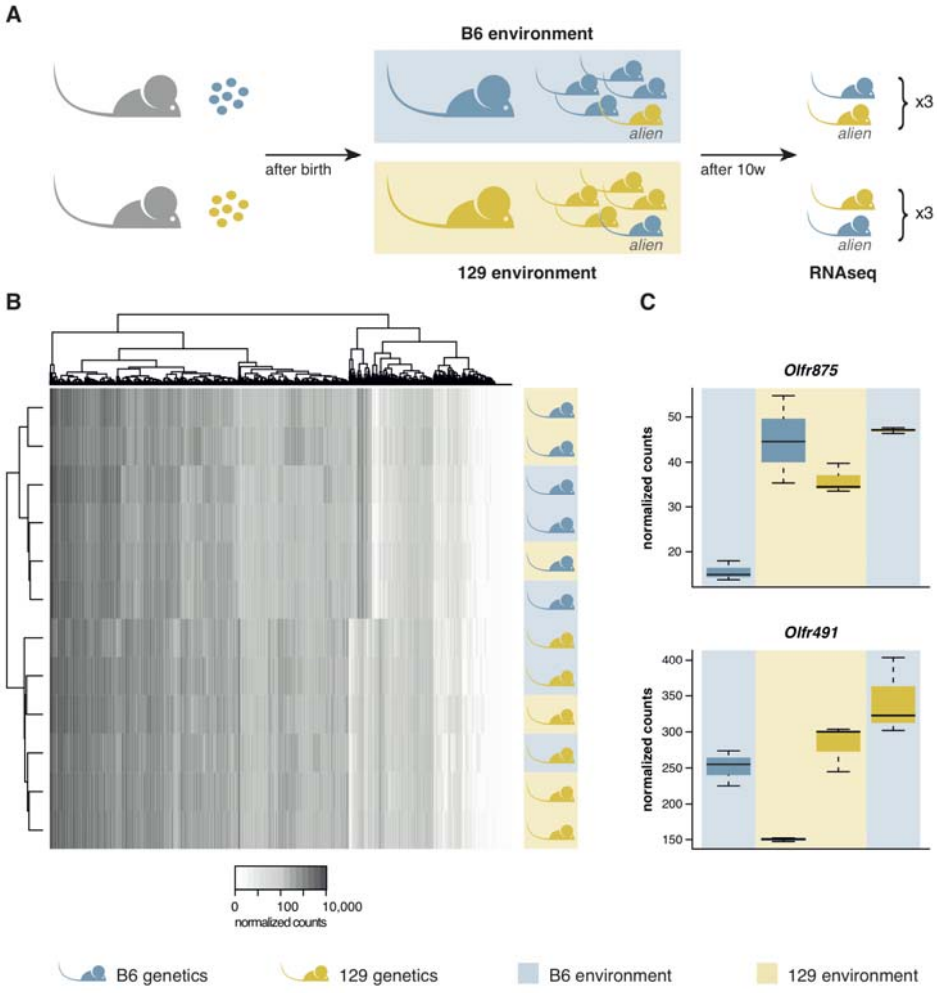

Figure 3: OSN diversity is determined by the genetic background and not by the olfactory environment. (A) Experimental strategy to differentiate genetic from environmental influences on OSN diversity. B6 (blue) and 129 (yellow) embryos, depicted as circles, were transferred into F1 recipient mothers (grey). After birth, the litters were cross-fostered to B6 and 129 mothers respectively. Each B6 litter received on 129 pup (the alien) and vice versa. After 10 weeks the WOM was collected for RNAseq from three aliens from each strain, and one cage-mate each. **(B)** Heatmap of the expression of the OR genes (columns) in all 12 sequenced animals (rows). Samples cluster by the genetic background of the animals. The strain and environment of each mouse is indicated through shading (right). **(C)** Differential expression analyses revealed mRNA from only two genes, *Olfr875* and *Olfr491*, that are significantly altered based on the olfactory environment. Expression values are shown

for each group. Blue and yellow boxes indicate B6 or 129 animals respectively, and the background indicates the olfactory environment. See also Figure S3.

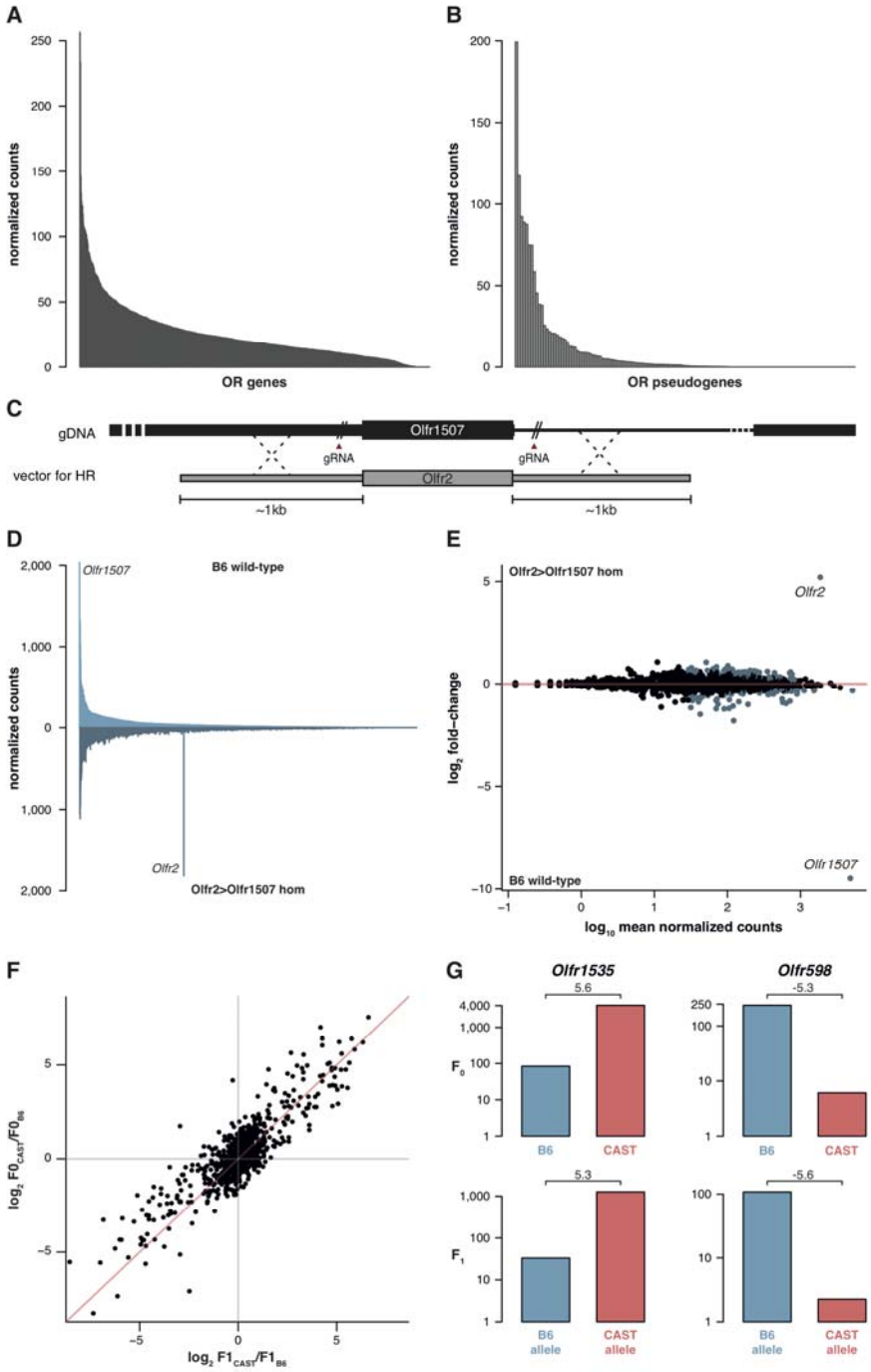

Figure 4: OSN diversity is independent of OR activity and is controlled in *cis*. (A) Mean normalized expression of the OR mRNA in the WOM of newborn B6 animals, arranged from most to least abundant (n=3). **(B)** Mean normalized mRNA expression

of 134 annotated OR pseudogenes in the B6 adult WOM (n=6). **(C)** A genetically modified mouse line was produced that contains the coding sequence (CDS) of *Olfr2* in place of *Olfr1507* (*Olfr2>Olfr1507*). The strategy combined the use of CRISPR-Cas9 technology to create double-strand breaks on either side of the *Olfr1507* CDS, and a DNA vector containing the *Olfr2* CDS along with ~1kb homology arms for homologous recombination. **(D)** Mirrored barplot of the mean normalized mRNA expression values for the OR repertoire in B6 animals (light blue, top; n=4) and in *Olfr2>Olfr1507* homozygous mutants (dark blue, bottom; n=4). *Olfr2* becomes the most abundant OR and *Olfr1507* is no longer expressed in the genetically modified line. **(E)** Scatter plot of the mean normalized counts (x-axis) of OR genes versus the $\log_2$ fold-change between *Olfr2>Olfr1507* homozygotes and WT controls (y-axis, n=4). OR genes that are significantly differentially expressed are represented in blue. *Olfr2* and *Olfr1507* are strikingly different whereas the rest of the repertoire is equivalent or very slightly altered. **(F)** Comparison of the fold-change of the CAST versus B6 OR expression (y-axis) to the fold-change between the CAST and B6 alleles in the F1 (x-axis). The genes fall on the 1:1 diagonal (red line) indicating the mRNA expression pattern observed in the parents is preserved in the F1 and thus OR abundance is controlled in *cis*. **(G)** Examples of the normalized mRNA expression in the parental strains (top) of an OR gene that is more abundant in CAST (*Olfr1535*) or in B6 (*Olfr598*). The corresponding mRNA abundance of each allele in the F1 (bottom) is preserved. The $\log_2$ fold-change is indicated for each comparison. See also Figure S4.

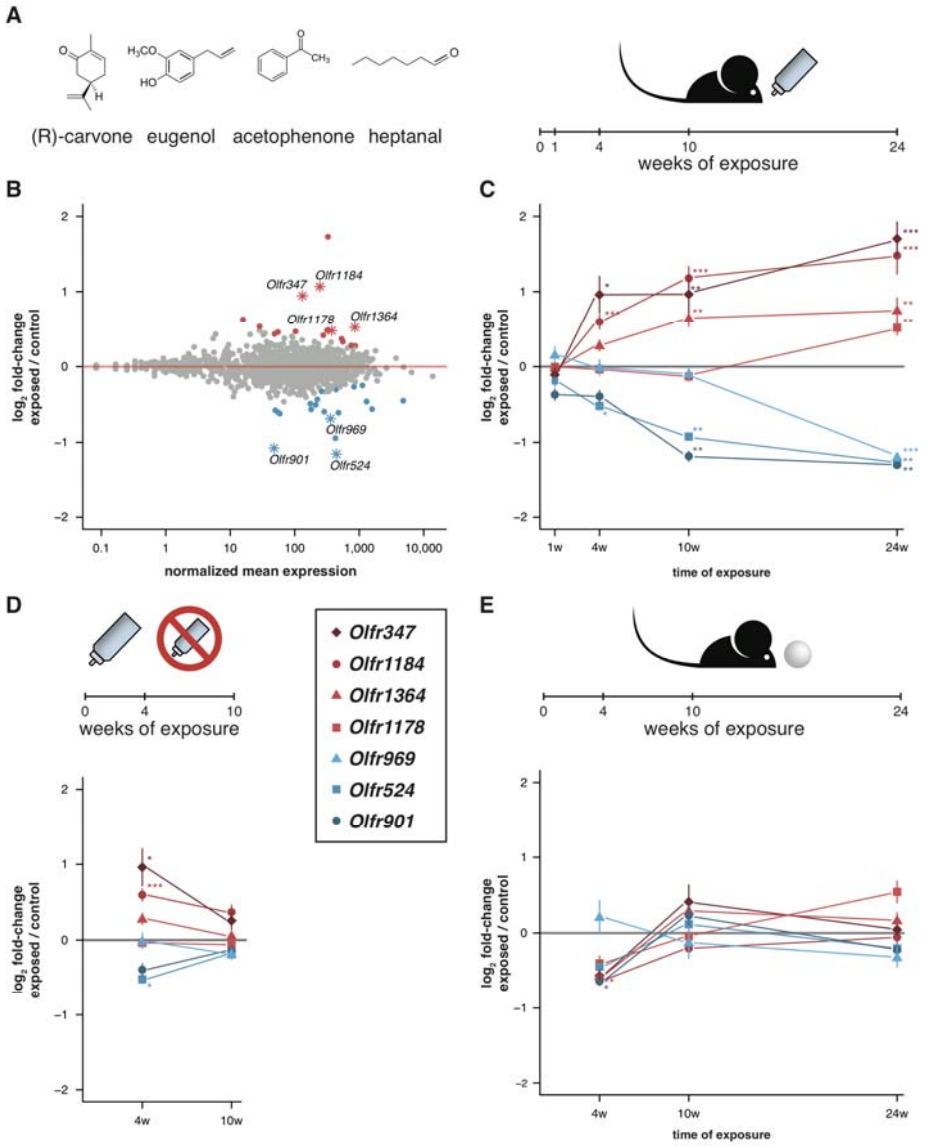

Figure 5: Acute but not chronic exposure to odors alter OR mRNA abundance. **(A)** Four different odorants were mixed together and used to stimulate B6 animals. In an acute paradigm, the odor mix was added to the drinking water supplied to the animals and WOM was collected at different time-points. **(B)** WOM from animals exposed for 24 weeks and matched controls were sequenced (n=6). The plot shows the normalized mean mRNA expression value (x-axis) for each OR gene compared to its fold-change in exposed versus control samples (y-axis). Genes highlighted in red or blue have significantly up or downregulated mRNAs, respectively. OR genes represented by an asterisk were selected for further validation. **(C)** qRT-PCR validation of the DE genes highlighted in (B). The mean fold-change between

exposed and control samples is plotted for animals exposed for differing periods of time (x-axis). After 24 weeks of exposure, all the genes are significantly DE (n=8-13). **(D)** Animals were acutely exposed to the odor mix for 4 weeks and then the stimulus was removed for 6 weeks. After the recovery period none of the OR mRNAs are significantly different from controls (n=8). **(E)** A chronic exposure paradigm was tested by presenting the odor mix on a cotton ball, placed in the cages of the animals for 24-hours a day. The WOM was collected at different time-points. The genes previously shown to be DE were tested by qRT-PCR and none show consistent changes in mRNA levels across time (n=3-10). T-test, FDR < 5%; * < 0.05, ** < 0.01, *** < 0.001. Error bars are the standard error of the mean. See also Figure S5.

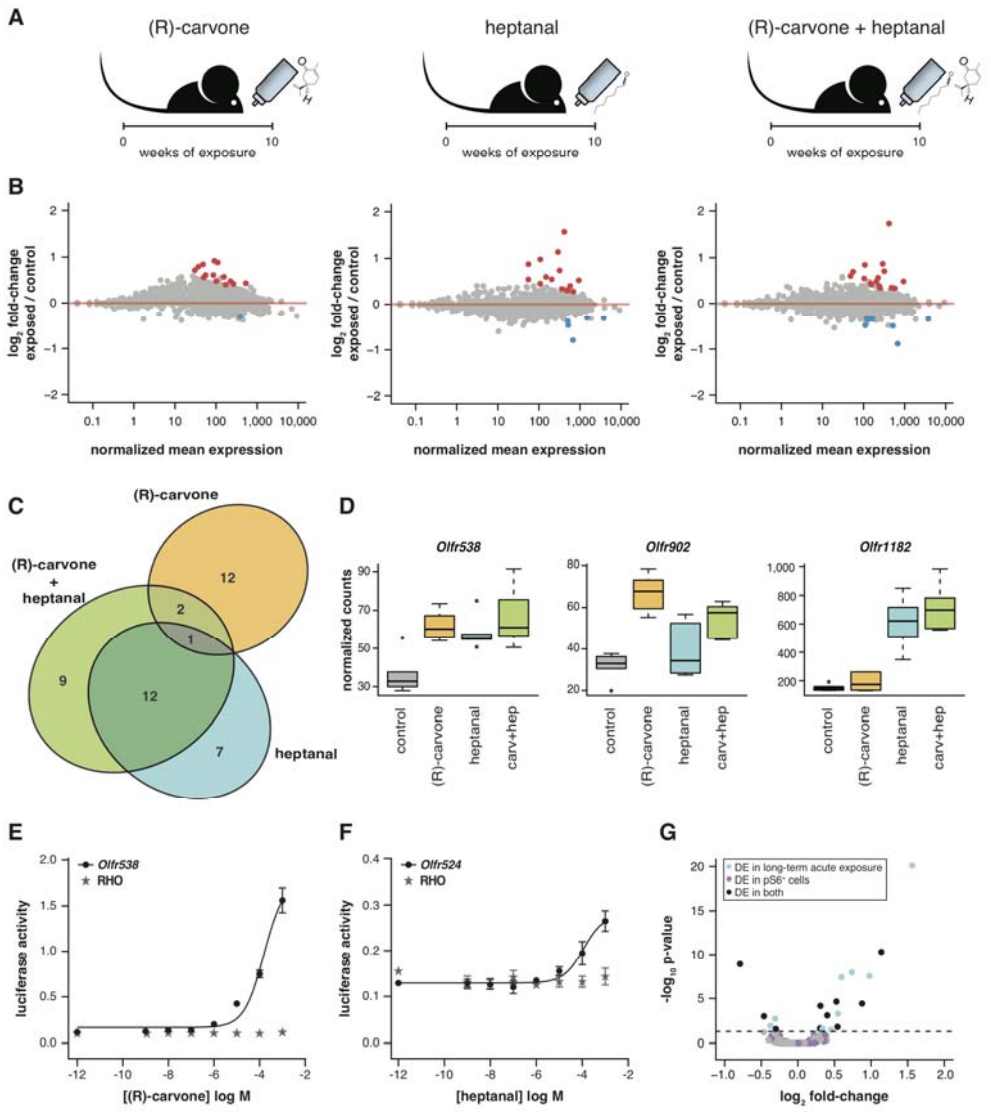

Figure 6: Odor-mediated changes in OR mRNA abundance changes are receptor specific. (A) B6 animals were acutely exposed for 10 weeks to (R)-carvone, heptanal or both. **(B)** The fold-change of exposed compared to control animals based on RNAseq data (y-axis) is plotted against the OR genes mean mRNA abundance (x-axis), for each of the experimental groups (n=6). Genes in red or blue have significantly up or downregulated mRNAs, respectively. **(C)** Venn diagram showing the intersections of the DE OR genes in each of the exposure groups. Only one OR mRNA changed in all groups; all the other are specifically altered upon exposure to (R)-carvone or heptanal. **(D)** Examples of an OR mRNA that changes in all groups (*Olfr538*), one that is specific to stimulation with (R)-carvone (*Olfr902*) and one that responds only to heptanal (*Olfr1182*). **(E)** Dose-response curve for HEK293 cells

expressing Olfr538 (black) and challenged with increasing concentrations of (R)-carvone. HEK293 cells expressing a RHO-tag only (grey) were challenged with the same concentrations of (R)-carvone as a control. **(F)** Dose-response curve for cells expressing Olfr524 (black) and challenged with heptanal, control cell responses are represented in grey. Error bars are the standard error of the mean. **(G)** Comparison of DE genes identified after 10 weeks of acute exposure to heptanal to those found via an *in vivo* deorphanization strategy. On the x-axis is the fold change of acutely exposed versus control animals with the corresponding p-value on the y-axis. The horizontal line represents the cutoff for significance. Each dot is an OR gene; those called significantly DE in both assays are shown in black, while those responding in only one experiment are in blue and purple. Half of all the DE genes in the acute exposure experiment are identified in the deorphanization assay, suggesting that the changes are indeed mediated by OSN activation by heptanal. See also Figure S6.